Sound and Meaning in Auditory Data Display

Thomas Hermann and Helge Ritter

***Abstract*— Auditory data display is an interdisciplinary field linking auditory perception research, sound engineering, data mining and human-computer interaction in order to make semantic contents of data perceptually accessible in the form of (non-verbal) audible sound. For this goal it is important to understand the different ways in which sound can encode meaning. We discuss this issue from the perspectives of language, music, functionality, listening modes and physics, and point out some limitations of current techniques for auditory data display, in particular when targeting high-dimensional data sets. As a promising, potentially very widely applicable approach we discuss the method of model-based sonification (MBS) introduced recently by the authors and point out how its natural semantic grounding in the physics of a sound generation process supports the design of sonifications that are accessible even to untrained, everyday listening. We then proceed to show that MBS also facilitates the design of an intuitive, active navigation through "acoustic aspects", somewhat analogous to the use of successive 2D views in 3D-visualization. Finally, we illustrate the concept with a first prototype of a "tangible" sonification interface which allows to "perceptually map" sonification responses into active exploratory hand motions of a user, and give an outlook on some planned extensions.**

***Index Terms*— Sonification, Exploratory Data Analysis, Auditory Perception, Human-Computer Interaction**

I. Introduction

Auditory data display denotes a rather young and rapidly evolving set of techniques also known under the term sonification to make data from a wide range of application domains accessible to auditory inspection, analysis and summarization [1]. Creating auditory data displays thus challenges us with the task to devise mappings from data to sound patterns in such a way as to exploit the highly developed capabilities of the human auditory system to uncover meaning in sound by detecting a rich variety of auditory patterns and "gestalts" (see Sec. IV-A). In this way, auditory data display offers a new and very promising tool to uncover hidden structures and meaning in massive collections of data that would be difficult to scan, explore or summarize by more conventional means.

With this goal, auditory data display can be seen as a highly interdisciplinary field at the interface between research in auditory perception, sound processing algorithms, data mining and human-computer-interaction [2], [3], [4]. From the perspective of this special issue, we will be particularly interested in the connections between sound semantics and musical listening and further, basic forms of human listening. From a more application oriented point of view, we will argue that a particularly promising aspect is the use of auditory data display techniques to aid and enhance the currently much wider established techniques of data visualization for the purpose of interactive, or *exploratory data analysis* [5], [6]. A major reason for this is that the specific properties of sound perception as compared to visual perception make auditory data displays highly suited to offer an additional route to meaning in data that is both synergistic and complementary to visualization. Particular strengths in this regard are (a) the capability of our auditory system to process several streams of information in parallel, (b) to offer a high temporal resolution, (c) its high sensitivity for structured motion, in particular rhythm and (d) its ability to function well even in noisy contexts.

Regarding the task of creation of auditory displays that are easily interpreted by human listeners, we discuss the issue of meaning in auditory displays from a number of different perspectives, ranging from language and music, function, listening modes and finally, to the semantic grounding of sounds in the physical process of their generation. After a review of existing approaches in the field we then present an approach based on a concept of user-controlled, virtual sound objects. This technique of *Model-Based Sonification (MBS)* has been introduced by the authors [7], [8] and allows for a very intuitive design of a wide class of sonification interfaces that can take important dimensions of sound semantics into account by grounding them in physical sound generating processes in a natural and user-transparent way. Whereas in the articles cited above the technical aspects of sonification systems dominate, here the relation of sound and meaning in auditory display is focused explicitely and brought in relation to the meaning of sound in other domains.

Section II discusses the meaning of sound from different perspectives including music, language, function and physics. Section III summarizes existing sonification techniques and describes the listening type used for interpreting the sound. Section IV then presents the framework of model-based sonification and contrasts it to the approaches in the previous section. The particle trajectory sonification model is presented to highlight various aspects of MBS, including the relation of sound and meaning. Section V addresses the topic of interaction with sound, caused by interaction with sounding objects. A haptic controller is presented as a means for manipulating sonification models, to control sonifications in real-time while maintaining the high-dimensional expressiveness that human hands provide. The paper closes with a conclusion and summary.

II. Sound and Meaning

Meaning in sound is what makes ears useful to their owners. The often amazingly highly developed auditory sense and its ubiquity in the animal kingdom provides telling evidence about the richness of acoustic information that can be conveyed and extracted in this important sensory domain, even in the absence of the very special capacities of language and music that give us an even enhanced perspective on sound as a carrier of

semantics. We are aware that the issue of meaning in sound is of extremely wide scope and that we can in the following only touch on a very limited part of the rich levels of meaning unfolding in the brain of a human listener

To disentangle the multitude of semantic dimensions offered by our auditory sense, let us perhaps start with the highest levels, spoken language and music, which are also evolutionarily most recent. Taking a perspective motivated by ecological acoustics [9], we will then gradually work backwards in evolutionary history to bring into view increasingly more basic constituents of auditory perception that became particularly apparent as "basic expression", and will connect these to more elementary dimensions of meaning, whose deepest roots ultimately can be seen in physics, reflecting very fundamental laws that connect physical and geometrical properties of our environment to sound characteristics in a rather universal manner, invariant over a wide range of conditions and time scales, so that evolution found ample occasion and time to imprint these regularities deeply into the brains of our predecessors and ourselves.

A. Sound and Meaning in Speech and Music

We usually find it extremely easy to listen to the narrative of another person that is using our native language. Moreover, we have the impression to listen to the same story when the speaker is reading the same text to us again, even though the visual inspection of the two sound pressure curves – which is what arrives at our ears – would hardly give us a clue to the fact that they contain the same meaning. The pressure-curve based comparison would become even more hopeless if the second pass through the story were made by a different voice although this would hardly make any difference for our immediate perception.

This example illustrates the extreme culmination point reached in our ability to extract meaning from sound patterns, provided they are drawn from a certain family of "privileged" encoding schemes delineated by the phonetic and syntactic structure of human language. If this requirement is fulfilled, our auditory system can bridge an incredible gulf that exists between the raw waveform of the auditory signal and the extremely rich semantic level of meanings that can be expressed in spoken language.

A significant part of this capacity is most likely genetically encoded in the brain areas that process language. However, another significant part is the result of learning and requires a sufficiently long prior listening experience of our native language. The same learning capacity permits us even in later life to implement a remarkable variety of different mappings, at least from the family of sound patterns spanned by the structure of human languages into the rich semantic space spanned by human narrative.

While the learnt part of meaning in spoken language is encoded in the largely conventional association between phonetic patterns and their word meanings (with the exception of some words that mimic acoustic features of processes or events that they denote, e.g., "to scratch", "to bounce", "to sizzle" etc.), there is also a substantial amount of information that is encoded along further non-verbal dimensions that are largely orthogonal to the verbal meaning of text and, therefore, can be accessible to a considerable extent also to a listener not familiar with the particular language.

While language itself is already some, albeit very coarse, indicator of membership to a particular community, finer delineations are superimposed by the different dialects, which can enable experts to spot the origin of speakers to geographic regions of remarkably restricted extent. Even without training, we can easily classify most voices as being male or female and we are accustomed not only to recognize individual persons in a highly selective way from their voice, but also to infer important additional aspects of their emotional state and even their health or momentary condition, such as being tired or out of breath.

Prosody is a major channel across which many of the above features become transmitted. It is the major feature that makes speech more impressive than writing by allowing us to annotate narrative with emotional contents that is not encoded in the choice of words alone, but in the way they are spoken. By its capacity to encode emotional information it also plays an important role in providing us with clues about the emotional state of the speaker himself.

Prosody shares its major elements with *music*: intensity, melody, articulation and rhythm. Obviously, this close relationship comes most vividly to the fore in human singing, where we see the smooth perfection from prosody to music: while most forms of singing still stick to language, the importance of the verbal layer now falls by a large margin behind the suitability of the used language as a carrier medium for melodic sounds, leading to an interesting differentiation of languages according to that criterion. Too much attention to the verbal layer may even lead to distractive interference with the musical experience itself. This, together with the particular musical characteristics of some languages, such as Italian, may explain why the inability of understanding the language of a song even may increase our readiness for its musical appreciation.

Another important layer of musical meaning may be understood from its production process: a performer controls a sounding object or instrument with the aim of expressing his or her emotions and intentions in sound. The activity of performing is in a way similar to story-telling. Meaning then becomes "condensed" in interrelation of musical elements, e.g. in harmonic, rhythmic or melodic structures. A change of tension and relaxation is created similarly as in telling a story. Musical relations, perhaps through their close relation with prosodic elements, are able to provoke emotional reactions. By the same token they are able to particularly strongly activate the listeners memory. Recognition of themes is important for binding meaning to musical sounds and most pieces of music include repetitive structures and transformations of central themes to evoke memorization [10]. Some elements of music can be easily related to emotional value, for instance consonance/dissonance (pleasant/unpleasant) or major/minor harmonies (happy/sad), but this contributes only marginally to an explanation of the relation between sound and meaning.

A feature that music shares with language is the strong role

of culture imprint for the constitution of meaning. However, in contrast to language, the interpretation of musical meaning can be extremely subjective. Besides the musical semantic value, the listener can attend other meaningful aspects of musical sounds, e.g. the quality of a musical instrument.

B. Meaning from the Perspective of Function

Meaning usually is closely related with *function* [11]. Considering language and music, their predominant functions may be seen as communication and enjoyment. However, beyond language and music, our daily life is pervaded with a rich variety of further acoustic experiences, bringing to the fore layers of meaning that stem from additional functions not primarily encountered in language or music, or exemplifying in a more genuine way functions which partly play also a role in language or music, but largely hidden under their more typical and predominant functional layers of communication and enjoyment.

The simplest and oldest function of sound is *alerting*. While for the simplest forms of alerting, such as being shocked by a very loud and sudden sound, already very simple processing can be sufficient, the high value of alerting gave rise to the evolution of much more sophisticated capabilities for extracting additional meaning from sound events that might indicate a potential threat.

A first example is the capability of *auditory localization*. Localization of sound sources is a complex computational process, and yields geometric information of crucial relevance for the rapid assessment of the closeness of danger and choice of a safe escape route. The same capability can then also be used for other means, e.g., for localizing prey, or, a task of not always entirely different character, for localizing a mating partner. In both cases, the ability of localization can benefit significantly from the ability of acoustic recognition. Already in insects we see highly developed auditory systems specialized on a remarkably accurate recognition and localization of sound signals from conspecifics and even the extraction of features correlated with important properties of the emitter, such as fitness or size.

In humans, but also in many higher animals, we find the ability not only to discriminate a very large number of different sound events, but also to rapidly learn new ones. This permits single sound events to attain *iconic meaning*, indicating events such as the slamming of a door, the arrival of a particular person from the sound of her footsteps, or the starting of a car. We also encounter conventionalized forms of acoustic icons, such as the use of bells or sirens for various signalling purposes.

Complex mixtures of natural or artificial acoustic events can be perceived as an “acoustic scenery”, telling us about the current weather, the situation on a busy city place, or what is happening in a forest. Well-trained listeners, such as blind people, impressively demonstrate the wealth of information that can be extracted in such sceneries.

A different function of sound is to aid *coordination of actions*. A classical example is the coordination of footsteps of marching soldiers. During other activities, such as brushing our shoes or when locking a door, the associated sounds provide us with feedback confirming the orderly progression of an intended chain of events. Numerous simpler interaction sounds that occur when we put two rigid objects into contact share this function of confirmative “acknowledgement” that one phase of an action, such as setting a cup onto its plate, has been successfully completed.

C. Meaning from the Perspective of Listening

Listening is an active process and humans can use auditory perception in different modes. For example, a listener can direct auditory attention to a single instrument in an orchestra performance; but he can also focus on the symphony “as a whole”. Such categories are referred to as *listening types*. Again, there are several aspects along which such categories can be formed. We will here follow a classification of Gaver (see [12], [13]) into musical listening and everyday listening, since it proves helpful for the later discussion.

If listeners attend the pitch, melody or harmonic organization or rhythmical patterns of a sound signal, they use *musical listening*. In this mode of listening, properties of the sound itself are attended to. The sound properties are not accessed to obtain knowledge about the object or instrument itself, e.g. its tension or excitation, but listening is focused on the sound itself: sound is attended to as the end and not as the means. This type of listening is investigated in psychoacoustics. Musical listening is not limited to music. For instance, listening to a bouncing ball, we can attend the rhythmical changes, the brightness of the sound and its level.

However, in everyday life, we usually experience sound in a quite different way: the very first thing we usually try is to identify the sound source and to generate a mental model about what interaction could have happened to cause the sound. At the same time, we identify the relative location of the sound source and are possibly concerned with an appropriate reaction. From the perspective of evolution, this source-oriented interpretation appears highly plausible. People that are asked to tell what they hear, frequently use a description of an imagined sound source or process and only rarely a characterization of acoustic properties as they are addressed in musical listening. For example, “the sound of a big metal gong” as a reply is more common than “a mixture of decaying tones with decreasing brightness”. Everyday listening is performed permanently without directing any effort to the listening process.

Besides these two types of listening, a third type shall be introduced now, *analytical everyday listening*[8]. In contrast to everyday listening, here the focus is not on an adequate reaction, but on learning about properties of the sound-producing process. When we shake an opaque box and try to guess its contents from the sound, we use analytical everyday listening. Listeners are quite good in discerning various attributes in analytical everyday listening, like size, shape, velocity, material of colliding objects, or the underground of rolling objects [14]. In contrast to everyday listening, a high amount of attention is directed to the event that caused the sound and the object is explored by using its sound. Obviously, this type of listening becomes very central when considering sonification.

D. Meaning from the Perspective of Physics

While we have seen that many aspects of meaning in sound, particularly in language or when using sound to transmit signals, have their origin in conventions, we also saw that there are numerous other layers of meaning whose origin appears to be less arbitrary. This is particularly true, when the meaning of an acoustic event is primarily rooted in conveying information about important physical properties of an object or process.

A major class of such events are *interaction sounds*. Beyond their already mentioned significance of providing confirmatory feedback, they also allow us to discriminate a remarkable number of object properties, including material, such as metal, plastic or wood, geometric properties, such as wall thickness of a drinking glass or grain size of gravel in a box, state properties, such as a filled or an empty bottle or even the presence of a crack in a plate.

Additionally, interaction sounds also convey important information about the relative movements bringing the objects into contact. We get important clues about the forcefulness of the event, and we can distinguish different geometric motion patterns, such as hitting, sliding, rolling, tottering etc.

The roots for our ability to access these many facets of meaning can be found in the laws of physics. Any sound that is generated is the product of an oscillatory process in the physical environment. Mechanical excitation and energy transfer from the object via air pressure waves to the listener's ears are the fundamental connection between physics and listening. In the case of a contact sound, the impact excites two physical objects. The stronger the impact, the more energy will be exchanged between the objects, and the higher will be amplitude of the objects vibrations, leading to sounds of a higher level. The frequency spectrum of the caused vibrations can be a complicated function involving the stiffness of the involved material, its density and its geometric shape, but also the locus of the impact point. Further properties, in particular, energy dissipation due to internal or external friction, become reflected in the sound amplitude envelope.

Given a detailed specification of the sound generating event, the laws of physics provide all the necessary information to compute the generated sound from first principles [15]. The resulting computational link between the aforementioned situational features and the emitted spatio-temporal sound pattern constitutes a so-called *forward model* of the sound generating process. The situation is analogous to computer graphics, where physical laws for light reflection can be used to compute the visual appearance of objects to a high degree of accuracy. One drawback is that such models, by their use of first principles, can be computationally too heavy for many purposes, e.g. real-time operation at a high frame rate. This has motivated techniques for creating more approximate models, often working directly on more global sound features, such as the temporal shape of the energy distribution in different frequency bands, or even on short patches of recorded real sounds that are then suitably filtered and blended together.

However, to uncover meaning in sound requires the inverse modeling path, i.e., to infer from sound patterns the features that caused their emission. This is more complicated than forward modeling, since, as in other modalities as well, connection between an effect and its cause usually is non-unique: different causes can produce mutually indistinguishable sounds. Resolving this ambiguity succeeds only with additional a-priori-information (or, in their absence, by making assumptions) about the sound source. For instance, when hearing repetitive noises, many interpretations are possible. With the additional information of being in a stairway, a likely cause are footsteps of a person, and if we additionally know that we are in our own house, the required inverse model may be restricted even further to the identification of a member of our family.

From a more extreme position, the laws of physics themselves can be viewed as a kind of context information for extracting meaning from sound events. Compared to other contexts, the context given by physical laws was stable all the time, so that evolution had ample time to adapt our brains extremely well to the ways how physics links sounds and their causes. This is reflected in a number of rather "universal" relationships that are deeply engrained in the way we — usually subconsciously — pick up meaning from sound events. They involve a number of very basic sound attributes, such as intensity, frequency, envelope and further temporal aspects of a sound signal that contain cues about a situation.

Intensity is a very direct signifier of the amount of power (in the very literal sense) that is in the cause of a sound event. This has biased our perception towards associating danger with very loud sounds.

Frequency is strongly correlated with two different features of a sound source: the natural oscillation frequencies of an object decrease with its mass and its size; they increase with its stiffness and its tension. Therefore, high frequency alone could signify high tension and therefore danger, but also a small and therefore relaxing harmless sound source, while low frequency tones would signify big and potentially dangerous sound sources or low tension and soft material and therefore low danger.

These opposing interpretations can be disambiguated by the simultaneously observed intensity. As a result, pitch at the extremal ends of the frequency spectrum reinforces the threatening character of intense sounds and the comforting character of weak sounds.

Additional clues are provided by the sound envelope. A short and sharp envelope indicates rapid change and high dissipation and is typical for situations involving high forces and stiff materials, factors tending to be correlated again with danger. Sounds of long duration, with only weak gradients of change are an indication of the stability of a situation and thus may – contingent on other context factors – be felt as comforting.

Further strong cues are contained in the temporal evolution of a sound. Since size and stiffness of an object usually are rather constant, an increase of frequency of a tone is an almost certain indicator of a built-up of force and tension and, therefore, can be a warning that we may be approaching a critical event, such as the breaking of some support structure. Conversely, a decaying pitch signals that we may be receding from a critical situation. The same pattern is also caused

by the Doppler effect [15] (although the underlying physical mechanism is entirely different): sounds of a very rapidly approaching object are shifted towards higher frequencies, with a rapid drop in frequency when the object has passed by and is receding. For similar reasons, fluttering noises indicate an element of undecidedness or uncertainty by tending to be correlated with causes in which some weak material is involved. Analogous remarks can be made regarding gradients in the temporal spacing of discrete sound events.

By virtue of their strong signalling character, these very basic patterns are also present in our prosody when we express emotions, and they are consciously exploited in music in order to convey a built-up of tension (increasing pitch, loudness; speedup of rhythm) or provoke a calm and comforting atmosphere (”warm” sounds of a low frequency, slow-down of rhythm).

Below, we will argue that the same universal relationships provide important design guidelines for the creation of auditory displays in such a way that they facilitate an immediate and natural perception of meaning event without prior training. Training then can serve the purpose to enhance our discrimination with the aid of additional attributes which – similar as in language – may be have their origin in pure conventions. An interesting intermediate position is occupied by sounds that derived their semantic significance not by the above, very universal physical relationships, but still from conditions which have already become either “hard-wired” into our brain or learnt from extensive everyday experience. A rich reservoir for such sounds is provided by human language, which certainly comprises many learnt features, but most likely also a considerably number of perceptual patterns rooted even more deeply by evolutionary processes.

III. Auditory Displays

The oldest approach and most direct approach to obtain an auditory display of a given data set is to use the data values directly as a series of sound pressure values. This technique is called *audification* [1], and is usually applied to time series data, where the data set is naturally sorted by a time attribute, e.g. seismic data [16]. Necessary parameters are a time compression factor and a level scaling factor. Filters are usually applied to preprocess the sound further. The technique can be extended to a high-dimensional data display either by mixing different audifications together or by using a multichannel sound system.

Although the generation of audifications is very simple, it already makes a number of useful data properties directly accessible to the human ear: the variance of the data becomes audible as sound level, data set size as duration, and pitch and timbre can reflect many aspects of the detailed time-resolved variation of the series. Obviously, by attending to these attributes, musical listening is used to interprete such sounds.

However, due to its simplicity audification is only applicable for limited sorts of data sets and requires many data points to deliver reasonably long sounds. Adapting the generated sounds to the perceptual characteristics of the ear is restricted to scaling and filtering. Therefore, audification is mainly useful for data in which important regularities are already reflecting temporal variations which happen to match well with the perceptual capabilities of the human ear, such as, e.g., periodicities.

Significant more flexibility for tailoring sonifications to the capabilities of the human ear is gained with *parameter mapping* [17], which is the currently dominating sonification technique. Parameter mapping sonifications are generated by superimposing data-driven sound events, e.g. instrument sounds, according to given parameters like onset time, duration, pitch, amplitude. Each data point now is mapped into the parameters of a separate sound event, which gives the method its name and offers much more flexibility than audification, since both, the underlying instrument sounds as well as the data-to-parameter-mapping can be specified by the designer of the sonification according to the special needs of the data analysis task at hand. Obviously, parameter mapping sonifications again are based on passive, musical listening and they can equally easily also be generated for data points of arbitrary size and dimensionality. However, the increased flexibility also comes at a price: without explicit knowledge of the employed mapping a parameter mapping sonification may be very difficult to interprete. Moreover, the specification of a good mapping can turn out to be a non-trivial requirement in many applications and the dimensionality of the display is fixed and given by the number of parameters of the chosen mapping.

With increasing complexity of the auditory events, they may be recognized and used in isolation to convey meaning in an auditory display. This idea is followed in *Earcons*, a very different sonification technique [18]. Earcons are auditory patterns usually composed of musical sounds, that represent a message in a short musical motive. Therefore, the association from an earcon to its meaning has to be learned. Again, musical listening is used to process earcons. Themes recognition is required to infer an earcon’s meaning. With regard to semantics, earcons are similar to linguistic sounds: each earcon represents an entire message of its own; several earcons can be combined into a sequence to represent more complex messages just as words can be combined to generate a sentence. This makes earcons very suitable to convey symbolic messages, but limits their use for displaying continuous-valued or high-dimensional data items. However, sonifications of such data by other means can benefit from earcons by embedding them as symbolic acoustic markers to annotate particular parts of the underlying continuous sonification.

Auditory icons[19] follow the same purpose as earcons, to convey abstract symbolic messages by using non-speech audio. In contrast to earcons, they do not base their meaning on a mere convention (that can only be acquired by learning), but instead employ a crisp sound metaphor to encode their message. For example, a trash can sound can be an auditory icon to confirm the deletion of a file on the computer desktop. This kind of encoding also offers an additional benefit: unlike the so far discussed auditory displays, which all require rather attentive musical listening, interpretation of auditory icons succeeds already with the less demanding everyday listening.

The main problem with auditory icons is that for many messages (e.g., "silence") it can be very difficult or even impossible to find an adequate sound pattern. Similar as for earcons, this auditory display is not really suited for presenting high-dimensional data sets.

Parameterized auditory icons are an extension that borrows some additional features from parameter mapping in order to convey additional analog information by suitably controlling the parameters of the icon sound [19]. In the example above, the two parameters sound level and sharpness of the trash can sound could be made to reflect the size of the deleted file and the elapsed time since the most recent modification date. Parameterized auditory icons preserve the advantage of easy understandability by their users since the metaphorical association facilitates the reference from the sound to its meaning. This can be made true even for their analogical part if the parameter mapping succeeds to reflect physical properties that admit a natural relationship to sound attributes, as discussed in subsection II-D.

Although useful in many situations, the above sonification techniques still suffer from some significant limitations: Audifications, earcons and auditory icons are not suited for generic high-dimensional datasets, since they can reflect only a small set of carefully selected attributes. This limitation is not shared by parameter mapping sonifications, but only at the price of burdening the user with a complicated mapping specification that must be kept in mind by a highly attentive and musical listener in order to interprete the sound w.r.t. the data. Even then, the simultaneously displayable number of dimension usually is limited to about 20.

In the next section we will describe *Model-based Sonification* (MBS), a very versatile framework for sonification that the authors have developed recently in order to cope better with most of the above limitations. MBS can be applied for a wide range of data types and application situations. It offers a very high amount of flexibility to create sonifications that can be made well-adapted to the discrimination and learning abilities of human listeners.

IV. Model-Based Sonification

The motivation for *Model-based Sonification* (MBS, [8]) was the desire for a principled connection between data and sound, a generic strategy which on the one hand allows auditory displays for arbitrary data sets concerning dimensionality and size, and on the other hand to provide – from the design of the sonification technique – a natural means for interacting with a sonification system.

In the new framework of model-based sonification these two objectives are achieved by using a *parameterized sound model* as the central device to create the auditory display. This sound model can be imagined as a *virtual object* responding with sound, for instance when being "struck" by the user. This offers a very flexible, two-level design approach for a sonification: the specification of (i) the virtual sound object (characterized by a range of acoustic modes and involving the specification of how the data determines the concrete setup), and (ii) the specification of how the user interacts with the virtual object in order to query and explore its properties from the sound. To get a first picture of the range of possibilities opened up by MBS, let us look first a bit more closely at possible ways to fill the design steps (i) and (ii).

A first way chooses the virtual sound object and its modes of interaction in close analogy to familiar physical situations, such as, e.g., striking of a drumhead. Even in this case, the interaction rules need not be precisely confined to what physics would permit in the real world. Instead, one may introduce modified or additional laws to accentuate the perceptibility of particular data properties. Examples might include data-dependent modifications of the drumhead's shape, membrane tension or damping properties, effects, that would be difficult or impossible to implement in reality. Here, a major important point is, that any such modifications can act in a familiar context of a physics-based sound generation process. This can significantly aid the understandability and learnability of the resulting MBS.

A second way would exploit the freedom of creating sound generation processes in a virtual world more aggressively by lifting restrictions such as the three-dimensionality of ordinary space, the limits of familiar materials, their internal dynamics as well a constructive constraints, such as limits of realizability of unorthodox, e.g. fractal, geometries and the like. Even then, significant parts of such models can still embody general process structures of familiar physical processes, although in a virtual world of otherwise possibly strange "physical" laws. This gives them a decisive advantage over purely abstract parameter mapping techniques while at the same time offering a tremendous amount of freedom in "sculpting" the sound generation process in a cognitively penetrable manner.

Both of the above two approach styles permit sonifications that are well suited for analytical everyday listening. However, from a perspective of music, the specification of the virtual sound object(s) also shares many analogies with the construction and tuning of a (in this case virtual) musical instrument, whose detailed properties are, however, determined and parameterized by the data set at hand.

This musical analogies provides designer and user of MBS with further rich possibilities for selecting model classes in such a way that they can benefit from musical listening skills as well. In this way, the new framework of MBS can address most of the problems encountered with parameter mapping and provides a qualitatively complementary link between data and their acoustic representation. However, as a subset of MBS, the data may be transformed to entities in model space that act on other acoustic model components, such that actually the data stream literally "plays the virtual instrument", a perspective which relates MBS and parameter mapping sonification.

Specifically, MBS offers the following advantages:

- **Limited number of parameters:** a sonification model can be formulated so that only few parameters need to be tuned. The number of parameters only depends on the model. In contrast, parameter mapping needs as many parameters as available sound attributes and is only capable to represent data of that dimension without loss.
- **Semantic grounding of parameters:** while parameters in parameter mapping are related to sound attributes,

parameters in MBS control *physical* source attributes. They may affect the sound in a complex way, but since the model is always grounded in a physical sound generation process that can be familiar from everyday experience, the connection between sound and data can be made to appear natural and easy to pick up.

- **Good Learnability:** As a consequence of the previous point, MBS inherits all the strengths of parameterized auditory icons while lifting their limitations through the strongly increased flexibility offered by the two-stage MBS design process. Compared to parameter mapping, the sounds of a sonification model are much more coherent in structure with different data sets. Thus the listener can rapidly become familiar with the sounds of a model and improve in perceiving subtle patterns.
- **Generality:** sonification models can be formulated so that they operate on data of arbitrary dimensionality and dataset size.
- **Intuitive time axis:** time matches to temporal evolution of the model and is thus intuitively related to changes or events with the process described by the model.
- **Intuitive interface:** sonification models can offer many flexible and natural "handles" for the access to and manipulation of the sound generating process and can use concepts grounded in a physical world. Sound is used as a feedback to user actions, which matches to our expectations from manipulating objects in the real world.
- **Active User:** interaction rules connect the user's actions with the sound feedback. Since user interaction may not only provide excitations of the model but also continuously control model parameters, MBS supports a new style of active data exploration with the model-driven auditory display in a closed loop with the user.
- **Ergonomic factors:** Avoiding annoyance by auditory data display is a crucial issue. If the sound is the system answer to a user's action, as in MBS, the annoyance is reduced – in fact, users may get so much used to the auditory feedback that it is missed if it is absent.
- **Symmetry:** sonification models may be designed to be invariant to transformations of a data set that have no semantical relevance, such as global rotations or scaling, which is impossible, e.g., for audification. Their design makes it also easy to respect symmetries in data space. In parameter mapping, such symmetries are for instance broken by assigning a single attribute to the time axis.

The detailed specification of a sonification model and its parameterization principles which describe how to incorporate the data set, may at first sight appear complicated. However, one should note that one needs to go through the full development cycle only in rare cases – typically only once, when creating a new model. From a practical view point, one would start with a library of different types of sonification models, geared towards different families of data types and analysis purposes. For a concrete application then one would only need to tailor a chosen model to the detailed specifics of the task at hand after which the sonification could start. Still, when need arises, MBS offers the necessary breadth to develop highly optimized sound models that can then act as very specialized "resonators" to endow the user with an "acoustic fovea" that can support his natural acoustic perception with highly domain specific auditory zooming capabilities that may be required to solve very delicate data analysis tasks.

In the following section we will illustrate some of these aspects more closely with a concrete example intended to support interactive cluster analysis. The example is intentionally chosen as rather simple in order to exemplify how already very few ingredients suffice to obtain an acoustically rich and versatile sonification model.

A. The Particle Trajectory Sonification Model

Our example uses as its sound generating process a model of the motion of a number of ficticious particles under the influence of a force field that is created from the data points of the data set under analysis [7]. We use this example for a concrete illustration of five general aspects that must be specified for the complete definition of a sonification model: (i) the *setup*, (ii) the *dynamics*, (iii) the *sound-link variables*, (iv) the *listener characteristics* and (v) the *interaction types*.

Setup. In the present case, we define the model setup a potential function V, which we choose as a superposition of distance-dependent potentials

$$V(\vec{x}) = \sum_{i=1}^{N} \phi(\vec{x} - \vec{x}_i) \ , \tag{1}$$

centered at the given data points $\vec{x}_i,\ i = 1, \ldots, N$.

Intuitively, it makes sense to restrict each data point's potential contribution $\phi(\vec{x} - \vec{x}_i)$ to the vicinity of its location $\vec{x}_i$. This motivates the choice

$$\phi(\vec{x}) = -m_p m(\sigma) \exp\left(-\frac{\|\vec{x}\|^2}{2\sigma^2}\right), \tag{2}$$

where m_p is a particle mass, $m(\sigma)$ the mass of a data point and σ is a bandwidth parameter. Different from the gravitational $1/r$ law, here a negative Gaussian is taken for two reasons: numerical instabilities are avoided since ϕ has no singularity, and the approximately parabolic shape of ϕ close to the origin gives rise to harmonic (pitched) sounds as will become clear soon[1].

Dynamics. In the present model, we specify the dynamical elements with a set of ficticious (test) particles, injected into data space to probe the potential V. For the particles' dynamics we choose Newton's law of motion with a damping term

$$m_p \ddot{\vec{x}}(t) = -\nabla_{\vec{x}} V(\vec{x}(t)) - R\dot{\vec{x}}(t) \tag{3}$$

where R is the resistance constant and m_p the particle mass. Due to the damping term $-R\dot{\vec{x}}$ the particles' kinetic energy decays until they come to rest in a local minimum of V. If the dataset exhibits a cluster structure, such minima will tend to be located near cluster centres. The parameter σ will control the scale at which clusters are seen: potential "valleys" of data

[1]The deviation from parabolic shape at larger distances, however, is important. Without it (i.e. all $\phi()$ purely quadratic), $V(\vec{x})$ would be quadratic as well and thus most of the information in the positions $\vec{x}_i$ would become "averaged out".

points closer then σ will "fuse" into a common, large valley while empty regions extending over distances significantly larger that σ will separate different clusters.

Sound-Link Variables. The sonification is simply obtained by adding the kinetic particle energies $E_i^{kin} = m_p \dot{x}_i^2/2$ of all particles $i = 1, \ldots, N$, giving $\sum_i E_i^{kin}$ the role of the sound-link variable. Since the kinetic energy is always a non-negative number, the sound signal will show a DC bias, which can easily be removed using a high-pass filter.

Listener Characteristics. Although the sonification model uses a spatial description, in this model the listener shall not be located into listening space: all kinetic energy terms contribute with the same weight to the sound and therefore the model may be denoted as non-spatial.

Interaction Types arise from the model definition: one can either 'throw' particles into model space, or 'hit the model' to increase all particle energies. Further possible interactions are discussed in Section V. Currently, only the first excitation type is implemented and the resulting sounds will be described.

To get an intuitive picture of the sound generation process, imagine a single particle moving around in V of a data set with Gaussian distributed data points. Fig. 1 shows a typical 2D-projection of the particle trajectory. If the particle passes through a minimum of V, its kinetic energy has a maximum. Like in a pendulum, kinetic and potential energy are transformed periodically so that the kinetic energy as a function of time shows an oscillatory behavior, audible as sound. If the potential function V would be a harmonic potential, i.e. V is a quadratic form, the newtonian dynamics would lead to damped sinusoidal sounds [15]. The nonlinearities of ϕ however cause the period to be longer if the particle reaches the tails of a potential trough, since the restoring force decays with distance to the data. So the sound of a particle is characterized by pitched sounds with an increasing pitch, converging to a pitch value that is determined by the curvature of V near a cluster center. Sound example S1 illustrates this behavior[2].

Let us assume that a particle moves around in a data space with data distributed according to a mixture of two normal distributions with different mean and covariance. Furthermore

[2]The sounds can be found on the web site [20]

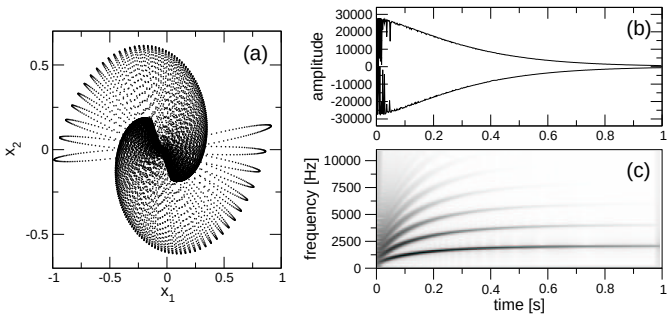

Fig. 1. Particle trajectory sonification connects system dynamics and auditory representation. The plot (a) shows 5000 steps of a typical particle trajectory in the data potential V for a clustered dataset in 2D. (b) shows the obtained sound signal by lowpass filtering the instantanous kinetic particle energy. A spectrogram is shown in (c). The pitch stabilizes during convergence of the trajectory to the mode of V.

assume that σ is large enough so that the individual point potentials "fuse" to yield only two separate potential troughs in V. Then a particle of energy $E(t) = V(x) + E_{kin}(t)$ will be able to move within the limited domain where $V < E(t)$. Initially, it will perform quasi-chaotic motions that contribute to the sound signal with a noisy chaotic pattern. Gradually the initial energy decays so that the particle becomes caught in one of the two potential troughs. As a result, the sound pattern turns into an increasingly harmonic oscillation that finally fades out as a pure sinusoidal with a frequency proportional to the curvature at the mode. Thus clusters of higher mass lead to increased attraction on the particles, the increased tension resulting in higher pitched tones, while the broader valleys of larger sized clusters will give rise to lower sounds, again in accordance with the physical semantics of everyday sounds.

From single particle sounds only limited information about V can be withdrawn. This changes when using an ensemble of particles, since there will likely be particles that converge to different clusters and thus contribute to the sonification with a different characteristic sound, making the clustering structure audible from the polyphony of the sonification. Sound examples S2 and S3 (see [20]) illustrate such sounds for a dataset with one and with three clusters.

Obviously, a very limited number of parameters are required to control the model: the particle mass m_p, the initial energy E_0, the resistance constant R and the bandwidth σ. They all have a clear meaning for the model, and the resulting sound changes on parameter variation are intuitively understood. E.g. it is obvious that by increasing R the sounds will decay faster while choosing a larger particle mass will shift the sounds towards a lower frequency. A particularly interesting parameter is the bandwidth σ that controls the spatial resolution of probing the data. With very large values of σ, V looks like a single (scaled) ϕ potential. At intermediate values, V reflects data clusters as separated smooth potential troughs. Finally for very small values, V contains as many local minima as there are data points, all having the same shape of ϕ. Such a resolution parameter is well suited to be controlled interactively according to interests on the data. Fig. 2 shows a sonogram of a sequence of 30 particle sonifications, obtained for a geometrically spaced set of σ-values decaying with time, remaining constant during the single particle sonifications whose start can be seen from the vertical bars in the figure. Since the sonification model mimics a physical process, meaning and sound are related as in a physical analogue: excitations cause an acoustic feedback and the sound level decays with time. From model design and from understanding how the model works, it is evident to relate the perceived pitch to the cluster mass (number of data points that contribute to a cluster) and to the cluster variance: the higher the cluster mass, the stronger the restoring force that attracts a particle, the larger the cluster variance, the lower the pitch. Similar as with real-world object interaction, stronger excitations will cause louder sounds, and thus interacting with the sonification model addresses the same perceptual skills as we use in analytical everyday listening.

A second important way - complementary to analytical listening - is based on auditory gestalt perception which can

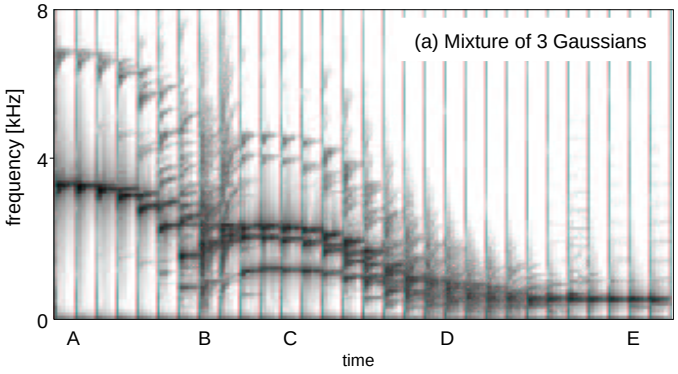

Fig. 2. Sonogram of a bandwidth sweep of particle trajectory sonifications. As σ is decreased the clustering structure becomes audible. A plateau of the pitched particle sound contributions, visible near 'C' reflects the presence of stable clusters at that length scale. The corresponding sonification contains are polyphonic texture at the corresponding time.

occur as a result of repeated experiences with a sonification model. The concept of auditory gestalts is in analogy to visual gestalts: a subset of acoustical elements perceptually bound together into a "unit" as a result of a particular coherence, characterized by one of the "gestalt laws", e.g. similarity (e.g. of timbre), good continuation (e.g. of pitch), common fate (similarity of changes, common onset of tones). For more on gestalt laws and a further discussion, see e.g. [2]. Any sound pattern that is discerned as such a gestalt can be related to knowledge about the underlying structure of the data. From experiencing the sonification model with a large number of different data sets, the listener can gradually learn to relate the perceived sound to the known structure of the data, and in this way develop semantic categories that are not only related to a single acoustic attribute but to the sound as a whole. Sonification models support this learning processes by supplying an invariant process to be used in the same manner for very different data sets.

V. Sound and Interaction

Most everyday activities are accompanied with sound feedback. Every keystroke on the keyboard as well as any foot step causes an acoustic result. Humans can actively elicit acoustic feedback by interacting with objects in various ways, such as hitting, rubbing, scratching, plucking, shaking or deformation of objects. Most of the interactions can be varied in strength, duration or location and thus represent multi-dimensional queries to the object properties. In real-world situations, the following aspects are most relevant: (a) immediate response - the sound corresponds directly to actions with a latency of less than 10 ms, e.g. the contact sound when putting an object on a table signals that a motion is finished, (b) information - the sound delivers often useful information for performing a task, e.g. using a drill – absence of sound would cause vagueness, (c) control loop - acoustic feedback sounds can provide valuable cues to refine an action or to "to keep it on track", e.g. filling a bottle with water.

An interesting analogy can be drawn with visualization. In sonification, each single interaction sound contributes a single "auditory aspect" of a situation and, therefore, appears as analogous to a visual 2D view on a scene. We are accustomed to an active navigation through a sequence of such views in order to pick up the 3D layout of the depicted situation in computer-assisted visualization. Very similarly, we should expect that active user-control for navigating through a sequence of auditory aspects, offered in the form of interaction sounds in a sonification, will play an important role in gaining a more complete understanding of a collection of data items from listening.

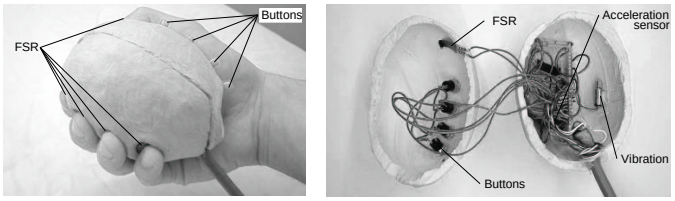

Fig. 3. Picture of the haptic interface ball for controlling and interacting with sonification models. It contains force sensitive resistors (FSRs) two 2D accelerometers and a motor to generate active vibration.

While in vision it is sufficient to make view point and view direction controllable, the different nature of sonification may benefit from the simultaneous, coordinated control of a larger number of parameters. While a suitable set of widgets in a computer screen GUI may provide an obvious starting point, a much better interface would directly use our ability to carry out complex hand movements that involve the rapid coordination of more than 10 degrees of freedom in each hand (this is a conservative estimate, taking couplings among joints into account).

While a dataglove provides an obvious input device a more challenging approach uses visual-based hand posture recognition for realizing such an interface. For an initial prototype example system, see e.g. [21]. Here, we wish to describe results of recent work along a different line, aiming at complementing the camera-based approach with an elastic, palm-sized interface ball. While the camera-based approach is contact-free, here we pursue the goal of an interface that offers a more "physical", mixed-reality interface for interacting with data. Thus, instead of providing just a controller for manipulations in the computer, the interface shall take the role of a *tangible, physical representation* of the data. To achieve this, accelerations imparted to the interface ball must be mapped onto suitable "acoustic aspects" of the data, and these response patterns must be computed with perceptually negligible delay in order to create for the user a convincing perceptual illusion that the sounds are to be attributed to the motion of the interface ball.

The current prototype of the device is shown in Fig. 3. In an ergonomically shaped housing formed with air-drying plasticine, force resistive sensors (FSRs) for each finger and two 2D accelerometers are mounted orthogonally. Additionally, four buttons are provided which may be programmed to provide additional commands for the user-interface. The FSRs permit to sense when the ball is being squeezed, while the accelerometer allows to track (by temporal integration) orientation (rotation) and (with a high-pass filter) to detect sudden impacts like hitting or striking the ball. In a future

version, we also will mount a pad connected to four piezo-mechanical sensors in order to resolve spatio-temporal impact patterns.

Our first sonification model for this ball, described in [22], uses the model of a *data-solid* to explore high-dimensional datasets from binary classification problems trough actively elicited interaction sounds. We employ a growing neural gas [23] to condense the given data set to a more manageable number of prototypical data items. Each data item is considered as a small 'data grain' with material attributes assigned according to the predominant class label among all data points which have the data grain as its nearest neighbor. The data grains are considered to be elastically bound to their equilibrium positions in the high-dimensional feature space. User-imparted shaking motion of the interface ball will set the data grains into corresponding oscillatory motion. The resulting contact sounds of colliding data grains are rendered in real-time according to the properties, with timbre determined from the grain material of colliding objects and level dependent by the relative grain velocities. In this way, the interaction generates an acoustic feedback that provides the user with the perceptual illusion that the data items are "inside the ball" and that the collisions are the direct result of his shaking movements. This allows him to probe in a very direct and intuitive way the nearness of adjacent cluster borders and their population density with data items from different classes.

The above sonification model based on particle trajectories will permit as a next step to compare different potential shapes with regard to their utility for listening-based, physical exploration of interesting features of data distributions in an abstract feature space. In the simplest case, the potential V would become fixed to the coordinate frame of the ball. Translations then cause a shift of the probing particles in the model. Shaking interaction removes the particles from their equilibrium position and makes them contribute to sound. Hitting the ball just provides additional kinetic energy to the particles. Spatially resolved hitting can be used to activate particles in different regions of the data space, e.g. by identifying the ball axes with the first two principal axes of the dataset. Squeezing the ball may be assigned to controlling the bandwidth parameter σ – another mapping of a highly intuitive character. Implementation of these modes of interactions with the ball are currently on its way and sound examples will be reported on the web site [20].

VI. Conclusion

This paper has addressed some aspects of the important relation between sound and meaning in auditory data displays and their role for using auditory data displays to support exploratory data analysis of high-dimensional data sets. Having started with the perspectives of speech and music, we have considered the connection between various functional roles of sound and its meaning, taken a closer look at different forms of listening and finally discussed which aspects of meaning in sound events can find their ultimate semantic grounding in the laws of physics that govern their creation.

From a brief review of existing techniques of auditory displays we derived the conclusion that most of them so far focus on the use of musical listening, while offering only little possibilities for everyday listening which would have the advantage of a reduced need of training. More seriously, the majority of techniques faces problems for the sonification of high-dimensional data sets and no existing technique so far considers the important issue of active, user-controlled interaction with sound.

As a promising new alternative, we have presented a new approach to data sonification which offers a generic strategy for creating auditory displays for arbitrary datasets in close combination with natural means of interacting with the sonification system. This new framework of *model-based sonification (MBS)* achieves these two objectives with a parameterized sound model deriving its strength from its grounding in an intuitive, physical picture of the sound generation process. We argued why it permits insightful sonifications also for analytical everyday listening and can help to solve some of the problems pertaining to the existing approaches discussed previously.

We have illustrated the MBS approach with an example using particle trajectories for sonification. Considering the issue of interaction, we compare sequential, user-controlled interaction with sonification models to navigation in visual scenes and report about a first prototype of a "tangible physical representation-interface" in the form of an elastic ball equipped with pressure and acceleration sensors, which allows to "perceptually map" sonification responses into active exploratory motions of shaking and squeezing. We conclude with an outlook on ongoing work towards the use of the trajectory model in conjunction with the interface ball.

From our experience so far we are very confident that the approach of MBS is very suitable to open up interesting new directions for utilizing sound in exploratory data analysis. We hope that the method can stimulate the invention of new families of sonification models, their refinement and their final organization into useful and versatile toolboxes offering researchers new ways to explore data for various properties like clustering, intrinsic dimensionality, non-linear dependencies, class borders and the like.

References

[1] G. Kramer, Ed., *Auditory Display - Sonification, Audification, and Auditory Interfaces*. Addison-Wesley, 1994.

[2] A. Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound*, MIT Press, Cambrigde Massachusetts, 1990.

[3] M. Kahrs, Ed., *Applications of Digital Signal Processing to Audio and Acoustics*, Kluwer, 1998.

[4] U. M. Fayyad et al., Ed., *Advances in Knowledge Discovery and Data Mining*, MIT Press, 1996.

[5] J. W. Tukey, *Exploratory Data Analysis*, Addison-Wesley, 1977.

[6] S. H. C. du Toit, A. G. W. Steyn, and R. H. Stumpf, *Graphical Exploratory Data Analysis*, Springer-Verlag, New York, 1986.

[7] T. Hermann and H. Ritter, "Listen to your data: Model-based sonification for data analysis," in *Advances in intelligent computing and multimedia systems, Baden-Baden, Germany*, G. E. Lasker, Ed., pp. 189–194, Int. Inst. for Advanced Studies in System research and cybernetics, 1999.

[8] T. Hermann, *Sonification for Exploratory Data Analysis*, Ph.D. thesis, Bielefeld University, Bielefeld, Germany, June 2002.

[9] K. Wrightson, "An introduction to acoustic ecology," *Soundscape - The Jounal of Acoustic Ecology*, vol. 1, no. 1, pp. 10–13, 2000.

[10] M. Minsky, "Music, mind, and meaning," *Computer Music Journal*, vol. 5, no. 3, 1981.

[11] G. Kramer, "An introduction to auditory display," in *Auditory Display*, G. Kramer, Ed. ICAD, 1994, pp. 1–79, Addison-Wesley.
[12] W. W. Gaver, "What in the world do we hear? An ecological approach to auditory source perception," *Ecological Psychology*, vol. 5, no. 1, pp. 1–29, 1993.
[13] W. W. Gaver, "How do we hear in the world? Explorations in ecological acoustics," *Ecological Psychology*, vol. 5, no. 4, pp. 285–313, 1993.
[14] M.M.J. Houben, D.J. Hermes and A. Kohlrausch, "Auditory perception of the size and velocity of rolling balls," IPO Annual Progress Report 34, pp. 86-93, 1999.
[15] P. M. Morse and K. U. Ingard, *Theoretical Acoustics*, McGraw-Hill, New York, 1968.
[16] F. Dombois, "Using Audification in planetary seismology," in *Proc. of the 7th Int. Conf. on Auditory Display*, N. Zacharov J. Hiipakka and T. Takala, Eds., Helsinki University of Technology, 2001, ICAD, pp. 227–230, Laboratory of Acoustics and Audio Signal Processing.
[17] C. Scaletti, "Sound synthesis algorithms for auditory data representations," in *Auditory Display*, G. Kramer, Ed. 1994, Addison-Wesley.
[18] S. A. Brewster, P. C. Wright, and A. D. N. Edwards, "A detailed investigation into the effectiveness of earcons," in *Auditory Display*, G. Kramer, Ed. ICAD, 1994, pp. 471–498, Addison Wesley.
[19] W. W. Gaver, "Using and creating auditory icons," in *Auditory Display*, G. Kramer, Ed. ICAD, 1994, pp. 417–446, Addison-Wesley.
[20] T. Hermann, "Sonification for exploratory data analysis – demonstrations and sound examples," `http://www.techfak.uni-bielefeld.de/~thermann/projects/index.html`, 2002.
[21] T. Hermann, C. Nölker, and H. Ritter, "Hand postures for sonification control," in *Gesture and Sign Language in Human-Computer Interaction, Proc. Int. Gesture Workshop GW2001*, Ipke Wachsmuth and Timo Sowa, Eds. 2002, pp. 307–316, Springer.
[22] T. Hermann, J. Krause, and H. Ritter, "Real-time control of sonification models with an audio-haptic interface," in *Proc. of the Int. Conf. on Auditory Display*, R. Nakatsu and H. Kawahara, Eds. Int. Community for Auditory Display, 2002, pp. 82–86, Int. Community for Auditory Display.
[23] B. Fritzke. A growing neural gas network learns topologies. In G. Tesauro, D. Touretzky, and T. Leen, editors, *Advances in Neural Information Processing Systems*, volume 7, pages 625–632. The MIT Press, 1995.

Thomas Hermann received a Masters in Physics 1997 from Bielefeld University working on femtosecond laser pulse analysis and frequency conversion. Afterwards, he joined the Neuroinformatics Group of the Faculty of Technology, Bielefeld University, where he started the research on sonification for exploration and process-monitoring of high-dimensional data. In 1998 he became a member of the Graduate Program "Task-Oriented Communication". He received a Ph.D. with the thesis "Sonification for Exploratory Data Analysis" in June 2002 and is currently continuing his research with a focus on interactive human-computer interfaces (e.g. audio-haptic controllers) and techniques for multimodal data exploration.

Helge Ritter studied physics and mathematics at the Universities of Bayreuth, Heidelberg and Munich. After a Ph.D. in physics at Technical University of Munich in 1988 he visited the Laboratory of Computer Science at Helsinki University of Technology and the Beckman Institute for Advanced Science and Technology at the University of Illinois at Urbana-Champaign. Since 1990 he is head of the Neuroinformatics Group at the Faculty of Technology, Bielefeld University. His main interests are principles of neural computation, in particular for self-organization and learning, and their application to build intelligent systems. In 1999, Helge Ritter was awarded the SEL Alcatel Research Prize and in 2001 the Leibniz Prize of the German Research Foundation DFG.