MULTI-TOUCH INTERACTIONS FOR MODEL-BASED SONIFICATION

René Tünnermann, Thomas Hermann

Ambient Intelligence Group
Cognitive Interaction Technology - Center of Excellence (CITEC)
Bielefeld, Germany
{rtuenner|thermann}@techfak.uni-bielefeld.de

ABSTRACT
This paper presents novel interaction modes for Model-Based Sonification (MBS) via a multi-touch interface. We first lay out details about the constructed multi-touch surface. This is followed by a description of the Data Sonogram Sonification Model and how it is implemented using the system. Modifications from the original sonification model such as the limited space scans are described and discussed with sonification examples. Videos showing examples of interaction are provided for various data sets. Beyond Data Sonograms, the presented system provides a basis for the implementation of known and novel sonification models. We discuss the available interaction modes with multi-touch surfaces and how these interactions can be profitably used to control spatial and non-spatial sonification models.

1. INTRODUCTION

Exploratory Data Analysis aims to develop techniques for users to better grasp the hidden structure in complex data. If we take this statement literally, we might not only ask how we could implement techniques to manually interact and get our hands on data, but also, how it sounds or should sound if we interact physically with data. Real-world acoustic responses that we experience when touching (hitting, scratching, ...) an object or surface are often very useful and reveal a whole range of information about the object’s properties (material, stiffness, surface properties, etc.). We often underestimate the utility of such direct feedback since it is omnipresent and at the same time effortlessly integrated into our multi-modal perceptions.

The arising questions are how can we inherit the benefits of action-perception loops for a better understanding of complex data and how can we structure surface-based interfaces so that users can obtain an informative acoustic reaction on arbitrary interactions? Model-Based Sonification takes these aspects of interaction particularly into account [10]. Sonification models according to MBS can be excited by the user. For this excitatory process many different interaction interfaces beyond the mouse, such as the audio-haptic ball interface, or the malleable user interface have been presented [11, 12]. These are primarily input interfaces and the sonification in many implementations has been the only output modality in this interaction loop.

In this paper, we investigate the above research questions by using multi-touch interfaces. We start by presenting a device developed within the Ambient Intelligence Group that combines the possibilities of Multi-Touch Interactions and Tangible Interactions in a desk-based system for simultaneous multi-user use. Our multi-touch system allows to create tightly coupled audiovisual interaction loops to represent the temporal evolution of sonification models while at the same time allowing real-time complex manual interaction with a sonification model. The system has been developed to serve as a sound basis to fuse and explore the potential of multi-touch interactions together with tangible interactions, while using truly multimodal output media.

In Sec. 4 we provide categories for flat surface-based interaction and then use these to discuss how interactions can be connected to the excitation for Model-Based Sonifications. We demonstrate the system together with a specific sonification model, the Data Sonogram Sonification Model, where the user can use multi-touch interactions to set centers of excitation waves that spherically pass through data space. In comparison to the mouse-based interaction used previously, the multi-touch interaction provokes new interaction styles such as rapid A/B-comparison and simultaneous excitations in different regions. Furthermore the real-time visualization supports a better cross-modal binding.

Beyond the demonstration of new interaction modes for holistic data experiences as exemplified with the use of our system for the interaction with Model-Based Sonifications, we see diverse application fields where sonification can be plugged-in to enhance the experience. For instance, in didactic multi-touch applications such as an interactive visualization of electromagnetic fields, sonification may represent the field strength as sound while the user moves electric charges or touches the surface. In the area of interactive games, sonification could enable games between sighted and visually impaired users where each receives the modalities she could use best.

2. MODEL-BASED SONIFICATION (MBS)

Model-Based Sonification is a framework for the development of sonification techniques [9]. MBS starts from the observation that humans are well trained to interpret the complex acoustic signals in the world with respect to sound source characteristics. To give an example in everyday interaction, imagine to fill a thermos flask with water. By the pitch rise, due to the changing resonance of the bottle, we are aware of the flask’s fill level. There is a large variety of situations when we use sound to gain insight into complex systems (e.g. engineers listening to machine sounds or physicians using the stethoscope to support diagnosis) [16, 17].

Most important, sound is connected to the underlying physical system by means of a dynamic (physical) model. The model mediates between a system’s state and its acoustic response. The human brain is trained to infer source properties from sound that results from physical models. This principle provides the basis for Model-Based Sonification which defines in analogy dynamic processes between elements that are typically parameterized by the data. As in physical systems, a sonification model is silent without any excitation. Via interaction, the user is able to excite the model which connects MBS to the field of interactive sonification [13]. Guidelines and examples for creating interactive sonification models are provided in [8, 9, 10].

For the definition of a sonification model according to MBS, six aspects need to be addressed: At first, the setup of dynamical
elements and the initial state of the model have to be determined. This is followed by the dynamics, which define how dynamic elements behave in time. The excitation defines how the user is able to interact with the model. The Link Variables used as transducers between the model and the audible domain have to be chosen and designed before the listener can be positioned in the setting. Finally, depending on the model, the listener needs to be positioned with respect to the data.

2.1. Excitation Modes for Sonification Models

The above definition for MBS has already shown that excitation plays a crucial role in the use of sonification models. Designers can take inspiration from all real-world interactions where the world responds acoustically, from 'foot steps' to 'hitting of objects'. If we focus on manual interaction we see that in most situations we either hit, pluck, shake, deform/squeeze or scrub objects. Most of these interactions have several degrees of freedom, e.g. the location of the interaction, the strength, the detailed direction relative to the surface. Depending on the details, interaction with real-world objects provides sonic feedback which includes information about the object. Obviously this richness is far beyond what can be obtained by simple mouse or keyboard interfaces. For those reasons, new interaction devices have been developed to better explore our manual interaction abilities [11, 12].

If we consider the interaction with surfaces in general (e.g. consider to search a wall for hidden holes) we often use tapping, scratching (to examine the surface) and (think of drumheads) bimanual interactions where one hand hits while the other changes an aspect of the surface. Similarly, interactions are natural for interacting with surfaces, and with interactive multi-touch systems we now have the chance to define audiovisual surface reactions, so that an almost as natural utilization of manual interaction procedures may occur by users that explore complex data.

Our long range aim is to implement examples for all available surface-based interactions to explore the potential of MBS to connect manual interactions with exploratory excitations to support the understanding of data under analysis. In this paper we start this investigation with tapping interactions as excitations for the Data Sonogram Model in Sec. 5.

3. MULTI-TOUCH TECHNOLOGY FOR THE TANGIBLE DESK (TDESK)

As a basis for the development, we started off using the tangible desk (tDesk) [1], a tabletop environment for tangible interaction (see Fig. 2). The tDesk is assembled using aluminium strut profiles. It consists of a 70 cm × 70 cm glass surface resting on four aluminum poles in about 80 cm height (see Fig. 2). The chosen extent of this table allows to conveniently work either alone or collaboratively within a group on and with touchable and tangible applications. Any spot on the surface can be reached with ease regardless of the user’s deskside position. Since modularity was a major design issue of the tDesk, the current glass surface is easily exchangeable. We designed a drop-in replacement surface, enabling the tDesk to recognize fingers touching the surface.

The used setup consists of the tDesk platform, the constructed acrylic surface with attached aluminium frame, lighting modules covering the pane edges, projector, camera, speakers and a computer system for image processing and multi-modal feedback (see Fig. 1). Basically, the constructed surface is a spatially resolved 2D-sensor recognizing multiple touch inputs. The physical sensor pane is made out of acrylic glass. The display is provided using a screen foil and an inside mounted projector.

The designed surface allows simultaneous interactions by at least four people in a closed-loop and direct manner. When designing the surface the following aspects were considered:

- **Low Latency**
  Closed-loop interactions require a low latency and therefore a high sampling rate. Reducing latency is of utmost importance, as rising latency can disturb closed-loop human-computer interaction. There has to be immediate feedback while the user interacts with the surface, otherwise the feedback can not be associated with the previous action.

- **Input Points**
  The sensor should support at least 40 input points so that up to four users can use all fingers simultaneously.

*ICAD09-2*
Proceedings of the 15th International Conference on Auditory Display, Copenhagen, Denmark May 18 - 22, 2009

4. MULTI-TOUCH INTERACTIONS FOR MBS EXCITATION

Since the appearance of multi-point displays, more advanced displays allowing absolute-spatial, multi-point and multi-user interaction by the use of one’s fingers are announced frequently, but the characteristics of touch (applied force, angle of approach) are mostly ignored. Whereas orientation and force are commonly exploited in graphics tablets such as Wacom’s devices, these characteristics are not used in multi-touch pointing surfaces. The touch-input of a finger is mostly limited to binary pointing, yet. In general surface input interfaces can be considered as transducers between the digital and the analog interaction space. The following categories provided us a useful scope to better differentiate and discuss characteristics of surface-based interaction, particularly as a source to excite sonification models from MBS. Some categories are influenced by a loose collection of ideas by Buxton, listed on his website [2].

Point vs. Touch: Existing ‘multi-touch’ displays often offer multi-point instead of multi-touch input. The touch of a finger’s tip is only used for a mere pointing, neglecting the details of touch. In addition, touching the surface with the hand or the arm will often lead to undefined behavior.

Single- vs. Multi-Spot: Old fashioned touchpads, which are still quite common, support only single-point input. Whereas single connotes just one spot input, multi-spot refers to devices capable of sensing more than one spot, for example all of the user’s fingers. With single- and multi-spot as two sides of the continuum, in between there are n-spot devices capable of sensing a fixed number n of spots.

Collaborative Use: Even though newer notebook computers offer multiple input surface devices, these can hardly be used by more than one person at a time. Even if those pads theoretically could be used by more than one person at a time, in most cases this will lead to odd experiences.

Degrees of Freedom: When using spots on the surface only as pointing input, the surface provides input with two degrees of freedom. The transducer gains degrees of freedom by adding information about the pressure-vector of touch and direction of approach or other information.

Feedback: Traditional touchpads give no active feedback at all. Touchscreens and multi-spot displays feature visual feedback and thereby create the illusion of being able to manipulate digital items directly. Visual feedback can be enhanced by sound to intensify the degree of immersion. Digital objects with an auditory behavior can create sounds when triggered or when several objects interact with each other. Additionally, vibration motors could be used to create haptic feedback.

Relative vs. Absolute: Touchpads are, like mice, relative input devices: If touched, they take the cursors position. The position in the real world is set equal to the position of the cursor in the screen space. Touchscreens on the other hand feature absolute input. The user does not have to move the mapping (see Fig. 3). Based on the experience with different setups, a RMSE of about 5.5 pixels can be expected for a calibrated system. In comparison, a usual fingertip covers an area with a radius of about 20 pixels. Additionally users induce a mean error of about 1.9 pixel when asked to repeatedly touch a spot on the surface. Considering this, the remaining error is sufficiently small for closed-loop interactions on the surface.

Figure 3: Corresponding test data point pairs in camera (+), display (*) coordinates and the resulting coordinates (△) transformed by the LMA optimized mapping (setup A). Note: display coordinates are of a higher extent ($[0..1000]^2$) than camera coordinates $[0..640] \times [0..480]$.

- High Resolution
  The sensor’s resolution should be close to the display’s resolution to support detection of the characteristics of touch, such as the shape or the orientation of the touched spot on the surface.

- Backprojection
  To intensify the degree of immersion, the projection should have its source inside the table. Top projection would lead to disturbing occlusions caused by the user’s hands and fingers.

We used the FTIR technique (as proposed by Han [5]) to sense contacts on the surface. By using this technique we were able to sense almost as many input contacts as fit on the surface, achieving a high resolution and sufficient latency. To apply FTIR sensing, an acrylic pane is flooded with near-infra-red (NIR) light. When a finger touches the surface a bright NIR blob is caused on the surface. A camera, mounted underneath the surface, is used to capture the resulting blobs. To reduce the latency in the loop, a firewire camera capturing images at a frequency of 60 Hz is used. To improve the image quality, an optical bandpass filter was mounted in front of the camera. Finally, the display is provided by an underneath mounted projector.

To calibrate the camera and projector coordinate system we have chosen a mapping to resolve:

- camera trapezoid and pincushion distortion
- translation, rotation and scaling
- projector trapezoid and pincushion distortion

Optimal parameter values of the transformation map are determined by minimizing a quadratic error function using a least-squares minimization. Matching point-pairs from both coordinate systems are needed to compute the residuals for each iteration of the optimization process. The Levenberg-Marquardt Algorithm (LMA) [14, 15] was used to train the mapping’s parameters. To evaluate the learned mapping, a test data set is recorded. The test data set contains 16 random point pairs. These data points are used to determine the remaining root mean square error (RMSE) of the learned
cursor from its current position to the target, but approaches the object directly by using a finger or a tool.

**Direct vs. Abstract**: When an object is moved with a finger or a pen-like tool, the interaction with it can be in a direct manner. If there is a relative transducer in the input chain, for example a mouse, the interaction becomes more abstract. There are a lot of discussions about when interactions are to be seen as abstract or not. We do not doubt that for someone who is familiar with mouse interaction, the relative transducer is ubiquitous and therefore virtually ready-to-hand (see dimension Tools and [7]). In this work, the term direct is used if the input chain of an interaction is free of relative transducers and the application allows the user to touch or move digital items.

**Discrete vs. Continuous**: Discrete interactions can be seen as single actions or events. A continuous interaction can be described by a trajectory of actions or events. Imagine typing on an on-screen-keyboard, or pressing displayed buttons, the interaction would be discrete. By moving an object from one position to another, the interaction becomes continuous.

**Tools**: Surface input devices can be designed to be used with different parts of the body such as a finger or with external tools. A tool can be ready-to-hand or present-at-hand to the user. A pen for example, when used for sketching or drawing tasks, is ready-to-hand. The user does not have to think about how to handle the pen, he just spends time on the drawing task itself [7], [3].

**Setup**: Data points are used as point masses in a model space of the same dimension as the data space. The Data points’ coordinates are used as fix point location for a virtual spring connected to the point mass.

**Dynamics**: (a) oscillation of spring-mass systems modeled by classical mechanics, and (b) the propagation of shock waves in the high-dimensional model space.

**Excitation**: The user can excite multiple ‘shock waves’ to emanate at certain locations in the model space.

**Link Variables**: The kinetic energy of all point masses is used to generate the sound signal, which represents the sonification. Alternatively the elongation from equilibrium can be used as link variable.

**Listener**: A two-dimensional view of the data and visual controls to navigate the data exploration are provided to the listener on the interactive surface. In the original model, the virtual listener  is positioned in model space at the point where the shock wave is initiated. In this implementation, however the listener is centered at the x-axis of the 2D plot in front of the table, since only a stereo panning is used for sound spatialization.

A two-dimensional scatter plot of the data serves as the interaction area for the user to excite data sonograms by selecting a position on the plot. The speed of the shock wave can be adjusted by the user interactively. The resulting shock wave then passes through the data set within some seconds, so that the user can examine the data step-by-step.

**5. THE DATA SONOGRAM SONIFICATION MODEL**

**5.1. Overview**

The *Data Sonogram Sonification Model* can be described by the following five categories:

---

**Figure 5**: The graphical user interface of the Data Sonogram application. The 13D wine data set is selected. The dimensions *Ash* and *Magnesium* are chosen for the visual feedback. Dimensions are chosen by a tap on a button in the button group for the corresponding axis. In a 13-dimensional data set, 13 buttons are displayed along each axis. Two shock waves are active, propagating through the two-dimensional data space.

**Figure 4**: Multi-Touch Interactions as melting pot for multi-modal coherent binding. The interface causes visual, tactile and auditory perceptions which are bound together via their synchronization to multi-modal perceptual entities.

According to the above definition we have implemented a direct and absolute-spatial, multi-point version of a data sonogram, which allows discrete and continuous use in a collaborative manner. We plan to exploit attributes such as force sensitive input, the use of tools and touch characteristics since the constructed surface already provides these informations.

**5.2. Multi-touch adaptations for Data Sonograms**

We added several features for the multi-touch Data Sonogram implementation. At the point of excitation in a sonogram, a virtual
shock wave is initiated (see Fig. 7). For the sake of usability and in contrast to physical constraints, the speed of the propagating wave can be adapted while the wave is traveling. It is important to mention that such crucial changes are only applied on the user’s request. Unnoticed changes in the synthesis would lead to false judgments. A shock wave has a center and a source point. The distinction has to be made because the point of excitation is not always the center of the shock wave. In case the user initiates a high-dimensional shock wave, the shock wave center will be located at the coordinates of the data point that is nearest to the excitation point in the two-dimensional display.

In our original implementation [9], no visual update occurred during the excitation of the data sonogram. Here we have added an interactive visualization of the shock wave front while the sonification is computed in real-time. Thereby the user knows where the sound comes from at the time they are perceived. A visual shock wave front is only meaningful in the case of a two-dimensional shock wave expanding on the table. In case that an arbitrary scatter plot of the high-dimensional data is shown, data points might be reached by the shock wave in a less predictable pattern. For this case we update the data points’ visual blob by increasing their intensity at the time where the high-dimensional shock wave passes through them. Again, these synchronized visual and auditory events help the user to better interpret the sound with respect to the data.

With the increased directness that multi-touch interactions serve, we discovered that users were interested to ‘probe’ the data sets frequently at different locations. The continuation of the shock wave after it has been triggered is then irritating and thus not helpful for this interaction pattern. We therefore implemented a new interaction style, where a shock wave immediately stops at the time when the user lifts the finger. In result, users can tap on the scatter plot of the high-dimensional data is shown, data points might be reached by the shock wave front. This behavior turned out to be a useful optional feature for a step by step exploration of the data space.

The interface now offers several options for adjustment of parameters, navigation of data in a coherent surface (see Fig. 4). Parameters such as the wave velocity, the sonic dimensionality and other features of the sonification model can be adjusted by the user interactively. The user is able to navigate in the data and can choose the to be displayed dimensions.

5.3. Auditory Components for Data Sonograms

The sonification sound signal is the superposition of all instantaneous elongations of masses from their equilibrium position. Since the spring forces the masses to a damped oscillation their representation becomes audible as a decaying sine tone. For the implementation of the spring-mass systems, unit generators for spring-mass-systems, since they are well decoupled.

A stereo speaker setup is aligned towards the listener (as shown in Fig. 1). The virtual listener is centered in front of the table. When the shock wave front passes a data point, a sound event is spawned via the OSC protocol. Since the class label is used as spring stiffness, it can be perceived as pitch. The spatial location of the sound source can be estimated via the stereo panning.

5.4. Visual Components for Data Sonograms

A two-dimensional plot of two user-selectable data dimensions (see Fig. 5) is presented on the interactive surface. Each dimension can be selected by a single touch on a corresponding button.

5.5. Example Data Sets

There are three data sets available to the user by default:

- Wine data set: These data are the results of a chemical analysis of wines. They are derived from three different cultivars in the same region in Italy. The analysis measured the quantities of 13 constituents found in each of the three types of wines [4].
- Iris data set: This is perhaps the best known data set to be found in pattern recognition literature. It contains three classes of iris flowers, with four features each. 50 instances are included for each of the three classes. One class is linearly separable from the other two. The latter are not linearly separable from each other [6].
- Random data: The random data set contains uniformly distributed random data in four dimensions. It serves here as a benchmark distribution to train audiovisual exploration, i.e. to better learn to associate auditory and visually perceived elements in the multi-modal system.
Every time the data set is changed by the user’s demand, a short description of the data set is displayed. This message shows the origin and history of the data set, its dimensionality and cardinality.

5.6. Interaction Examples

To discuss the approach, we provide videos of a user interacting with the application on our website. In the first video the user demonstrates different functions and explores the data space. The user chooses displayed dimensions and triggers shock waves in the high- and two-dimensional space.

In the second video you can see the iris data set and hear how the data sonogram depends on the location in a sequence of sound events, starting with one of two pitch levels, depending on where the shock wave is initiated. Thereby the regions of the three classes in the iris data set can be well discerned. Furthermore overlapping classes and class boundaries can be perceived.

6. DISCUSSION AND CONCLUSION

In this paper we have presented a multi-touch interface for the excitation of sonification models. In detail we have reimplemented the Data Sonogram sonification model and demonstrated multi-touch multi-user exploration of scientific data via surface-based interaction.

The main advantage of our approach is that a very natural contact between the user and the surface (as the physical representation of the data) can be established. Interaction modes with typical real-world surfaces such as tapping, hitting, pushing and scratching provide examples of how interactions can be profitably used in the context of sonification models. With the data sonogram model we have given a first example that shows how spatially-resolved tapping on the surface can be utilized as a tapping into data spaces, using quasi-physical dynamic processes in the space of the sonification model to associate meaningful acoustic responses which then represent the data to the user. In result a qualitative experience is created from the ongoing continuous interaction.

An important aspect is that the interface connects the auditory and visual representation and binds them via the surface to multi-modal data perceptualization units. Synchronization is a key component for the user to be able to connect visual and auditory elements. Since the interaction occurs in the same frame of reference, and tactile sensations complement the experience, a tightly closed interaction loop is created.

In our future work we will particularly focus on sonification models that allow to explore yet untouched aspects of continuous interaction with data distributions. Instead of providing a trigger only, we want to enable users to continuously deform data representations in order to perceive the resulting tension by these deformations as informative sound. Sonification can be used in various multi-touch applications. For instance, for didactic applications, the real-time sonification of variables (e.g. stress, magnetic field strength, etc) while interacting with a simulation of a system can deliver complementary information to what is visible on the surface. Also, auditory games where the goal is to competitively or jointly shape sounds via physical interaction with the surface offer a great potential to explore tactile computing in a yet unseen way.

In summary, the presented multi-touch sensitive surface enriches the available modes to interact with complex data and to perceive structure-related features as sound via Model-Based Sonification. The tight coupling of visualization, sonification, tangible interfaces and continuous interaction in one interface contributes to a truly multi-modal experience and shows the potential of an increased level of understanding of structures in the data. The scope of our ongoing research is to explore and quantify the possibilities in this direction.

7. REFERENCES


http://sonification.de/publications/TuennermannHermann2009-MTI/


