MODELING AND SONIFYING PEN STROKES ON SURFACES

Christian Müller-Tomfelde and Tobias Münch

IPSI - Integrated Publication and Information Systems Institute
Fraunhofer - IPSI, Dolivostr. 15, D-64293 Darmstadt, Germany
{mueller-tomfelde,muench}@ipsi.fhg.de

ABSTRACT

This paper will describe the approach of modeling and sonifying the interaction with a pen on surfaces. The main acoustic parts and the dynamic behavior of the interaction are identified and a synthesis model is proposed to imitate the sound emanation during typical interactions on surfaces. Although a surface is two-dimensional, modeling acoustical qualities of surfaces has to employ volumes to form resonances. Specific qualities of surfaces like the roughness and the texture are imitated by a noise generator which is controlled by the pen movement in real-time to achieve a maximum of acceptance of the sound effect. The effect will be used one hand to produce natural and coherent interaction on nearly silent electronic white boards or pen-tablets, i.e., reinventing of lost sound qualities. On the other hand modeling and sonifying pen strokes on surfaces allow to convey information about the properties of different areas or the current state of a windows of a computer display by using this sound feedback.

Keywords : sound model, human-computer interaction, real-time, disappearing computer, audio feedback, sonification, mixed reality, multi-modal

1. Introduction

As we can learn from writing with chalk on whiteboard, with pencil on the table-top or with a board maker on a flipchart, pen strokes in the physical environment produce specific sounds that are characteristic in multiple directions. The sound stimulated by the interaction is depending on the kind of surface we are writing on, the kind of pen we are writing with and, of course, the way of how we write. In a real-world situation all these sound cues are side effects that convey secondary and redundant information to the writer and listener. But especially this kind of information which is not in the focus of the listeners attention and which is perceived unconsciously could help to make the overall interaction in a virtual environment or mixed reality more natural and coherent.

Two tendencies in the development of current computer technology underline that in the near future the usage of secondary and redundant information of interaction could play a important role not only as nowadays in high-end virtual reality environments, e.g., [1] but also in standard off-the-shelf computer systems: on one hand touch sensitive displays and tactile input devices become more and more popular and the user interface becomes multi-modal. On the other hand due to the ubiquitous presence of computers and, at the same time, due to the miniaturisation of the computer equipment (see [2][3][4][5]) the traditional sounds of the computer will get lost. This disappearing computer allows us to rethink of the role of the sound for the interaction. The user interacts more with information objects than with the computer because the computer not only physically but also, and this is more important, mentally disappears. A consequence could be that distracting sounds will vanish and useful sounds can be reinvented or, on the next level of user interface design, invented new. Hence, a great field for situation and context dependent sound feedback will appear, when there is no more physical reason for sound emanation of a computer environment.

In a computer environment e.g. with an electronic white board where the finger or a dedicated pen is mainly used to manipulate virtual documents (see e.g. [6]) all the secondary and redundant sound information vanishes and gets lost, although the ability of people to differentiate between types of surfaces and identify the qualities of interaction is still available. By simply using a pen as shown in figure 1, but, of course, without ink would solve the problem of missing sounds on an electronic white board, but the idea of inventing new situation and context depending sound feedback possibilities would be impossible.

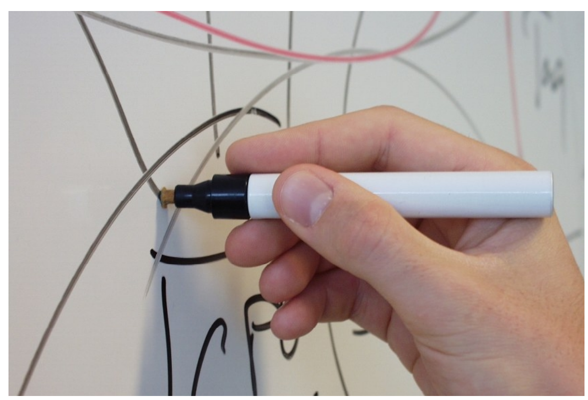

Figure 1. *An example of typical natural sonifyed pen strokes: board marker on a white board*

From the perception point of view the sound feedback on virtual surfaces is an example of a true cross-modal perception in a mixed-reality environment, i.e., hybrid world, where the missing tactile information is offered to be perceived by imitated natural sound cues. In our approach the goal is not to simulate the sound of a certain constellation of surface and pen, the focus is more to imitate learned patterns and utilise them in a hybrid computer environment (like cartoon sounds, sounds with enhanced characteristics). The sound processing parts of the modelled pen stroke, discussed later are simple, easy to implement and predictable in their behaviour, like noise generator, linear filter or modulator. The complexity and richness in the sound of interaction are built on the combinations of the sound processing units with the control parameters of the user input device. Again, the overall sound system is designed to react to the users interaction according to what the user expects, i.e., what the user has learned in the past while his interaction with real matter and based on these experiences new forms of sound feedback can be proposed in future.

1.1. Overview

The work is motivated and inspired by the research experiences of the authors of usage of large electronic whiteboards (see [7]), pen tablet displays in cooperative work situations and by the work of Bill Gaver (see [8]). Initial recordings of sounds of typical interactions were made to investigate the timbre characteristics of different types of pens and surfaces.
First we will describe the model for the interaction with a pen on a surface followed by a description of the output parameter of a standard computer drawing device, usually a mouse pointer, their derivatives and higher-level processed features. Then the main passive acoustical parts of the interaction and the sound stimulation are identified. The next section of the paper concerns more with the properties of surfaces and their audible characteristics. A model for the stimulation of the passive system and the implementation environment is introduced. Finally an overview of related work is described followed by the conclusions and future work.

2. Interaction with a pen

The standard situation that is the base for our considerations is the following: a user holds a pen in his hand and presses the tip of the pen on the two-dimensional surface of a table, or more general a corpus (see figure 2). The users writes, draws or annotates something on the surface, optional he can support his hand position by leaning on the surface. Depending on the application the pen could be a pencil, a ball-pen or a board marker etc. For our consideration we assume that the pen has a cylindrical body and a tip of sold material so that the tip is sliding on the surface.

2.1. Interaction model

A more detailed analyze of the pen interaction, described above leads to the interaction model. For building the model the most important area is, the so called, contact area where tip of the pen touches the surface. In a closer view the surface is not anymore an ideal and flat two-dimensional surface (see detail of figure 2). There are peaks of different heights and the distance between them could be either constant or irregular. When the user moves the pen over the surface the pen tip and these peaks collide in the contact area. When the input force is raised over a certain level the pen tip is released and hops to the next peak. Finally, these attacks and releases of the force at the peaks excite the volumes that participate the interaction. Mainly the body below the surface and the body of the pen start vibrating and emit sound waves. We do not consider the acting forces explicitly because the main goal of the model is to imitate gliding and sliding on surface and not scratching and manipulation of a surface. Only modeling a simple impact of the pen on the surface could be interesting from our application point of view. Since the force parameter of the interaction is not detectable, the speed of the tip relative to the surface is the most relevant interaction parameter. In the two-dimensional plane the velocity of the tip can be easily calculated from the current and the last position of the tip. The consideration of the surface asperities also leads to a relation between the velocity and the sequence of the pikes that excite the resonating volumes. The greater the velocity of the tip the more peaks are hit (this will be discussed in section 3.2).

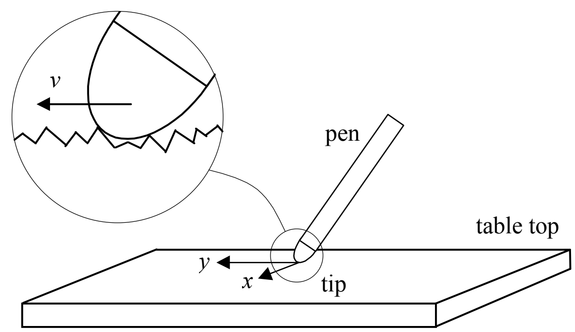

Figure 2. *The model of the interaction of a pen sliding over a surface. The pen tip is moved over the two-dimensional surface with a speed v. The attack and release at the peaks excites the acoustical system.*

2.1.1. Drawing device

The available data output of the standard computer drawing device is the coordinate of the current position of the pen or cursor and, after a certain processing, the derivatives like speed, angle and angle speed. All these control values are sampled at fixed sampling frequency which is usually below the audio sample frequency, e.g. 10Hz up to 100Hz. In general the output data of the drawing device can be characterized by primary and secondary events. The position over time of pen and pressure on the surface, as well as the pen-down or pen-up events are understood as primary events. While the speed and the direction of the movement of the pen are secondary events to control the sound generation of the interaction with the surface in real-time.
On a more higher level of processing one can differentiate between translation, i.e., no change of the angle and rotation. The rotation is characterized by a constant angle speed or rotation frequency, i.e., drawing a circle with constant radius. Unfortunately not detectable with standard pointing hardware is the angle between the axis of the pen and the table-top. This angle has certain effect to the sound of the sliding tip. It can be simply described as the difference between moving the pen forward and backward or pulling and pushing. The reason for this effect can not be found in the quality of the surface. In our realization the exact pen-table-top-angle over time has minor

importance and is substituted by the angle in the horizontal plane of the interaction. While using a three-dimensional stylus or digitizer it would be possible to detect the correct pen-table-top-angle. A maximum of naturalism of the sound feedback can be mainly achieved by providing a feedback to the interaction with low latency and high precision.

2.2. Model of the passive acoustical parts

The two dominant volumes that are involved in sound emanation during the interaction is the volume below the surface, i.e., the table-top and the pen. The first volume is assumed to be a rectangular one and the second one will be approximated by the cylindrical form (see also figure 2). The sound characteristics of both shapes are well studied in literature. Because these shapes can be easily described and are not as complex as other real shapes they are named basic shapes (see [12]). Their spectral acoustical behavior is entire mathematically solved without applying sophisticated algorithms designed (see [13][14]), like those to predict room acoustics, like ray-tracing or image source. But of course, this description in the frequency domain does not allow to simulate the spatial sound propagation within or outside the shape. For a rectangular volume or resonator the equation

$$f_r^r(l,m,n) = \frac{c}{2}\sqrt{\left(\frac{l}{X}\right)^2 + \left(\frac{m}{Y}\right)^2 + \left(\frac{n}{Z}\right)^2} \qquad (1)$$

allows to calculate the resonance frequencies f_r, where X,Y,Z are the length, width and height of the shape and l, m and n denote the orders of the particular mode. The speed of the propagation of the sound waves c is depending on the material. In the case of the cylindrical one the model consists only of one dominant frequency

$$f_r^c(l) = \frac{c}{2}\frac{l}{X} \qquad (2)$$

where X denotes the length of the tube. The diameter of the tube can be used to imitate the refection coefficient, but is not important for modeling the cylinder. For both models exist simple, efficient and easy to implement signal processing structures that base on the rich and complex sounding comb filter structure and low-pass filter (see [12][15][16]). These linear-time invariant systems (LTI) are well-known and predicable in their real-time behavior and the ratio between computational load and sound complexity is good due to the recursive filter with infinite impulses responses (IIR).
The goal the application of these passive acoustical models in the context of interaction would be that the user identifies these volumes as different ones in the overall output sound. The output sound should be perceived as a superposition of two excited and resonating volumes. As in the discipline of room acoustics the addition of the space or volume information does not affect the perception of the input sound, i.e., the listener is able to separate sound from the space where it is performed.

2.3. Stimulation of the system

While the modeling of the resonating parts (as in 2.2) describes the influence of the passive and static parts to the sound, the role of the stimulation of the system by the movement pen is more dynamic and time-variant and requires more attention. The most important interaction parameter that controls the stimulation is the speed of the pen sliding over the surface. Two examples should motivate our approach.: In comparison to a musical instrument like the violin, the table top and the pen function as the instruments body which resonates due to the excitation by the movement of the pen tip on the surface (e.g. in [17] the transfer function of the body of a violin separated of the excitation by the string). The detection of a music sound signal from a record with a phonograph needle has also some parallels: invisible small vibrations during a more or less simple movement stimulate a resonating system to emit sound waves. All information about the qualities of the contact and the sliding has to be modeled in the sound stimulation while the pen and the table top function finally only as the "loudspeaker of the interaction".

3. Sound of surfaces

Usually the qualities of surfaces can be named by terms like, rough, coarse, soft or smooth. Although these qualities could also correspond to visual properties of the surface areas they are mainly related and determined by the haptic experiences during the interaction. As in [18] proposed the human visual perception is not correlated either spatial or temporal to the haptic perception surface details at a certain size. The individual representation or interpretation can be different but we assume there could be also a tendency that allows to specify objective categories of surface, like in the context of room acoustics, where a common vocabulary can be used to characterize the reverb.

3.1. Audio surface texture

During an interaction of a pen with surface the haptic experience is accompanied by auditory cues that correspond in general to tactile cues. For our approach we suggest a perceptual model for an audio surface texture consists of three sub models to form a convincing sound feedback without a real tactile feedback: a micro, meso and macro surface texture. We assume that the visual feedback in this context is less relevant and significance only appears when strong correlation between all modes occurs, e.g. at the boarder of two different types of surfaces or at edges of an object. Hence, the original bi-modal feedback is reduced to a single modal feedback with the goal of persistence of the surface quality illusion.
Micro surface texture: This category of the surface is supposed to be homogenous in the two dimension of the interaction surface. The single peaks of the surface (see figure 2) do not have any directionality and the peaks can not be identified spatially as single objects or events. Their spatial distribution is assumed to be stochastic. When a probe is sliding over these peaks the lateral force stimulating the acoustical system inherits these property and the emitted sound also appears to have gaussian characteristics. The variance to this gaussian process is invariant of the speed v of the probe because the histogram of the process stays the same. As in [19] the timbre of this process especially the dimension of the fractal noise could be used to control the auditory roughness of a surface.

Meso surface texture: In contrast to the latter definition of the micro surface texture the meso surface texture concerns more with the isolated perceptible peaks within the surface like the grain of wood or sandpaper. The pattern of these textures could have a preferred orientation. Thus depending on the direction of the sliding interaction the auditory feedback can vary and can convey information about the quality of the surface. In our model the spatial distribution of theses peaks are stochastic as well as those of the micro surface texture .
Macro surface texture: The macro surface texture is very close related to the visible pattern of the surface. Like on a chess-board the areas of different surface qualities have determined boundaries and have spatial a fixed position. The category of the macro surface texture is built by the definition of the model parameter over the whole interaction area of the two preceding categories.

3.2. Sound generation

The core sound generator, the sound stimulator located at the pen tip is formed by the superposition of the models for the first two of the three suggested categories while the macro surface texture is realized by controlling models of the micro and macro surface texture. The sound generation is controlled in real-time by two kind of parameter: the interaction input parameter like position and speed of the pen tip and the surface model parameter.

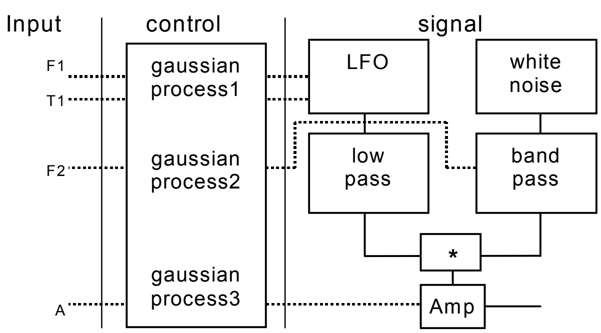

Figure 3. *The abstraction of the sound generator: a LFO modulates a white noise generator. The input parameters indirectly control the signal processing. Their adjustments are processed to obtain gaussian characteristics.*

The micro surface texture is mainly modeled with a white noise generator which is filtered by a band pass (with F2 the center frequency of the band) (see figure 3 right side). To add characteristics of a meso surface texture the output of the low frequency oscillator (LFO) is modulating the filtered white noise (see left part of the signal domain in figure 3). The LFO can be controlled in frequency F1 and in width of the pulses T1. The parameters of the oscillator are adjusted to the corresponding output parameter of the drawing device, i.e., the speed of the pen is mapped to the frequency F1 of the oscillator. The differentiation between micro and meso surface textures can also be understood a separation in the frequency domain of the signal processing: micro textures are processed in high audio frequencies while meso texture produce also audio signals but in a lower frequency range. To imitate the influence of the stochastic spatial distribution of peaks on the surface the oscillator frequency parameter F1 of the LFO is controlled by a gaussian process (GP) so that an stochastic change of the frequency can be adjusted by the variance of the GP. For example, with the variance of the GP1 one can control the irregularity of asperity of the meso surface texture. To achieve a maximum of flexibility nearly all parameter like the duration resp. length T1, the center frequency F2 and the amplitude A of the output are modulated by gaussian processes (see figure 3). This allows to map properties like pattern directivity, irregularity, etc.. The gaussian process is realized by applying the Box-Muller transformation to a uniform distributed white noise signal. In the case of a surface with homogeneous distribution of peaks it becomes clear that the faster the pen tip is moved over the surface the more often collisions with peaks become audible. Hence the frequency of the LFO must be proportional mapped to the speed of the pen to ensure the coherent perception of the surface characteristics during interaction. Finally, the cut-off frequency of the low pass filter (figure 3) smoothes the edges of the rectangular pulse before it is multiplied with a white noise signal. This can be used to modify and control the attack of collisions.

4. Implementation

The initial prototyping of the models were done under Max/MSP on a Macintosh G4 [20]. The graphical programming of both signal objects and control objects allows fast and rapid prototyping of sound synthesis on one hand and easy assessing of the sound effect (as proposed in [8]). The Max object 'mouse' gives access to the actual mouse coordinates and thus allows to calculate the speed of the mouse cursor. An Max abstraction called 'Drawingdevice' provides all succeeding models with all information about the users interaction. Different mouse devices were tested, a standard desktop computer mouse, a large touch screen device and a graphic tablet. For a good quality of the sound feedback the sampling rate of the pointing device should be as high as possible and a low system latency is important for acceptance of the feedback. In case of the white board sized touch screen the sampling rate of the point detection was not sufficient. The current work of the authors concentrate to implement and experiment with the sonifyed pen stroke on a Windows PC for a better and smooth integration into the application domain.

5. Related work

A lot of interesting and related work from different disciplines can be found in the research communities like those of computer music, human computer interaction and perception. Although all this research has specific applications and different points of departure it is important for this paper. Three areas of research are focussed and described but, of course the list stays uncompleted: research on haptic interaction, on sound synthesis and modeling and, finally auditory perception of the shape of objects. On of the most important work is the one of Bill Gaver. His publication on synthesizing auditory icons from 1993 covers all the three areas of research and can be understood as a foundation for this kind of work in which the auditory feedback is not designed merely to provide entertainment [8]. The work of Pai et al initial concentrates on the simulation haptic textures

[18]. The work gives an excellent analysis of the forces participating during the contact with surfaces. In various following papers new aspects of the haptic interaction are studied. Mainly it is the support and integration of sound effects for contact interaction by translating the normal and lateral forces into a so called audio force which is then responsible for the stimulation of the sound [21]. The application can be found in virtual realty and games, where the demand of efficient sound rendering in real-time with low latency has to be fulfilled in order to achieve high acceptance of the simulation resp. imitation [22]. Physical models of sounding objects and creation of virtual musical instruments are one aspect of the work of Cook. The research on physical inspired statistical particle models ([23]) are important for our models of surface sounds as well as the efficient and versatile implementation of the real-time sound synthesis on off-the-shelf-computers [24]. From the perception point of view this work on sonifyed pen strokes was influenced by research that investigates the sound characteristics of objects with a certain shape. These characteristics can be either added to an unechoic sound by filtering it with an impulse response, like in room acoustics [12] or the characteristics of a shape or object appear with in the sound during an interaction of the object with an other like impact, bouncing etc.. In the latter case the shapes could be two dimensional like drums etc. or three dimensional like spherical and cubic resonators[13][14].

6. Conclusions and Future work

In this paper we presented our work on a sound feedback that imitates the sound characteristics of the interaction of pen on a silent surface of interactive computer displays for direct manipulation. Initial recordings motivated the development of different models to imitate the sound cue. A simple and effective model was implemented using models of the passive acoustical parts and stochastic controlled sound generator, that allows to adjust surface qualities in different degrees. The sonifyed pen strokes will be implemented in future under Windows to experiment with this cross-modal feedback on the application layer.

7. REFERENCES

[1] Eckel, G., Application for the CyberStage Spatial Sound Server, In *Proceedings of the AES 16th International Conference*: *Spatial Sound Reproduction*, Rovaniemi, Finland, 1999 April 10-12, pp. 478-484.

[2] Shneiderman, B., Designing the User Interface - Strategies for Effective Human-Computer Interaction, 2nd Edition, Reading (Ma.) etc.: Addison-Wesley, 1992.

[3] Norman, D. A., The Invisible Computer. MIT Press, Cambridge, MA. 1998.

[4] Weiser, M., Designing Calm Technology, PowerGrid Journal, v101, July 1996, http://powergrid.electriciti.com/1.01.

[5] Weiser, M., The computer for the twenty-first century. *Scientific America*n, Vol. 265, No. 3, pp. 94-104, 1991.

[6] Streitz, N. et al, i-LAND: An interactive Landscape for Creativity and Innovation In: *ACM Conference on Human Factors in Computing Systems (CHI '99)*, Pittsburgh, Pennsylvania, U.S.A., May 15-20, 1999. ACM Press, New York, 1999, pp. 120-127.

[7] Müller-Tomfelde, C. and Steiner, S., Audio-Enhanced Collaboration at an Interactive Electronic Whiteboard. In *Proc. of 7th International Conference on Auditory Display*, ICAD, Espoo, 2001. Pp 267-271.

[8] Gaver, W., Synthesizing auditory icons. In INTERCHI '93,24-29 April 1993, pages 228-235, 1993.

[9] Gaver, W., Sound Support for Collaboration, In: *Proceedings of ECSCW ´91*, Amsterdam, 1991.

[10] Gaver, W., The SonicFinder: An Interface That Uses Auditory Icons, In *Human Computer Interaction*, 4(1), 1989, pp. 67-94.

[11] Gaver, W., Auditory Interfaces, In: M.G. Helander, T.K. Landauer, P.V. Prabhu (Eds.), *Handbook of Human-Computer-Interaction*, 2. Edition, Part VI "Multimedia, Video and Voice, North-Holland 1997.

[12] Dutilleux, P. and Müller-Tomfelde, C., AML: Architecture and Music Laboratory, In *Proceedings of the AES 16th International Conference*: *Spatial Sound Reproduction*, Rovaniemi, Finland, 1999 April 10-12, pp. 191-206.

[13] Rocchesso, D. and Dutilleux, P., Generalization of a 3-D resonator model for the simulation of spherical enclosures, *Applied Signal Processin*g, vol. 2001, no. 1, pp. 15–26, 2001.

[14] Rocchesso, D., Acoustic cues for 3-d shape information. In *ICAD 2001*, Espoo, Finland, July 29-August 1, 2001, pages 180-183.

[15] Roads, C., *The Computer Music Tutorial,* 3. Edition, The MIT Press, 1996.

[16] Zölzer, U., Digital Audiosignalverarbeitung Vorlesungsskriptum der TU Hamburg Harburg.

[17] Farina, A., Langhoff, A., Tronchin, L., Realisation of 'virtual' musical instruments: measurements of the Impulse Response of violins using MLS technique, in Proc. of CIARM95, *2nd Internat. Conf. on Acoustics and Musical Research*, Ferrara 19-21 May 1995.

[18] Siira, J. and Pai, D.K. Haptic Textures – A Stochastic Approach. *IEEE International Conference on Robotics and Automatio*n,Minneapolis MN, April 1996.

[19] van Doel, K., Sound Synthesis for Virtual Reality and Computer Games, Thesis at the The University of British Columbia, 1998

[20] Max/MSP, Cycling 74, http://www.cycling74.com.

[21] DiFilippo, D. and Pai, D.K. Contact Interaction with Integrated Audio and Haptics, in Proc. *ICAD 2000*, Atlanta, Georgia, USA, April 2-5, 2000

[22] van Doel, K., Physically-based Sound Effects for Interactive Simulation and Animation, ACM SIGGRAPH 2001, 12-17 August 2001, Los Angles, CA, USA.

[23] Cook, P. R., Physically Informed Sonic Modeling (PhISM): Synthesis of Percussive Sounds. *Computer Music Journa*l, 21:3 1997, 38-49.

[24] Cook, P. R. and Scavone, G. P., The Synthesis ToolKit (STK), version 2.1. *In Proc. of the International Computer Music Conference*, Beijing, 1999.