Listen to your Data: Model-Based Sonification for Data Analysis

T. Hermann and H. Ritter
Department of Computer Science · University of Bielefeld · D-33615 Bielefeld
e-mail: {thermann,helge}@techfak.uni-bielefeld.de

Abstract

Sonification is the use of non-speech audio to convey information. We are developing tools for interactive data exploration, which make use of sonification for data presentation. In this paper, model-based sonification is presented as a concept to design auditory displays. Two designs are described: (1) particle trajectories in a "data potential" is a sonification model to reveal information about the clustering of vectorial data and (2) "data-sonograms" is a sonification for data from a classification problem to reveal information about the mixing of distinct classes.

Keywords: Sonification, Exploratory Data Analysis, Acoustics, Cluster Analysis

1 Introduction

The detection of hidden regularities in high dimensional data sets is one goal of the work in the research area of *datamining*[4]. Structures may occur as a clustering of the data, as hierarchical organization or in functional dependencies between the components of data. In high-dimensional data this organization is mostly not obvious. This has motivated the development of various visualization techniques such as multidimensional scaling[3] or projection on principal components[8] that attempt to create dimensionality reduced displays in which the "main" structure of the data is more discernable for humans. These methods are attractive, since they transform the given data into a format that allows us to invoke our highly developed capabilities for detecting even sutble visual patterns in images. However, we are also capable to detect very subtle patterns in acoustic sounds, which is exemplified to an impressive degree in the field of music, or in medicine, where the stethoscope still provides very valuable guidance to the physician. While these examples demonstrate that the use of sound for detecting subtle structures is an important praxis in several fields, it so far has found comparably little attention in the field of datamining, which may be due to the larger difficulty to communicate about sound in comparison to visualization.

This paper presents two new methods for acoustic data presentation: listening to particle dynamics in a *data potential* reveals information about the clustering of data. Listening to *data sonograms* gives an impression on results of a prior clustering analysis, e.g. on the class borders of a learned classification. Both methods provide an extension to earlier sonification approaches, which are mainly based on the following four techniques:

- **Auditory Icons:** A suitable classification process selects one of a set of sound pieces. These serve as an auditory sign[6], which must either be learned or intuitively understood. This method is often used for alarm signals and navigational cues.
- **Earcons:** here auditory signs are combined to form more complex messages, just as spoken words are combined to form sentences.
- **Audification:** here the data is directly translated to the audible domain[9], it is interpreted as a time series which directly controls the audio signal amplitude. This is particulary meaningful if the data originates from a system dynamic evolving in time, e.g. certain physical measurements.
- **Parameter Mapping:** here the data drives the parameters of a synthesizer, which may be imagined as an waveform-producing algorithm. For each data point one or more tones are generated where the parameters of the events, e.g. time stamp, duration, volume, pitch, envelope characterisics, brightness, etc., are controled by the data vector components [12]. The result can be called a multi-dimensional "sonic scatter plot"[10].

For high-dimensional data, both audification, auditory icons and earcons are very limited and parameter mapping is the richest method. Nevertheless, even parameter mapping has some problems:

Unique Mapping: there is no unique way of mapping between components and parameters. The listener therefore requires some learning time to get acquainted to a chosen mapping. The necessity of parameter assignment leads furthermore to a combinatoric explosion of possibilities with increasing dimensionality.

Limited dimensionality: The dimensionality of the sonification is limited by the number of parameters of the selected instrument.

Invariance: the sonification is not only sensitive to structures, e.g. it's not invariant to translations or rotations of the data.

Independence/Perceptual Uniformity: Some of the instrument parameters are perceptually coupled (e.g. duration and pitch), some are perceptually nonlinear.

Relationship: the sonification is just the superposition of independent events. This sonification method has no possibility to exploit the relationship between different data points, e.g. the local density of the data.

To circumvent most of these problems, we propose a different view to sonification: why not sonify data spaces by taking the environmental sound production in our real world as a model. Nature has optimized our auditory senses to extract information from the auditory signal that is produced by our physical environment. Thus the idea is: (i) build a virtual scenario from the data, (ii) define a kind of "virtual physics" that permits vibrational reaction of its elements to external excitations, (iii) let the user interactively excite the system and listen. We call this scheme "*model-based sonification*" because its basis is the imagination of a virtual data "material" for the development of the sonification. Depending on the choice of model, the sonification produces sound that may be more pleasant than audifications, if the dynamics follows analogous rules as real world sound generating systems. The model may be designed to reveal specific information, permitting a task-oriented model design. Furthermore, the model may be controled by a very limited number of parameters, making the sonification easy to use and to understand. The learning effort is as well reduced. Having knowledge of the model, an interpretation of the sound is simplified. The models may be designed to be assigned to data of different dimensionality, leading to a data presentation even for very high-dimensional data without dimensionality reduction.

This paper is structured as follows: section 2 presents the concept of model-based sonifications and discusses the benefits of virtual instruments. Section 3 presents an example model: particles in a data potential. Section 4 discusses this model and gives some sound examples. Section 5 presents the data sonogram model. The last section presents a discussion of the results.

2 Model-Based Sonification

In our world, normally passive objects are silent. So why should a data set itself produce sound? Sound occurs, if a system becomes excited. The system dynamics leads to a vibrational reaction of some parts of the system which is transmitted through the medium to our ears and we perceive that as sound. In many situations, when we examine an object (e.g. shaking a bottle, knocking on a table), we ourselves cause this excitation, and our auditory perception is optimized to use this interaction between action and feedback to draw information from this. An important aspect is, that the sound itself emerges from a process, which is defined by the underlying physics, whereas the instrument is given by its material structure.

Taking this view on sound production, the design of a sonification model consists in a "material design" in a data space. The material structure is not only determined by the setup of the elements, but also given by the interactions between the elements. A kind of "virtual physics" must be defined, that permits a vibrational process analogous as in real sounding materials. Thus the data more or less directly becomes the sounding instrument, which is examined, excited, or played by the listener. By a clever design of the virtual physics, the model can lead to sonifications that facilitate the auditory perception of important structures in the data, e.g. their clustering, the mixing of distinct classes in a classification task or the local intrinsic dimensionality of the data, etc. To define a physics on the abstract data space, theoretical acoustics can contribute. The virtual physics may be given by a set of differential equations, transfer functions can describe the signal spreading in space and time. For efficient implementation of the model, physical model sound-synthesis techniques can be applied [13]. Summarizing, the following steps have to be carried out to define a sonification model:

Setup:	define dynamic **elements**
Dynamics:	define **interactions** between elements and **initial state**
Excitation:	define **user interaction modes**
Auditory Observables:	define **observables** that contribute to the audio signal
Listener:	define **sound wave transfer** and **receiver characteristics**

A sonification model determines the structure of the "data material", e.g. its stiffness, damping, resonances, and other acoustic properties. Furthermore the modes of interacting with this new material or virtual object are determined. Like real objects, the virtual data object may be beaten, shaken, touched or squeezed. The multidimensionality of sound makes it a well usable medium to explore abstract data.

Theoretical Acoustics [11] and well established sound synthesis models like Digital Waveguide Theory [13] or Modal Synthesis [2] yield ideas for the design of virtual instruments and materials as well as their efficient implementation for sound synthesis.

3 Model I: Particle Trajectories in a Data Potential

Here, our goal is to receive information about the clustering of vectorial data. The sonification is determined by the following model: the data points are interpreted as "planets", held at fixed positions in data space. Each data point contributes to a global data potential. The potential leads to a "gravitational force" on particles which are injected into the data space to probe the system. Different from the $1/r$-law in gravitation, here the potential is chosen to approximate a harmonic potential close to the data point (which is related to harmonic oscillations and therefore pure tones), but which is localized so that it vanishes for large distances. Such a behaviour is shown by a negative gaussian bell. The width of the "planet potential" determines its "interaction length" σ and so the resolution of the data potential. To render the sonification, some dozens of particles (test masses) are injected into the data space to probe the potential. The audio signal is taken as the superposition of their kinetic energy. The probing is repeated with different interaction lengths σ, e.g. by decreasing σ exponentially.

The potential can be compared with kernel density estimation, a method to fit the multivariate probability density. In this method, the width parameter (or interaction length) must be properly assigned to avoid overfitting and oversmoothing. Indeed, the interaction length in the sonification model is a kind of blurring parameter.

The data potential for a data set with M data points is given by

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

with the data point potential or kernel function

$$\phi(r) = -\exp\left(-\frac{r^2}{2\sigma^2}\right) . \tag{2}$$

As an alternative, $\phi(r) = -1/\sqrt{(\sigma^2 + r^2)}$ could be used, which shows a similar behaviour. Both functions are shown in fig. 1.

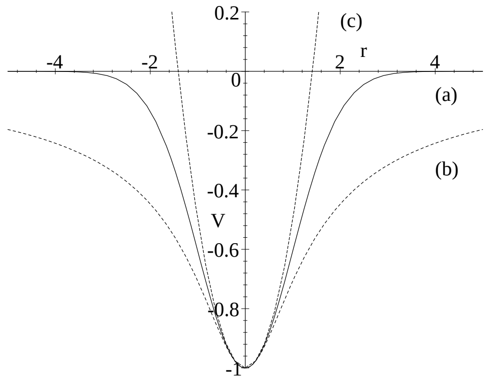

Figure 1: The data point potential: (a) negative gaussian bell ($\sigma = 1$), (b) modified coulomb potential and (c) harmonic approximation. Higher particle energies lead to oscillations in the non-harmonic regime leading to longer oscillation periods. This causes the sawtooth shape of tones in the spectrogram in fig. 3.

Now a set of particles $\alpha = 1 \ldots N$ with given mass m are injected into the data space at location $\vec{x}_\alpha$, subjected to Newton's law with a damping term

$$m\ddot{\vec{x}}(t) = -\nabla_x V(\vec{x}) - \gamma\dot{\vec{x}}(t) \tag{3}$$

where γ is a friction constant leading to exponentially decaying particle energy. [11]. The initial states of the particles are chosen randomly, each with the same total energy

$$E_0 = V(x_0) + \frac{1}{2}mv_0^2. \tag{4}$$

The observable of each particle α is its kinetic energy $\frac{1}{2}mv_\alpha^2$ and the over-all signal is taken as their sum. The systems dynamic is integrated in a straightforward manner using the Runge-Kutta method. Figure 2 illustrates a data potential and a typical particle trajectory.

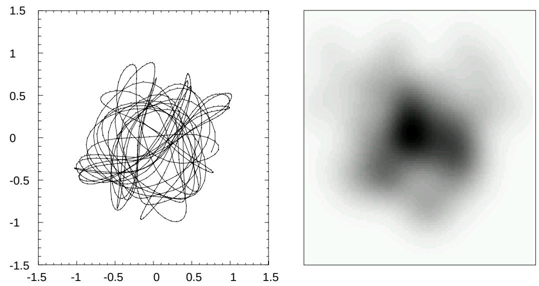

Figure 2: An example trajectory (left) of a particle in a gaussian distributed cluster in 2d and the data potential (right) for a given σ.

Now the system is controled by 5 parameters: the friction constant γ, the interaction length σ, the particles initial energy E, particle mass m and timestep Δt. Δt, m and γ are scaling factors for the time axis, whereas σ and E control the qualitative behaviour of the sonification. Decreasing σ lowers the central force, thus the particle masses are scaled with $m = k/\sigma^2$, which is appropriate to compensate that effect.

The next section will present sound examples and discuss the qualitative system behaviour.

4 Example: Listen to Three Data Clouds

As an example data set, three gaussian distributed clusters with different a priori probabilities were taken in 2d space. For each probe 50 particles with identical mass were injected and 4000 steps were calculated, corresponding to an audio signal of 0.25 sec at a sampling rate of 16000 Hz. The interaction length σ was exponentially decreased from a σ_1 greater than the largest distance between data points to σ_2, smaller than the smallest data point distance in 25 steps. Fig. 3 shows a short time fourier transform (STFT) of the resulting audio signal, using a window size of $N = 512$ with 400 samples offset between successive frames. You can listen to the result, by playing the demo file on the web page [7].

For large σ (large against the data set variances) the particles will do nearly harmonic oscillations in the data potential. All particles thus produce tones of the same pitch. The total curvature of the potential minimum is proportional to the number of data points that contribute to the potential, so that the tone is higher pitched for a larger data set. With decreasing σ more and more structures get audible. First, there is a tonal split for particles oscillating orthogonal and parallel to the main covariance direction. Next, there is a critical range of interaction length, where potential hills grow between clusters, but the particle energy is still high enough to move over this walls. The resulting trajectories are more or less chaotic and contribute a complex mixture of low pitched and noisy sounds, which are easily recognized. These sounds indicate the separation of clusters at this length scale. At some σ, the potential throughs of the three clusters separate from each other. Some particles end up with their motion in one of the clusters. This is perceived as a tonal split, which occurs if the clusters have different "mass". It can be shown that a cluster with M data points yields a tone with frequency $\omega = \sqrt{M/\sigma^2 m}$. Thus, clusters of the same size and shape cannot be acoustically distinguished. With decreasing interaction length, the tones keep their pitch until the data point potentials split again, leading to a plateau for each cluster, whose duration corresponds to the "cluster compactness". With further decreasing σ, the granular substructure of each cluster gets audible. At the end, all data point potential troughs are separated and each particle ends in a same shaped potential, which again results in tones of one pitch.

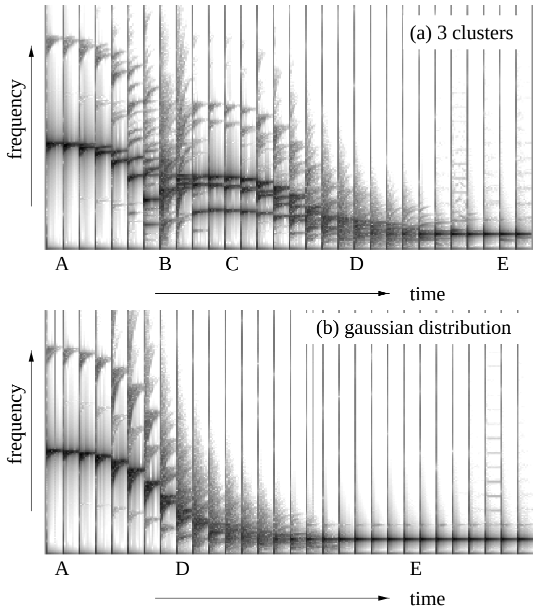

Figure 3: Spectrogram of the data potential sonification for a data set with (a) three point clouds with differing a priori probabilities: (A) one tone in overall harmonic potential, (B) chaotic trajectories between potential troughs, (C) plateau with 3 tones, one for each cluster, (D) separation of cluster member potentials, (E) one final pitch as all potentials have separated. (b) one gaussian distributed cluster: here no pitch plateau is obvious. The upper curves are just the higher harmonics of the waveform. The sawtooth shape of the tones is explained in fig. 1

This sonification is of rather short duration (7.5 sec) and communicates a lot of structural information. Surely a little training time is improving user performance. For comparison, a sonification for a single gaussian distributed cluster is shown there as well and may be listened to on the web page [7]. As there is no clustered substructure, the tones directly drop to the final pitch.

In its current form, the method is computationally quite demanding, because for each time step ∇V must be evaluated, which depends on all particles involved. Look-up-tables and particle potentials $\phi(r)$ with limited support may help to reduce computation time.

5 Model II: Data Sonograms

As a second model, *data sonograms* are presented. This sonification model was developed to listen to vectorial data of a classification problem. Furthermore it is applicable to explore the results of a prior classification process. A common question in classification problems is the separability of distinct classes, which is a prerequisite for the selection of a classifier (e.g. a perceptron is well-suited for linear separable classes), and which should become audible with this sonification.

The sonification is determined by the following model: each data point is fixed with a spring to its position in the n dimensional data space. Thus each point can perform vibrational movements around its position which is determined by an external excitation, the mass of the point, the stiffness of the spring and the specified dissipation rate.

In our model we give all points the same mass, but we let the stiffness of the spring be controled by the local environment of the point. Specifically, we choose it proportional to the local class entropy S. Seeking the k nearest neighbours and calculating a class member histogram, the local class mixing entropy is given by

$$S \propto -\sum_{\alpha=1}^{N} p_\alpha \ln p_\alpha, \tag{5}$$

where N is the number of distinct classes, and p_α is the probability density for a member of class α at this point. The probabilities can be estimated by the number of class members in class α among the kth nearest neighbours. The class entropy is low for data points surounded by members of the same class and high in regions where the members of different classes mix. The friction force is given by the local data density, which may be estimated by the reciprocal size of the sphere, in which the k nearest neighbours were found. To excite the data, a shock wave is emanated from a certain position, e.g. the centroid of all members of a selected class, which spherically expands through the data space, as shown in fig. 4. The data points are excited to vibrations around their position as the shock wave arrives. The listener is assumed to be positioned at the shock wave center and the sonification is just the superposition of all sounding data points.

Again, the model is controled by a small number of parameters: the number k of neighbours, the mass m, a constant relating the spring stiffness to the entropy S and the shock wave speed which may be adjusted. Each parameter is easily understood and leads to expected changes of the sonification.

As an example, the "iris data"[1] set, which consists of 3 clusters in a 4 dimensional space, 2 of them having an overlap, is sonified. Here we only monitored the tones from members of the selected class, whose centroid is also the starting point for the shock wave. The sonification gives an impression about the class borders:

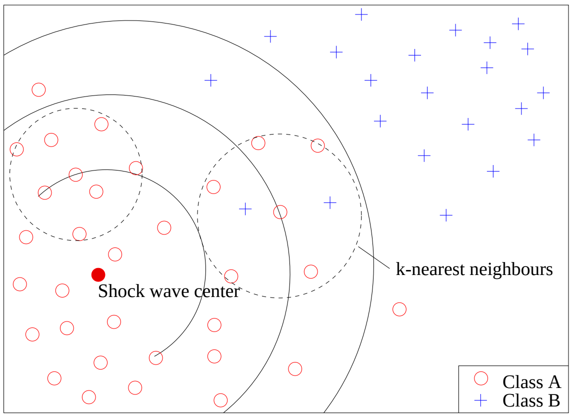

Figure 4: Illustration of an emanating shock wave. All data points are excited to vibrations around their position as the shock wavefront reaches them. The spring stiffness (and thus frequency) is determined by the local class entropy, the dissipation rate is given by the reciprocal volume of the k-nearest-neighbour-sphere. To examine class borders between two classes, the center of mass of the first selected class is taken as the shock wave center and only the tones for the first class are sonified.

the separation of class 1 and 2 is audible as all points have the same basic pitch of minimal entropy. The overlap between class 2 and 3 is audible: the higher pitched tones come from data points which are located in the class border. Sound demos may be found on the webpage [7]. Fig. 4 shows a short time fourier transform of the data sonogram.

A drawback is the missing spatial resolution. All signals from the shock wave sphere are perceived at the same time. Here a useful extension of the model could be to position the listener in the plane spanned by the first two principal components of the data set and synthesize stereo audio signals. With this extension it might be further possible to code positional information in phase differences between the left and right channel.

6 Discussion

Sonification addresses a so-far largely underused perceptual channel for man-machine-interaction. Data can thus be experienced in a new way, which bears the advantage of a deeper and possibly richer understanding of data structures. As sonification is a multidimensional presentation, it may prove especially useful for multivariate data analysis. The concept of Model-Based Sonification presented in this paper contributes an approach to design data sonifications that both permit intuitive interactions with the data and allow generalizations to ar-

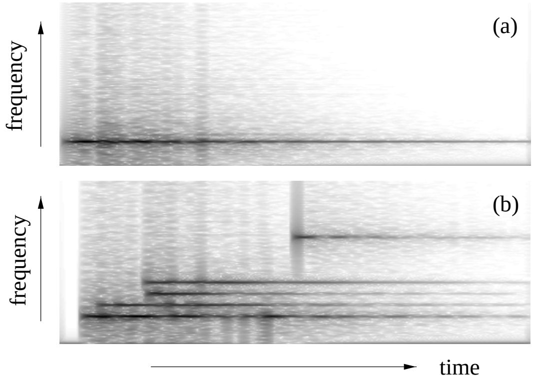

Figure 5: Data Sonograms of iris data set. The shock wave starts at centroid of class 2, (a) S only includes members of class 2 and 1, (b) S only includes members of class 2 and 3. Here the class overlap is audible as higher pitched tones.

bitrary data dimensionality. The model may further be specialized to highlight certain aspects of the data and therefore can form a good basis for an "acoustic toolbox" for data exploration.

Sonification may be a good aid for rapid screening of data since an auditory stream can be "consumed" with comparable little effort. We think that an even higher potential lies in its use as a new method for interacting with data, particularly when used in conjunction with visualization. However, it is rather easy to refer to details in visual displays simply by pointing, which makes communication with others easy and fast. This is different for auditory displays. Here, there is no straightforward equivalent to pointing. Modern computer techniques may offer new ways to overcome this limitation, which would enhance the usefulness of sonification further.

Up to now, research on sonification for data analysis is just beginning and further experience has to be gained. It should be mentioned that training is required to distinguish sonifications and to interprete them correctly. A good example is given by the car mechanic who often is highly proficient in diagnosing many causes of malfunction acoustically due to his long training. Likewise, in familiar domains, such as the perception of language, we all are well trained "experts" and can discover subtle features such as small but important differences in prosody. Therefore, in addition to the development of sonification models that convey useful information, empirical research about the learnability, reliability and usability will become an important requirement for a more quantitative assessment of the performance of different sonification strategies.

Like every method, sonification will not be the best choice for all problems, but already has shown to be superior than other methods for some questions and some kind of data[5]. Exploratory Data Analysis will gain the most profit, if the best suited methods are combined and used synergetically. We hope to address this point in more depth in future work focussing on the development of useful sonification models and their integration for multimodal data exploration.

References

[1] UCI repository of maschine learning databases. ftp://ftp.ics.uci.edu/pub/machine-learning-databases/iris/.

[2] J. M. Adrien. The missing link: modal synthesis. In *Representations of Musical Signals*, A. Piccialli et al.(eds), MIT Press, 1991.

[3] D. R. Cox and M. A. A. Cox. *Multidimensional Scaling*. Chapman & Hall, London, 1994.

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

[5] T. Fitch and G. Kramer. Sonifying the body electric: Superiority of an auditory over a visual display in a complex multivariate system. In *Auditory Display*, G. Kramer (ed.), Addison-Wesley, 1994.

[6] W. W. Gaver. Using and creating auditory icons. In *Auditory Display*, G. Kramer, (ed.), Addison-Wesley, 1994.

[7] T. Hermann. Model-based sonification - demonstrations. http://www.techfak.uni-bielefeld.de/∼thermann/projects/index.html, 1999.

[8] Jolliffe. *Principal Component Analysis*. Springer, New York, 1986.

[9] G. Kramer, (ed.) *Auditory Display - Sonification, Audification, and Auditory Interfaces*. Addison-Wesley, 1994.

[10] T. M. Madhyastha and D. A. Reed. A framework for sonification design. In *Auditory Display*, G. Kramer (ed.), Addison-Wesley, 1994.

[11] P. M. Morse and K. U. Ingard. *Theoretical Acoustics*. McGraw-Hill, New York, 1968.

[12] C. Scaletti. Sound synthesis algorithms for auditory data representations. In *Auditory Display*, G. Kramer (ed.), Addison-Wesley, 1994.

[13] J. O. Smith. Physical modeling using digital waveguides. *Computer Music Journal, Vol. 16, no. 4, pp. 74-91*, 1992.