

ON THE USE OF SOUND FOR REPRESENTING GEOMETRICAL INFORMATION OF VIRTUAL OBJECTS

Simon Shelley¹, Miguel Alonso¹, Dik Hermes¹

¹ Eindhoven University of Technology
Human Technology Interaction
P.O. Box 513, NL 5600 MB, Eindhoven
The Netherlands
s.b.shelley@tue.nl

Armin Kohlrausch^{1,2}

² Philips Research Europe – Eindhoven
Digital Signal Processing Group
High Tech Campus 36, NL 5656 AE, Eindhoven
The Netherlands
armin.kohlrausch@philips.com

ABSTRACT

This study is concerned with the use of sound in a multimodal interface that is currently being developed as an aid for product design. By using this interface, the designer is able to physically interact with a virtual object. The requirements of the interface include the interactive sonification of geometrical data, relating to the virtual object, which are otherwise practically undetectable. We propose a classification scheme of the sound synthesis methods relevant to this application. These methods are presented in terms of the level of abstraction between the virtual object and the sound produced as a result of the user's interaction. Finally, we present an example that demonstrates the advantages of sonification for this application.

1. INTRODUCTION

For several years now, computer aided design tools have helped product designers in representing new concepts and ideas. Nowadays, however, these tools are still highly technical and lacking in intuitive user interfaces. Moreover, in most cases vision is the only sense that design tools use to convey information and few attempts have been made to incorporate interactive sound. In the process of shaping objects, designers often feel the need to "touch & feel" the surfaces of their products, and so to check and evaluate their aesthetic quality. In fact, some important details may be very difficult to see or appreciate, and using only a computer screen can be misleading. The ability to touch and feel the objects they produce during the design process allows designers to better evaluate their shape in terms of their geometric properties. The research described in this article is being undertaken as part of the SATIN project, which consists of an augmented reality environment where the user can interact with virtual 3D objects. The user will be able to see the object and to both explore and modify the shape of the object through the use of touch. An important and innovative element of this project lies in the use of sound as a means to convey information and feedback about the virtual object and the user interaction. More specifically, the use of sound allows the designer to explore geometric properties of the object that are not detectable by touch or sight.

This paper presents an overview of the theoretical foundations behind the sound design in the context of the above mentioned project. We present a panorama of the sound synthesis methods relevant to the system described above and we select those that

we consider most adequate for the required task. The classification of these methods is based on the level of abstraction of the sound in relation to the geometrical object at hand. This research takes place in the tradition of ecological psychology [1] of hearing [2, 3], in the sense that two points of departure of this approach will be used as the basis for theoretically motivating the choice of the experimental design. First, in sensing the environment an observer uses multiple sources of coherent information, also if this information comes from more than one sense. Second, if the information comes in through a large part of the sensory array, the observer uses all of this information.

The article is organized as follows. In section 2 an overview of the sound synthesis techniques relevant to this project is presented and the classification scheme is proposed. Based on this classification and considering the different types of user interaction, the selection of the most suitable sound synthesis methods is discussed in section 3. Then, in section 4 an example is presented that illustrates the advantages of sonification for this project. Finally, in section 5 a summary of this work is presented and future work is discussed with an emphasis on the need for perceptual evaluation.

2. LEVELS OF ABSTRACTION

One of the key issues for the SATIN project is the level of abstraction used in the synthesis of sounds. For the purpose of this application, at the lowest level of abstraction we present exact physical models. This is the level on which the sounds are produced in such a way that they cannot be distinguished from sounds produced when a real object is touched. On the other side, we consider the earcons, which are relatively simple sounds, often consisting of rhythmic and melodic sequences of pure tones or noise bursts. These earcons represent completely abstract mappings between events and the generated sound, but no obvious relation exists, besides that a certain action causes the reproduction of a specific sound.

With these two different levels at the extremes, we propose the classification scheme illustrated by Figure 1 which presents the ordering of different implementations of sound generation along a continuum from the exact physical modelling to arbitrary sound mappings. For the sake of clarity, this continuum has been divided into four overlapping categories named *physical models*, *perceptually intuitive sounds*, *symbolic sounds* and *earcons*. The following two sections provide an explanation of these four categories.

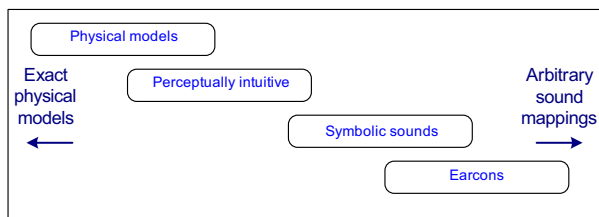


Figure 1: Levels of sound abstraction.

2.1. Physical models and perceptually intuitive sounds

Perhaps one of the most evident ways to generate the auditory properties of a virtual object is by using a model of that object to simulate its fundamental physical behaviour according to the way it is touched, or excited, by the user. We present these methods at the lowest level of abstraction, and we refer to them as *physical models*. Making a physical model is a commonly used technique with a large number of existing approaches [4]. Such models are based on the physical laws that govern the actions and reactions of the system, i.e. the virtual object, and express them using mathematical formulae. Applications of these approaches are presented by Van den Doel et al. [5, 6], and some of the Sounding Objects models [7]. In principle, if these models accurately describe the mechanical behaviour of the object, the generated sound should be identical to the sound produced by a real object. The big advantage of such models is, of course, the naturalness of the sounds produced. There are various disadvantages, however. The exact models are mostly very demanding from a computational point of view, so that only for very simple systems they will operate in real time and lead to perceptually acceptable sounds. More importantly, these physical models can only produce sounds that are in accordance with the underlying physical laws and, therefore, only information that is naturally mapped to the sound, e.g. material, roughness, can be conveyed to the user.

The next level of abstraction describes what we call *perceptually intuitive sounds*. These are sounds that the users can associate with sounds produced by touching and moving over virtual objects. Gaver [2] gave some very simple acoustic models to synthesize “everyday sounds”. In the European project “the Sounding Object” a number of synthetic sounds have been designed for use in animations and virtual reality situations [7]. A selection of demonstrations can be viewed and listened to at the web address: <http://www.soundobject.org>. Most of the synthesis algorithms are based on simplified physical models of the objects involved, and include pouring water, rolling wheels, foot steps, etc. In fact, these models are so simple that certainly the informed listener can in most cases easily distinguish the synthesized sounds from the real sounds. Although this approach does not aim at producing physically exact sounds, it has been shown that the generated sounds can convey important environmental information to improve perception and performance in virtual reality [8, 9]. In other words, when interacting with virtual objects these sounds help to experience their presence as realistic. These perceptually intuitive sounds may have cartoon-like characteristics: as the over simplified images in cartoons are perfectly suited to convey visual information, such simplified sounds appear to contain information on the acoustic environment and the objects placed in it. The models used for generating such sounds can be very simple, however, without necessarily becoming cartoon-like. For instance, wind can

be well simulated by passing white noise through a second-order resonating low-pass filter of which the resonance frequency randomly fluctuates in an appropriate frequency range [10]. In a similar way, the surf at the seashore can be modelled by passing white noise through a lowpass filter of which the cut-off frequency is slowly fluctuating [10].

2.2. Symbolic sounds and earcons

In many applications sound is used in a more symbolic way. In this case, such sounds are often part of an auditory display and we refer to them as *symbolic sounds*. In the sound abstraction scale presented in Figure 1 these sounds are located in the third position. For the symbolic sounds, the mapping between an object property or parameter and the generated sound is completely arbitrary. This sound mapping operation is usually referred to as *sonification* [11, 12]. In the broadest sense, sonification can be seen as a way of representing data using acoustic signals. In a sonification application, one thus needs first of all a mapping, i.e. a mathematical relation, between a parameter value to be sonified and some specific properties of sound, e.g. frequency or bandwidth. An important aspect of sonification applications, as stressed by Hermann [11], is the intention for generating the sound. The goal is to learn something about the data that are being represented by sound by listening to it. In general, when there is no mechanical or acoustic model that underlies the sonification process and the sound parameters, we speak of symbolic sounds. A specific property of such symbolic mapping is that it gives quantitative information about the sonified process. These symbolic sounds, however, do not have a direct relation with the actions of the users and the physical properties of the objects they are shaping. The sounds are essentially different from the sounds the users may naturally expect when they touch the (virtual) objects under consideration.

At the rightmost extreme of the sound abstraction scale depicted in Figure 1 are the “earcons”. These are sounds which often consist of a well defined succession of tones in a specified rhythm and tempo. The sounds produced by telephones are a typical example, not only the bell tones, but also the sounds that indicate that the number is engaged or out of order, etc. Other examples are the sounds produced by electrical door bells, alarm clocks, ambulances, or fire engines. Also in this case, the relation between an event and a sound is not based on the mechanical or acoustical properties of the objects involved in the event. Earcons are very often used for drawing the user’s attention, for warning the user, or as alarm signals. As such they have a categorical relation with the process in which they are used. A detailed review about the design of attention, warning and alarm sounds is presented in [13, 14, 15, 16].

3. SELECTION OF SOUND SYNTHESIS METHODS

For the system under consideration, involving the evaluation and modification of virtual objects, there exists a number of different possible modes of haptic interaction. In this system, the user decides what data or information are represented by the auditory display.

For the purpose of sonification, three possible modes of haptic interaction are considered, as illustrated in Figure 2. These modes are used to evaluate and explore the shape of a virtual model. The first involves the impact between the user’s hand/finger and the model at a certain point, for example by tapping the object, as

shown in Figure 2(a). The second, Figure 2(b), involves a sustained contact with the object without movement. Finally, the third type of interaction, Figure 2(c), involves sliding a hand or finger over the object.

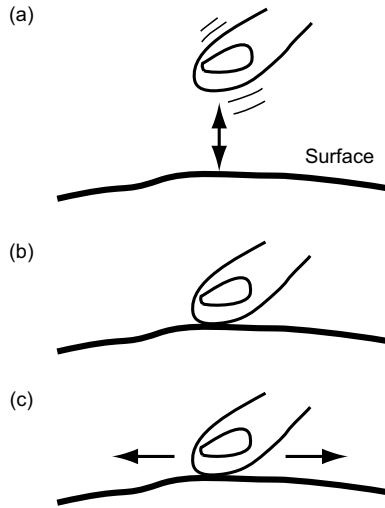


Figure 2: Modes of user interaction with the virtual model.

Considering these different modes of interaction, and the information that is to be represented, the type of sound synthesis can then be selected. Perceptually intuitive sounds can be used to intuitively convey information about surface attributes, such as material and texture. In order for the sounds to be perceptually intuitive, they must be modelled both according to the specific surface attributes, and also to the user's interaction. The sounds will therefore be triggered by the interaction of tapping and sliding. Strict physical models of the sound are not considered for this purpose because they are difficult to implement in a real-time system. Perceptually intuitive sounds on the other hand can be synthesised with relatively simple models.

In order to sonify data representing defined geometrical properties of the virtual model, symbolic sounds can be used. The sounds are again triggered by the user's interaction with the model. Since these sounds are not necessarily perceptually intuitive, they can be triggered by each of the three types of interaction described above. For this same reason, however, the user will require some prior knowledge about the sound feedback and mapping. A possibility for sonification is to use the models designed to create perceptually intuitive sounds. As they are not strictly dependent on the physical parameters of the virtual object, they offer degrees of freedom that can potentially be exploited to sonify surface or object properties using an abstract mapping.

As well as the evaluation and exploration of the object, the possibility to modify its shape by direct interaction is an important aspect of the SATIN project. During the shape modification process, the user will be able to choose to enhance the interactive process with the use of sound and also to select between different modes of sonification. The sonification in this case will consist of two different possible modes:

Perceptually intuitive sounds can be used to characterise specific user actions that are applied to the interactive device, such as enlarging, bending and twisting. These actions can be mapped to perceptually intuitive sounds that do not necessarily need to have

any relation to the physical system, but do have an intuitive association with the action that is being performed. This use of perceptually intuitive sounds can be compared to the sound effects often used to accompany simple actions in cartoons, such as bouncing and jumping for instance. Symbolic sounds can be used as an aid to the fine tuning of shape parameters, for example by using amplitude or spatial modulations in the audio output. This is especially useful as very small changes in shape parameters may be undetectable using the other sensory modalities. The use of symbolic sounds could therefore serve to guide the user in such a situation if necessary.

In summary, the sounds generated for this project will consist mostly of both perceptually intuitive sounds and of symbolic sounds. Perceptually intuitive sounds offer a more realistic experience to the user, and depend on certain attributes of the virtual object. However, they need only be intuitive, rather than derived from exact physical models. This means that there is the possibility with such sounds to sonify attributes that would not naturally be associated to the object. For symbolic sounds there is no requirement for an intuitive mapping and therefore any data or attributes of the object can be sonified. These sounds can be triggered by all types of interaction with the object, or even by no interaction with the object at all.

4. SONIFICATION OF CURVATURE DATA

A specific requirement of the SATIN project is the presentation to the user of *curvature data* relating to the surface of the interactive object. Curvature data is of particular interest to designers when considering the aesthetic quality of the surface. In the automotive industry for example, a smooth surface that is continuous in terms of curvature is highly desirable because discontinuities in curvature result in a disconnected appearance in the light reflected from the surface. Designers refer to such surfaces as *Class A surfaces* and they consider them as being aesthetically appealing.

Sonification is considered to be an advantageous solution for presenting curvature data. There are other potential media for communicating this information such as visualization and haptic devices. These media, however, are already employed in the SATIN project for a number of different tasks. For this reason, their use for this particular task has the risk of overloading the visual and tactile senses of the user. In addition, they may also require more expensive technology. The use of sound, on the other hand, is not employed for any other operation in the project and also only requires a PC with a sound card.

The sonification in this case is triggered by the exploratory interaction of the user. For this project, curvature is considered only for a cross sectional slice of the virtual object. For a plane curve C , the curvature at point p has a magnitude equal to the reciprocal of the radius of the osculating circle, i.e. the circle that shares a common tangent to the curve at the point of contact. The curvature is a vector that points to the centre of the osculating circle. For a plane curve given explicitly as $C = f(p)$ the curvature is given by:

$$\mathcal{K} = \frac{\frac{d^2 C}{dp^2}}{\left(1 + \left[\frac{dC}{dp}\right]^2\right)^{\frac{3}{2}}} \quad (1)$$

For this application, only the curvature magnitude is considered. A negative value implies the curve is convex and a positive

value implies that the curve is concave. Aspects of the curvature data such as discontinuities in curvature are difficult to detect by touching the object or through the visual feedback of the system. This is made clear in Figure 3. Figure 3(a) illustrates three different curve shapes, all of which are apparently smooth. The curves are labelled Curve A, Curve B, and Curve C. Figure 3(b) shows the curvature magnitude corresponding to these curve shapes. Note that the shape of curve C has two discontinuities in curvature data. However, the graph describing its shape is smooth and continuous along the entire length of the curve. In the SATIN project the visual feedback used by the system is concerned only with an accurate display of the virtual object, and the tactile feedback is concerned only with the reproduction of the object shape. Sonification can therefore be used to enhance the exploration by allowing designers to become more sensitive to such details that otherwise are practically undetectable. The sonification approach has the advantage that the visual and tactile outputs of the system are not compromised.

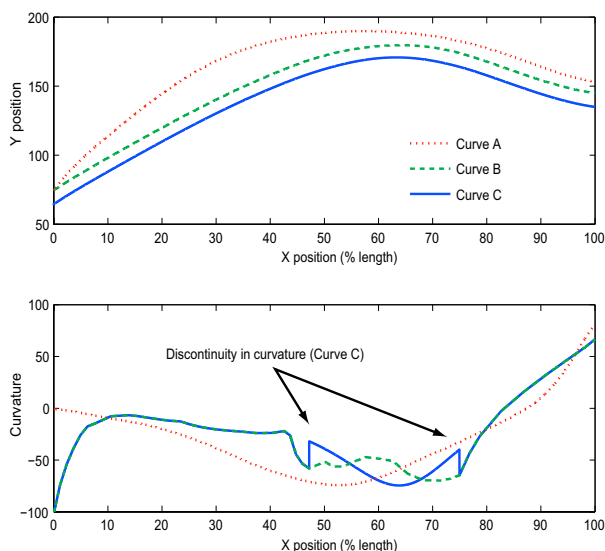


Figure 3: Shape and curvature.

Three distinctive aspects of these data are explicitly considered for curvature sonification. The absolute value of the curvature, indicating how strongly curved the surface is, the sign of the curvature, indicating whether the surface is concave or convex, and finally potential discontinuities in the curvature, which are of particular interest to designers using this system. These parameters are mapped to different acoustic variables used to control or modify a sound source. A straight forward example of how such a system can be implemented follows.

First a sound source or generator must be selected. In the simplest case a pure tone can be used, or other basic signals such as a sawtooth waveform or narrowband noise. This straight forward and purely symbolic sound synthesis approach is presented here as a proof of concept. However, we are currently developing more elaborate synthesis algorithms based on perceptually intuitive techniques and simplified physical models, for example using modal synthesis methods [5, 6].

In the current implementation, the absolute value of the curvature is mapped to the frequency associated with the source. The

mapping is implemented in such a way that the minimum absolute value, 0, of the curvature is mapped to a frequency of 200 Hz and the maximum absolute value found in the dataset is mapped to a frequency of 1600 Hz, resulting in a frequency range of 3 octaves. Frequency is a good choice for this mapping because humans perceive frequency changes with a relatively high resolution. Typically, frequency changes can be determined with an accuracy of up to 0.3% [17]. The mapping to frequency is also a robust choice because its perception is generally not dependent on the acoustic surroundings of the demonstration like reverberation or the audio apparatus used.

The sign of the curvature is mapped to the stereo panning of the sound output in the following way. If the data are negative, the output is weighted to the left channel. More precisely, if the absolute value of the negative data is more than 20% of the maximum absolute value of the curvature data, the sound is output entirely in the left channel, and no output is heard in the right channel. Similarly, if the data are positive, the output is weighted to the right channel, and if the value is more than 20% of the maximum absolute value of the dataset, the sound is output entirely in the right channel. Between these two reference points, the output is linearly cross faded between the two channels and a curvature value of 0 is mapped to a sound reproduced from the middle between the two loudspeakers.

To signify discontinuities in the curvature data, two different modulations are applied to the sound. When the finger is moved across a discontinuity position on the surface, the amplitude of the audio output is rapidly increased and then decreased, resulting in a click-like sound played in synchrony with the finger's movement across the discontinuity. This sound is easily distinguished from the pure tones associated with ordinary curvature values and gives the user initial feedback about the presence and the approximate position of the discontinuity. In order to allow a precise exploration of the position of the discontinuity, a specific sound is reproduced if the finger rests on the exact position of the discontinuity. This sound is made by rapidly alternating between the two frequencies which correspond to the curvature values on either side of the discontinuity. Discontinuity positions are thus the only points along the surface at which the reproduced sound is not stable over time, but is frequency-modulated. This modulation continues until the finger moves away from the spatial position of the discontinuity.

5. DISCUSSION AND EVALUATION

In this paper, we have shown one potential way of using sound for representing geometrical information of virtual objects in the context of the SATIN project. A specific implementation for the sonification of curvature data has been described. There is a practically unlimited number of methods and strategies, however, that can be used to sonify the data. There are also a number of factors to consider such as how effectively the data are represented, and the emotional effect the audio feedback will have on the user, e.g. is the sound annoying?

The implementation described here considers only the use of purely symbolic sounds, or sounds that have no intuitive relationship with the virtual model. As discussed in section 3, one potential alternative approach is to synthesize perceptually intuitive sounds as the source for the sonification process. These sounds will also have an intuitive relationship with the action of the user. For example the *brightness* of the sound can be mapped to the ve-

locity of the users' movement, while the *sharpness* can be mapped to the pressure the users exert with their fingers on the virtual surfaces. Other parameters of the synthesis method can then be mapped to the curvature data, for instance the intensity, the bandwidth or the modal frequency.

A perceptual evaluation study will be designed to test the different sonification approaches and mapping strategies. Test subjects will be asked to carry out tasks relating to the exploration of curvature data through interaction with the haptic interface. The usability of the different sets of sound will then be determined by the three standard measures of usability, efficiency, effectiveness, and satisfaction [18]. This means that in the experiments the time it takes to complete the tasks will be a measure for the efficiency of the auditory display. The number of errors made by the users will also be counted, and they will yield a measure for the effectiveness of the auditory display. A questionnaire will then be developed to measure the user satisfaction. This will include the experienced ease of use, intuitiveness and pleasantness.

In this paper, we have proposed a classification scheme of sound synthesis methods that are used as part of an interactive virtual reality environment. In the SATIN project, the user is able to physically interact with and manipulate a virtual object. The classification scheme we have introduced is based on the level of abstraction of the synthesised sound in relation to the virtual object and the user's interaction with it. Using this scheme as a starting point, we have presented an example that demonstrates the advantage of the use of sound and sonification in this virtual reality application and we have described a potential implementation.

In future work, we plan to implement a number of different sound synthesis approaches and sonification mapping strategies. We then plan to test these using a perceptual evaluation study, in order to measure them in terms of their usability, i.e. their efficiency, effectiveness and satisfaction in the context of the application.

6. ACKNOWLEDGEMENTS

The research work presented in this paper has been supported by the European Commission under the project FP6-IST-5-054525 SATIN – Sound And Tangible Interfaces for Novel product design (<http://www.sat-in-project.eu>).

7. REFERENCES

- [1] J. J. Gibson, *The senses considered as perceptual systems*. Houghton Mifflin, 1996.
- [2] W. W. Gaver, "How do we hear in the world? Explorations in ecological acoustics," *Ecological Psychology*, vol. 5, pp. 285–313, 1993.
- [3] —, "What in the world do we hear? An ecological approach to auditory source perception," *Ecological Psychology*, vol. 5, pp. 1–29, 1993.
- [4] T. Tolonen, V. Välimäki, and M. Karjalainen, "Evaluation of modern sound synthesis methods," Helsinki University of Technology, Department of Electrical and Communications Engineering, Tech. Rep., 1998.
- [5] K. Van den Doel and D. K. Pai, "The sounds of physical shapes," *Presence*, vol. 7, no. 4, pp. 382–395, 1998.
- [6] K. Van den Doel, P. G. Kry, and D. K. Pai, "Foley automatic: Physically-based sound effects for interactive simulation and animation," in *Proceedings of ACM SIGGRAPH*, Los Angeles, CA, USA, August 2001, pp. 12–17.
- [7] D. Rocchesso and F. Fontana, Eds., *The Sounding Object*. Mondo Estremo, 2003.
- [8] I. Díaz, J. Hernantes, I. Mansa, A. Lozano, D. Borro, J. Gil, and E. Sánchez, "Influence of multisensory feedback on haptic accessibility tasks," *Virtual Reality*, vol. 10, pp. 31–40, 2006, special Issue on Multisensory Interaction in Virtual Environments.
- [9] H. P. Kjær, C. C. Taylor, and S. Serafin, "Influence of interactive auditory feedback on the haptic perception of virtual objects," in *Proceedings of the 2nd International Workshop on Interactive Sonification*, York, UK, 2007.
- [10] S. Conversy, "Ad-hoc synthesis of auditory icons," in *Proceedings of the 5th International Conference on Auditory Display*, Glasgow, UK, 1998.
- [11] T. Hermann, "Sonification for exploratory data analysis," Ph.D. dissertation, Bielefeld University, Germany, 2002.
- [12] T. Hermann and A. Hunt, "An introduction to interactive sonification," *IEEE multimedia*, vol. 12, no. 2, pp. 20–24, 2005.
- [13] N. E. Stanton and J. Edworthy, Eds., *Human factors in auditory warnings*. London, UK: Ashgate Publishing, 1998.
- [14] S. A. Brewster, "Using non-speech sound to overcome information overload," *Displays*, vol. 17, pp. 179–189, 1997.
- [15] —, "The design of sonically-enhanced widgets," *Interacting with Computers*, vol. 11, pp. 211–235, 1998.
- [16] —, *The human-computer interaction handbook: Fundamentals, evolving technologies and emerging applications*. Mahwah, NJ, USA: Lawrence Erlbaum Associates, Inc., 2002, ch. Non-speech auditory output, pp. 247–264.
- [17] C. C. Wier, W. Jesteadt, and D. M. Green, "Frequency discrimination as a function of frequency and sensation level," *Journal of the Acoustical Society of America*, vol. 61, pp. 178–183, 1977.
- [18] *Ergonomics of Human-System Interaction*, International Organization for Standardization Std. ISO 9241, 2007.