Synthesizing Non-Speech Sound to Support Blind and Visually Impaired Computer Users

A. DARVISHI, V. GUGGIANA, E. MUNTEANU, H. SCHAUER

Department of Computer Science (IfI)
University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland

M. MOTAVALLI

Swiss Federal Laboratories for Material Testing and Research (EMPA)
Uberlandstrasse 129, CH-8600 Dubendorf, Switzerland

M. RAUTERBERG

Usability Laboratory, Work and Organizational Psychology Unit
Swiss Federal Institute of Technology (ETH), Nelkenstrasse 11, CH-8092 Zurich, Switzerland
Tel: +41-1-632 7082, email: rauterberg@rzvax.ethz.ch

Abstract. This paper describes work in progress on automatic generation of "impact sounds" based on physical modelling. These sounds can be used as non-speech audio presentation of objects and as interaction-mechanisms to non visual interfaces. In this paper especially we present the complete physical model for impact sounds "spherical objects hitting flat plates or beams." The results of analysing of some examples of recorded (digitised) "impact sounds" and their comparisons with some theoretical aspects are discussed in this paper. These results are supposed to be used as input for the next phases of our audio framework project. The objective of this research project (joint project University of Zurich and Swiss Federal Institute of Technology) is to develop a concept, methods and a prototype for an audio framework. This audio framework shall describe sounds on a highly abstract semantic level. Every sound is to be described as the result of one or several interactions between one or several objects at a certain place and in a certain environment.

Keywords: non speech sound generation, visual impairment, auditory interfaces, physical modelling, auditive feedback, human computer interaction, software ergonomics, usability engineering, material properties

1 Introduction

The following sections describe the background about using computers by blind computer users and the development of graphical user interfaces and their impact on this group of computer users. Subsequently two approaches for presenting non visual interfaces will be described. A short description about using non speech audio in different applications follows, two basic strategies for creating non speech audio and our approach to automatic generation of non speech audio are introduced. At the end this paper describes some steps which are carried out and those which are supposed to be done by the end of the project.

2 Background

For much of their history, computer displays have presented only textual and numeric information to their users. One benefit of this character-based interface, was that users who were blind could have fairly easy access to such systems. Users with visual disabilities could use computers with character-based interfaces by using devices and software that translated the characters on the screen to auditory information (usually a synthesised human voice) or/and tactile terminals and printers. One of the most important breakthrough in HCI (Human Computer Interaction) in recent years was the development of graphical user interfaces (GUIs) or WIMPs (Windows, Mice, Pointers). These interfaces provide innovative graphical representations for system objects such as disks and files, and for computing concepts such as multitasking by windows. GUIs are not powerful because they use windows, mice, and icons. Rather it is the underlying principles of access to multiple information sources, direct manipulation, access to multitasking, and intuitive metaphors which provide the power [Mynatt-91]. Since the mid 80's, the computer industry has seen a remarkable increase in the use of GUIs, as a means to improve the bandwidth of communication between sighted users and computers. Unfortunately, these GUIs have left a part of the computing population behind. Presently GUIs are all but completely inaccessible for computer users who are blind or severely visually-disabled. Today, there are some commercial products which are able to convert textual information on the screen of GUIs into synthetic speech, however they are absolutely insufficient. This is due to the fact that historic strategies for providing access for these users are inadequate [Boyd-90, Bux-86]. Today, there exists no simple mapping from graphical window based systems into the auditory or tactile domains.

3 Models for Non-Visual Presentation of GUIS

Based on different models there are two approaches for presenting information from GUIs in non visual forms [Crispien-94]. They are being investigated in two projects (see below). Both projects present prototypically information about graphical applications and also graphical computer environments in non visual form. These applications make use of on-screen graphical mechanisms (such as buttons and scroll bars) to control the application, and the environments provide an abstraction for the basic objects in the computer system, such as data files, directories, etc., and the basic computer operations, such as copying and deleting. The following diagram illustrates two strategies of deriving an audio/tactile perceptual manifestation of an interface display. Conceptual and perceptual mapping can be carried out either directly (e.g. a room environment representation - first approach) or indirectly (by using the visual model - second approach) [Gaver-89].

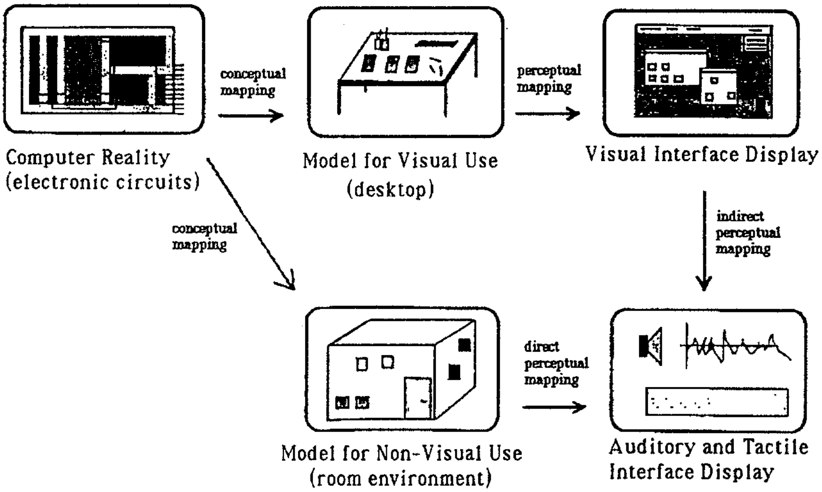
conceptual mapping
perceptual mapping
Computer Reality (electronic circuits)
Model for Visual Use (desktop)
Visual Interface Display
conceptual mapping
indirect perceptual mapping
direct perceptual mapping
Model for Non-Visual Use (room environment)
Auditory and Tactile Interface Display

Fig. 1. Two different mappings between computer world and user world

4 First Approach: Hierarchical Presentation of Interface Objects

The Mercator project [Mynatt-93] is an example for this approach. Its aim is to provide access to X-Windows and Unix workstations for computer users who are blind or severely visually impaired. The interface objects (such as icons, windows etc.) are organised in a hierarchical tree structure which can be traversed using a numerical key path. The primarily output modality is audio (synthetic speech and non speech audio) and since recently Braille output. The Mercator project uses the so-called "Audioroom" metaphors and spatial sounds to simulate the graphical computer environment.

5 Using Non-Speech Cues in the Mercator Project

The interface objects in the Mercator environment are called AICs (Auditory Interface Component). The type and attributes of AIC-objects are conveyed through auditory icons and so called "Filtears". Auditory icons are sounds which are designed to trigger associations with everyday objects, just as graphical icons resemble everyday objects [Gaver-89]. This mapping is easy for interface components such as trashcan icons but is less straightforward for components such as menus and dialogue boxes, which are abstract notions and have no innate sound associated with them. An example of some auditory icons are: touching a window sounds like tapping on a glass pane, searching through a menu creates a series of shutter sounds, a variety of push button sounds are used for radio buttons, toggle buttons, and generic push button AICs, and touching a text field sounds like an old fashioned typewriter.

6 Second Approach: Spatial Oriented Approach

The GUIB (Textual and Graphical User Interfaces For Blind People) is an example for this approach [see Crispien-94]. GUIB uses concepts and methodologies of visual interfaces and translates them into other modal sensories primarily Braille and also audio. Some special devices are built for input and output. Spatial 3-dimensional environments and auditory icons are also integrated.

7 Non Speech Audio

Non speech audio is being used in many fields including:

- Scientific audiolization [Blattner-92]
- User interfaces [Gaver-86] such as:
  - Status and monitoring messages.
  - Alarms and warning messages [Momtahan-93].
  - Audio signals as redundancy information to the visual displays to strengthen their semantics.
  - Sound in collaborative work [Gaver-91]
  - Multimedia application [Blattner-93]
  - Visually impaired and blind computer users [Edwards-88].

Just as with light, sound has many different dimensions in which it can be perceived. Visual perception distinguishes dimensions such as color, saturation, luminescence, and texture. Audition has an equally rich space in which human beings can perceive differences like pitch, timbre, and amplitude. There are also much more complex "higher level" dimensions, such as reverberance, locality, phase modulation, and others. Humans have a remarkable ability to detect and process minute changes in a sound along any one of these dimensions [Rossing-90]

8 Generation of Non Speech Audio

There is an increasing world-wide interest for generating and fast processing of audio data in many areas, especially for non visual interfaces. In principal there are two different ways of generating and processing audio data:

(1) digital recording, storing these audio data (sound samples), searching and playing them synchronously when needed;

(2) modelling audio data on a highly semantic level through extraction of relevant parameters of sound, storing the parameters that need a low rate of storage, generating audio data through these parameters.

We present results to go the second way [Rauterberg-94].

Every sound is to be described as the result of one or several interactions between one or several objects at a certain place and in a certain environment. Every interaction has attributes, that have an influence on the generated sound. Simultaneously, the participating objects, which take part in the sound generation process, can consist of different physical conditions (states of aggregation), materials as well as their configurations. In themselves also have attributes, which can have an influence on the generated sound.

The hearing of sounds in everyday life is based on the perception of events and not on the perception of sounds as such. For this reason, everyday sounds are often described by the events they are based on. First, a framework concept for the description of sounds is presented, in which sounds can be represented as audio signal patterns along several descriptive dimensions of various objects interacting together in a certain environment. On the basis of the differentiation of purely physical and purely semantic descriptive dimensions, the automatic sound generation is discussed on the physical, syntactical and semantic levels.

To be seen as physical descriptive dimensions are the sound pitch, frequency, volume and duration; as semantic dimensions, the differentiation of the interacting objects concerning their physical condition (state of aggregation) solid, liquid and gaseous has resulted.

Within the scope of this research project, we shall especially look for possibilities to describe the sound class "solid objects," in particular the class of the primitive sounds "knock" ("strike", "hit"), because this class of sounds occurs very frequently in everyday life and many blind persons orient themselves through auditory landmarks in everyday life, the interacting objects can be easily and well described by their material characteristics and the knowledge of solid-state physics and acoustics can be used. The generation of "impact sounds" based on physical models provides new possibilities to synthesise sounds for parametrized auditory icons [Gaver-93] especially for presenting objects and interactions in non visual interfaces.

In next sections we describe an experiment and the corresponding physical model for comparing the natural frequencies. The last section describes some phases of the project that are not carried out so far and still should be done in the next phases.

9 Experiments with Plates and Beams

We have investigated and modelled the sound patterns: "impact of a solid spherical object falling onto a plate or beam". A laboratory experiment (see fig. 2) was carried out to examine and optimise the derived theoretical model with the sound of a real impact. The experiments were done in an sound-proofed room.

Six different spheres (tab. 2) with different radii and one glass sphere were dropped from three different heights (tab. 1) on six different plates and beams (tab. 3). The signals were digitised via a DAT- Tape. 252 sound sequences (7 radii * 3 heights * (6 plates + 6 beams) = 252) were recorded. Hence some parameters which were derived from the theoretical model could be changed and examined.

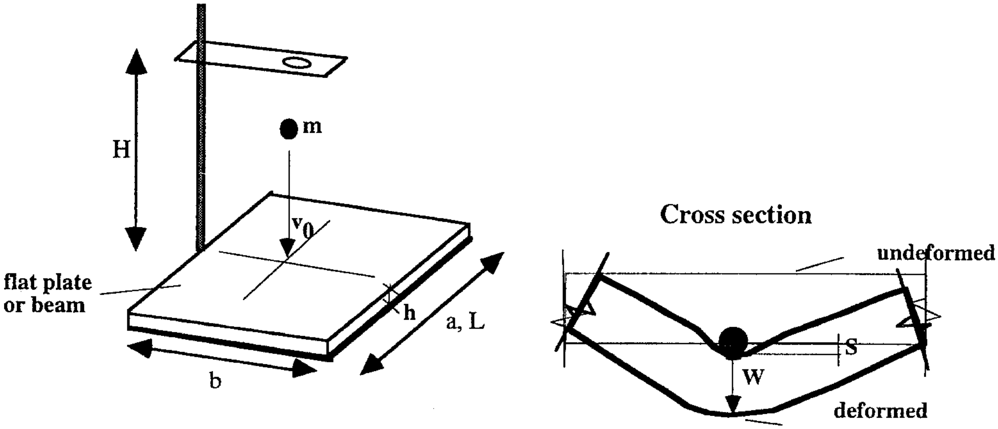

Fig. 2. Experimental Device

Height (H) [cm]	100	50	5

Tab. 1.: Heights

Name	g	s1	s2	s3	s4	s5	s6
Material	glass	steel	steel	steel	steel	steel	steel
Diameter [mm]	14.1	7.5	8.0	9.0	10.0	12.0	14.0
Mass [g]	3.65	1.72	2.09	2.98	4.07	7.02	11.16

Tab. 2.: Spheres

Material	h plate [mm]	h beam [mm]	density [kg/m^3]	Poisson's ratio	elasticity [Pa]
steel	2.96	2.96	7700	0.28	19.5E10
aluminium	3.98	3.98	2700	0.33	7.1E10
glass	7.94	7.94	2300	0.24	6.2E10
plexi	3.80	3.90	1180	-	-
PVC	6.00	6.12	-	-	-
wood	8.06	8.06	-	-	-

Tab. 3.: Properties

10 Physical Modelling of Thin Plates and Beams

The physical description of the behaviour of the plate's oscillations following the impact with the sphere provides variations of air pressure that we are able to hear. The natural frequencies of our small spheres are usually not in the audible range, therefore we don't care for the time being about this. However, we are concerned to include in our simulations the essential influence of the interaction on the impact sound.

How does the sound arise? The sphere hits the plate or beam and stimulates vibrations with the natural frequencies. These oscillations are transmitted to the medium as variation of pressure. The human ears receive these pressure waves and we interpret them as sound. The natural-frequencies of our small sphere are too high and they overtake the threshold of audibility. Thus, for the beginning we don't take them into consideration in our physical models.

General Notations:

E = Young's modulus	σ = Poisson's ratio	w = displacement
D = bending stiffness	ρ = density	h = thickness
I = moment of inertia	ρ' = mass per unit length	L = beam's length
a,b = length, width of plate	ω = angular frequency	f = natural frequency
w(x,y,t) = displacement	u(x,y)=static solution	m,n= integers

g(t) =function that describes the temporal component (damping)

Basic Relations: $\omega = 2\pi f$, $I = \frac{bh^3}{12}$, $D = \frac{Eh^3}{12(1-\sigma^2)}$, $\rho' = \rho \cdot b \cdot h$

10.1 Model of the Plate:

Equation of plate's motion:

$$D\left(\frac{\partial^4 w(x,y,t)}{\partial x^4} + 2\frac{\partial^4 w(x,y,t)}{\partial x^2 \partial y^2} + \frac{\partial^4 w(x,y,t)}{\partial y^4}\right) + \rho h \frac{\partial^2 w(x,y,t)}{\partial t^2} = 0$$

Solutions for natural frequencies in two particular typical boundary conditions:

clamped along all edges

$$f_{m,n} = 2\pi\left(\frac{m^2}{b^2} + \frac{n^2}{a^2}\right)\sqrt{\frac{D}{\rho h}}$$

simply supported at all edges

$$f_{m,n} = \frac{\pi}{2}\left(\frac{m^2}{a^2} + \frac{n^2}{b^2}\right)\sqrt{\frac{D}{\rho h}}$$

10.2 Model of the Beam:

Equation of beam's motion:

$$EI\frac{\partial^4 w(x,t)}{\partial x^4}+\rho' h\frac{\partial^2 w(x,t)}{\partial t^2}=0$$

Solutions for natural frequencies in two particular typical boundary conditions:

clamped at both ends

$$f_n=\frac{n^2\pi h}{L^2}\sqrt{\frac{E}{3\rho}}$$

simply supported

$$f_n=\frac{n^2\pi h}{4L^2}\sqrt{\frac{E}{3\rho}}$$

11 Software Synthesis of Impact Sounds

So far we have designed an audio signal generator for "impact sounds". This generator consists of object libraries to support the modelling of the audio signals which arise from the interaction of standard structures. These objects can be used to build up an infinite number of real models whose parameters are mechanical material properties and geometric dimensions. The concept "impact sounds" covers also audio signals such as: "scrapping", "rolling", "bouncing", "breaking", etc. They can be modelled using the same object libraries.

The components which are briefly described below the block diagram (fig. 3) are suitable for the planned architecture of the audio-generator as a stand-alone product.

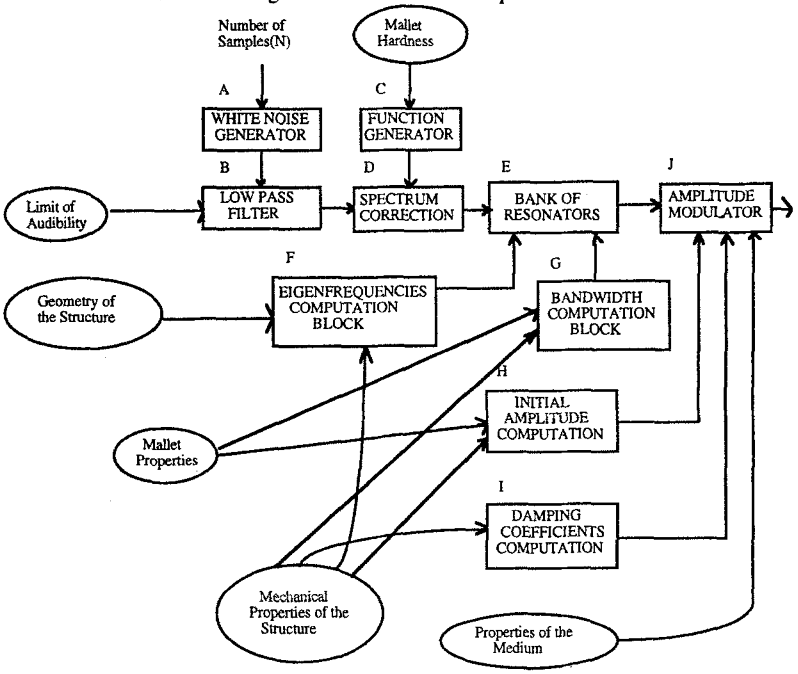

Fig. 3: Audio generator

A *Random White Noise Generator*

B *Low Pass Filter* cuts the frequencies above the audible range ($f_{\text{cut-off}} \approx 20$ kHz).

C,D *Function Generator, Spectrum Correction Block.* The analysis of "impact sounds" from tapping with hammers of varying hardnesses has shown that the energy released by the hammer into the structure is distributed over a certain frequency range. A hard hammer (of steel) transmits force into the structure and quickly deforms it, thus supplying a large portion of high frequency energy. Whereas a soft hammer (made of rubber for example) deforms the structure slowly and supplies mainly low frequency energy. This energy transmission function is calculated in Block C and drives the "spectrum correction" which is performed in Block D (see fig. 3).

E *Bank of Resonators* is a bank of filters which only let through natural-frequencies. The bandwidths are calculated in Block G.

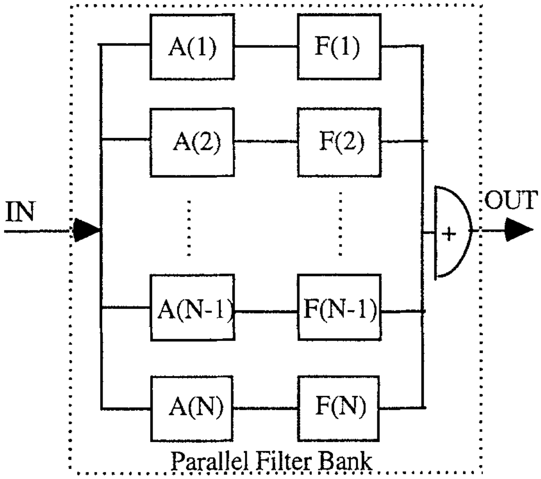

Fig. 4 Bank of Filters

This parallel implementation permits individual amplitude control for each natural frequency (Blocks A(1) - A(N)). The number of filters (N) is equivalent with the number of natural frequencies that are used in the model of vibrating structure. The transfer function of the bank of filters can be computed as follows:

$$H(z)=\frac{Y(z)}{X(z)}=\prod_{k=1}^{N}\frac{1-2e^{-2\pi B_kT}\cos(2\pi f_kT)+e^{-4\pi B_kT}}{1-2e^{-2\pi B_kT}\cos(2\pi f_kT)z^{-1}+e^{-4\pi B_kT}z^{-2}}$$

with f_k = the k-th natural frequency

B_k = bandwidth of f_k

T = sampling rate

N = number of resonators (poles)

By applying the inverse z transform to each term of the product, we obtain, the filtered signal. Samples of the output of every resonator k are computed from the input sequence $x[nT]$ by the recursive equation:

$$y_k[nT]=C_kx[nT]+D_ky_k[(n-1)T]+E_ky_k[(n-2)T]$$

where $y_k[(n-1)T]$and $y_k[(n-2)T]$ are the previous two sample values of the output sequence $y_k[nT]$.The constants C_k, D_k, E_k are related to the resonant frequency f_k and the bandwidth B_k of the resonator as follows:

$$E_k=-e^{-2\pi B_kT},\qquad D_k=2e^{-\pi B_kT}\cos(2\pi f_kT),\qquad C_k=1-D_k-E_k$$

F The *Eigen-frequencies Computation Block* calculates the natural frequencies of the particular structure (see chapter 10). These natural frequencies are dependent only on the structure.

G The *Bandwidth Computation Block* calculates the bandwidths of the bank of filters on the basis of the mechanical properties of the bodies which come in contact.

H The *Initial Amplitude Computation Block* calculates the initial amplitudes of the frequency components which essentially comprise the interaction effect.

I The *Damping Coefficients Computation Block* calculates the amplitude damping for the waves. It is dependent on frequency and the loss factor of the material of the structure. The following formula can be usefully applied for elastic materials:

$$\delta_n = \delta_0 \cdot \omega_n, \text{ where}: \omega_\text{n} = 2\pi \cdot f_n, \ \delta_0 = \frac{\tan\varphi}{2}$$

where tan φ = tangent of the loss factor

J The *Amplitude Modulator.* The values calculated under I and H modulate the amplitude of the signals provided by the bank of filters (Block E).

12 Comparison of Physical Model and Experiments

In the evaluation of the output of these two approaches we have to consider the inherent errors of input parameters as well as the inhomogenities of the materials. We compared the respective natural frequencies through spectrograms (see Fig. 2 and Fig. 3). The output error was situated in normal expectations as follows below. The initial amplitudes of the vibrations and the damping coefficients were introduced qualitatively. We have in view a quantitative, theoretic-modelled description of these two parameters. We will also study the possibilities to find the influence of the sphere in the sound generating event.

Errors in our parameters are given as percentile deviations:
E: ~ 1% σ: ~ 1% ρ: ~ 1%
h: ~ 1% a,b,L: ~ 1%.
Resulting errors in our solutions are given as percentile deviations:
- plate: ~ 4.5%
- beam: ~ 2.5%

Following activities should be carried out in the next steps of the project: (1.) description of a new model that contains all neglected aspects of interaction for impact sounds, (2.) implementation of the model, (3.) psycho-acoustical comparison studies. The subjects are asked to judge whether they can distinguish any difference between real and automatically generated sounds.

13 Conclusion

Having implemented our physical models in fast algorithms, we are now able to generate automatically the sound class of "spherical objects falling onto a beam or plate." It seems that our approach, to base our model on the physics of interacting objects is successful. The automatic generation of "impacted sounds" can be easily used to implement non speech audio cues for presenting objects and interactions in the non visual interfaces. The developed algorithms need low rate of storage and facilitate them to be incorporated in software systems.

References

[Blattner-92] Blattner, M. M., Greenberg, R. M. and Kamegai, M. (1992) Listening to Turbulence: An Example of Scientific Audialization. In: *Multimedia Interface Design*, ACM Press/ Addison-Wesley, pp 87-102.

[Blattner-93] Blattner, M. M., G. Kramer, J. Smith, and E. Wenzel (1993) Effective Uses of Nonspeech Audio in Virtual Reality. In: *Proceedings of the IEEE Symposium on Reaseach Frontiers in Virtual Reality*, San Jose, CA. (In conjunction with IEEE Visualization '93) October 25-26, 1993.

[Boyd-90] Bloyd, L.H., Boyd, W.L., and Vanderheiden, G.C. (1990) The graphical user interface: Crisis, danger and opportunity. *Journal of Visual Impairment and Blindness*, p.496-502.

[Crispien-94] Crispien,K. (1994) Graphische Benutzerschnittstellen für blinde Rechnerbenutzer. Unpublished manuscript

[Edwards-93] Edwards, W.K., Mynatt, E.D. and Rodriguez, T (1993) The Mercator Project: a non-visual interface to the X Window system. *The X Resource*,4:1-20.(ftp multimedia.cc.gatech.edu /papers/Mercator /xresource).

[Gaver-86] Gaver, W. W. (1986). Auditory icons: Using sound in computer interfaces. *Human-Computer Interaction.* 2, 167-177.

[Gaver-88] Gaver, W. W. (1988). Everyday listening and auditory icons. Doctoral Dissertation, University of California, San Diego.

[Gaver-89] Gaver, W. (1989) The SonicFinder: an interface that uses auditory icons. *Human Computer Interaction* 4:67-94.

[Gaver-90] Gaver, W. & Smith R. (1990) Auditory icons in large-scale collaborative environments. In: D. Diaper, D. Gilmore, G. Cockton & B. Shackel (eds.) *Human-Computer Interaction - INTERACT'90.* (pp. 735-740), Amsterdam

[Gaver-91] Gaver, W., Smith, R. & O'Shea, T. (1991) Effective sounds in complex systems: the ARKola simulation. in S. Robertson, G. Olson & J. Olson (eds.), *Reaching through technology CHI'91.* (pp. 85-90), Reading MA: Addison-Wesley.

[Gaver-93] Gaver, W. (1993) What in the World do We Hear? An Ecological Approach to Auditory Event Perception. *Ecological Psychology*, 5(1).

[Momtahan-93] Momtahan, K., Hetu, R. and Tansley, B. (1993) Audibility and identification of auditory alarms in the operating room and intensive care unit. *Ergonomics* 36(10): 1159-1176.

[Mynatt-92] Mynatt, E.D. and Edwards, W.K. (1992) The Mercator Environment: A Nonvisual Interface to XWindows and Workstations.*GVU Tech Report GIT-GVU-92-05*

[Mynatt-92b] Mynatt, E.D. and Edwards, W.K. (1992) Mapping GUIs to auditory interfaces. In: *Proceedings of the ACM Symposium on User Interface Software and Technology UIST'92.*

[Mynatt-93] Mynatt, E.D. and Weber, G. (1993) Nonvisual Presentation of Graphical User Interfaces: Contrasting Two Approaches. *Tech Report/93*

[Rauterberg-94] Rauterberg, M., Motavalli, M., Darvishi, A. & Schauer, H. (1994) Automatic sound generation for spherical objects hitting straight beams. In: Proceedings of "World Conference on Educational Multimedia and Hypermedia" ED-Media'94 held in Vancouver (C), June 25-29, 1994.

[Rossing-90] Rossing, T.D (1990) The Science of Sound 2nd Edition, Addison Wesley Publishing Company

[Sumikawa-86] Sumikawa, D. A., M. M. Blattner, K. I. Joy and R. M. Greenberg (1986) Guidelines for the Syntactic Design of Audio Cues in Computer Interfaces. In: *19th Annual Hawaii International Conference on System Sciences.*