AUTOMATIC IMPACT SOUND GENERATION FOR USING IN NON-VISUAL INTERFACES

A. DARVISHI, E. MUNTEANU, V. GUGGIANA, H. SCHAUER
Department of Computer Science (IfI),University of Zurich, Winterthurerstr. 190, CH-8057 Zurich, Switzerland

M. MOTAVALLI
Swiss Federal Labs for Material Testing and Research (EMPA),Uberlandstr. 129, CH-8600 Dubendorf, Switzerland

M. RAUTERBERG
Usability Laboratory, Work and Organizational Psychology Unit
Swiss Federal Institute of Technology (ETH), Nelkenstr. 11, CH-8092 Zurich, Switzerland

ABSTRACT

This paper describes work in progress on automatic generation of "impact sounds" based on *purely* physical modelling. These sounds can be used as non-speech audio presentation of objects and as interaction mechanisms to non visual interfaces. Different approaches for synthesizing impact sounds, the process of recording impact sounds and the analysis of impact sounds are introduced. A physical model for describing impact sounds "spherical objects hitting flat plates or beams" is presented. Some examples of impact sounds generated by mentioned physical model and comparison of spectra of real recorded sounds and model generated impact sounds (generated via physical modelling) are discussed. 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

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, different approaches for synthesizing impact sounds, a software architecture concept for automatic analysing and synthesizing of impact sounds and some implemented tools are described. A physical model, some implemented examples and their comparison with real impact sounds are presented. At the end, this paper describes some steps which are to be carried out and those which are supposed to be done by the end of the project.

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.

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 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
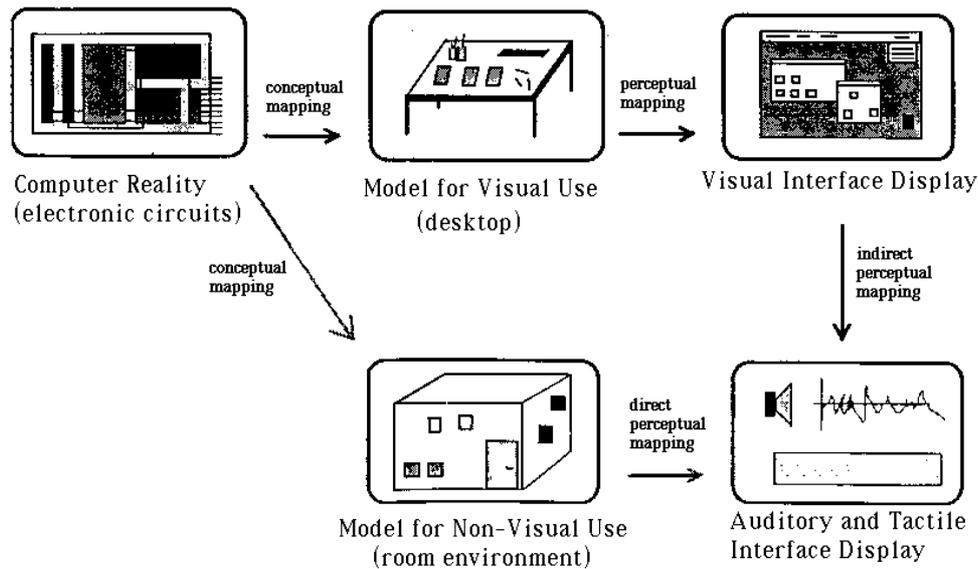
Fig. 1: Conceptual and perceptual mapping in user interface design

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 diagram (fig. 1) 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].

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.

-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.

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.

THE USE OF NON SPEECH AUDIO IN DIFFERENT COMPUTER APPLICATIONS

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].

DIFFERENT APPROACHES FOR SINTHESIZING IMPACT SOUNDS

There are two different approaches based on two different models for synthesizing impact sounds;
(1) empirical models, this approach is used by William Gaver in his work "Synthesizing Auditory Icons" [Gaver 93],
(2) purely physical models, this approach is used in our project to synthesize impact sounds by purely physical modelling.

RECORDING IMPACT SOUNDS FOR FURTHER ANALYSIS

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.
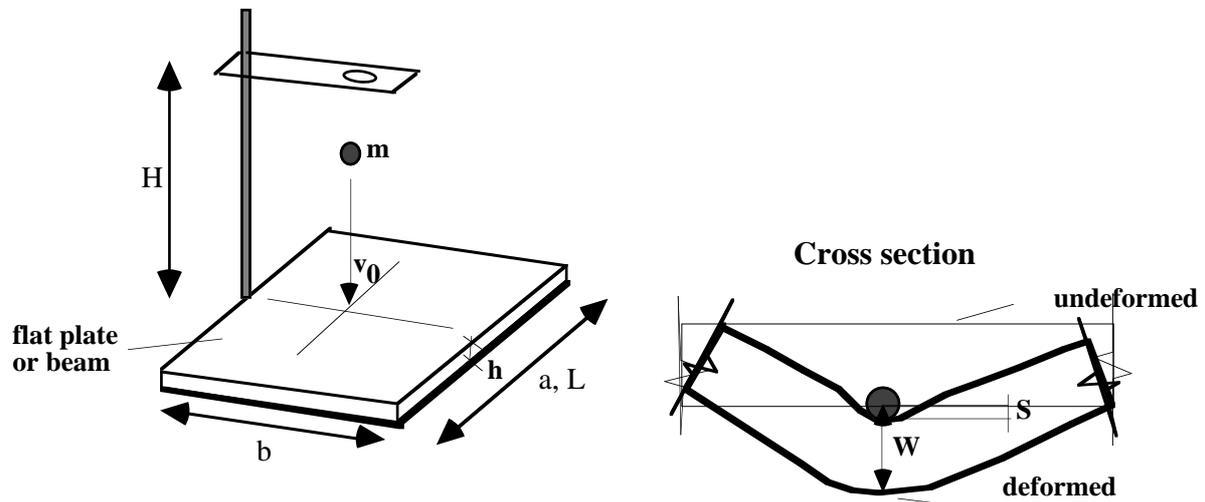

Fig. 2: experimental device

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

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

Basic Relations:

$$\omega = 2\pi f \qquad\qquad I = \frac{bh^3}{12}$$

$$D = \frac{Eh^3}{12(1-\sigma^2)} \qquad\qquad \rho' = \rho \cdot b \cdot h$$

Model of the Plate:

Equation of motion:

$$D\left(\frac{\partial^4 w(x,y,t)}{\partial x^4} + 2\frac{\partial^4 w(x,y,t)}{\partial x^2 \partial y^2} + \frac{\partial^4 w(x,y,t)}{\partial y^4}\right)$$
$$+\rho h\frac{\partial^2 w(x,y,t)}{\partial t^2} = 0 \qquad\qquad (1)$$

Solutions for the natural frequencies of plate 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

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

Now, we have a complete solution for impact sounds of small spheres onto flat plates. Every sound, that is caused by an impact of a small spherical object onto a flat plate of different material, can be automatically generated using equation (1). We have to specify the physical dimensions of the plate, the materials' properties, the boundary conditions (clamped along all edges or simply supported at all edges), the position where the plate was hit and the height from where the sphere falls.

In the case of beams, the equations become less complex:

Model of the Beam:

Equation of motion:

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

Solutions for natural frequencies of beam in two particular typical boundary conditions:

clamped at both ends: $\quad f_n = \frac{n^2 \pi h}{L^2}\sqrt{\frac{E}{3\rho}}$

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

SOFTWARE ANALYSIS OF IMPACT SOUNDS

The main purpose of the analysis is to extract the meaningful parameters from the recorded sounds. These are for instance: natural frequencies and their initial magnitude and damping constant. The analysis of the recorded signal through means of Fast Fourier Transform provides the natural frequencies in the audible domain. The spectrum is computed successively over a short temporal segment from the beginning of the signal. To determine the initial magnitude of spectral components we make use of the first spectrum (attack portion). The spectrum of next interval (decay portion) is analysed in the same manner. Consequently, the set of damping constants for all eigen-frequencies can be determined with the following relation:

$$\delta^{(k)} = \frac{-\ln\dfrac{A_t^{(k)}}{A_0^{(k)}}}{t}$$

where: t   time;
            k   number of natural frequency;
            $\delta^{(k)}$ damping constant;

$A_t^{(k)}$ magnitude at time t;

$A_0^{(k)}$ initial magnitude.

In order to minimize the error, we compute the damping constant at different moments. Then, averaging these values we get the value that we use in synthesis.

SOFTWARE GENERATION OF IMPACT SOUNDS WITH FILTER BANKS

This kind of approach has a good flexibility and lead to an efficient way to simulate a lot of complex interactions as "breaking", "bouncing", "rolling", "scrapping".

The filter bank consists of a set of bandpass filters (resonators) and can be build in two configurations: cascade (fig. 3 a) or parallel (fig. 3 b).

a)

b)
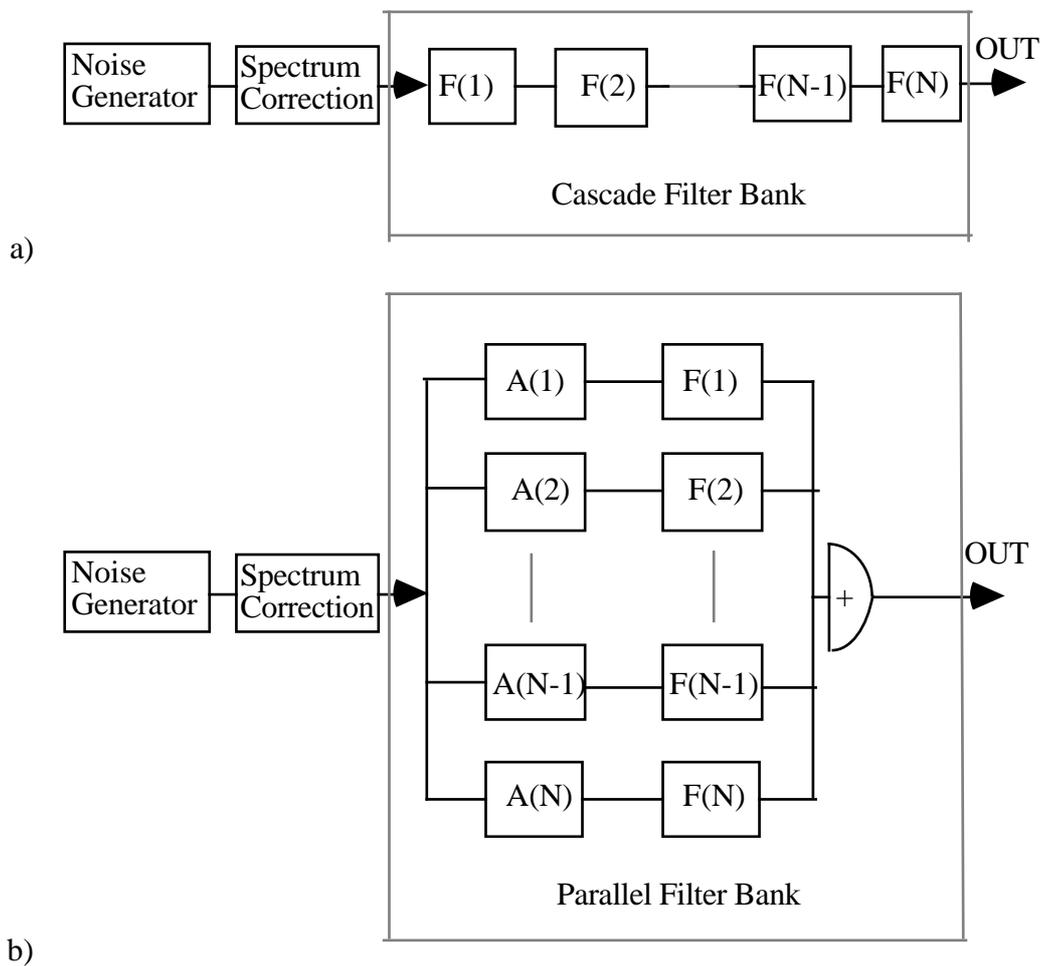
Fig. 3 Cascade (a) / Parallel (b) Synthesiser

The 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 bandwidth of each filter reflects the hardness of the material. 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.

The transfer function of one bandpass filter corresponding to the k-th natural frequency is:

$$H_k(z) = \frac{Y_k(z)}{X_k(z)} =$$

$$\frac{1 - 2e^{-2\pi B_k T}\cos(2\pi f_k T) + e^{-4\pi B_k T}}{1 - 2e^{-2\pi B_k T}\cos(2\pi f_k T)z^{-1} + e^{-4\pi B_k T}z^{-2}}$$

with: f_k = the k-th natural frequency

B_k = bandwidth of k-th filter

T = sampling rate

By applying the inverse z transform to the above equation, we obtain, the filtered signal (in discrete form). Samples of the output of such a resonator are computed from the input sequence $x[nT]$ by the recursive equation:

$$y[nT] =$$
$$C_k x[nT] + D_k y[(n-1)T] + E_k y[(n-2)T]$$

where $y[(n-1)T]$ and $y[(n-2)T]$ are the previous two sample values of the output sequence $y[nT]$. 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:

$$E_k = -e^{-2\pi B_k T}$$
$$D_k = 2e^{-\pi B_k T}\cos(2\pi f_k T)$$
$$C_k = 1 - D_k - E_k$$

COMPARISON BETWEEN REAL SOUNDS AND SYNTHESIZED SOUNDS

In order to find out how close are the sounds synthesized using the described physical model to the real ones, we compared them in two ways: 1) through means of spectra 2) hearing the two sounds repeatedly in order to find out whether human ear perceives a difference between them. The first way is illustrated below in the case of a steel sphere that hits a steel beam. The spectra for real and synthesized sounds are presented in fig. 4 and 5 and are very similar.
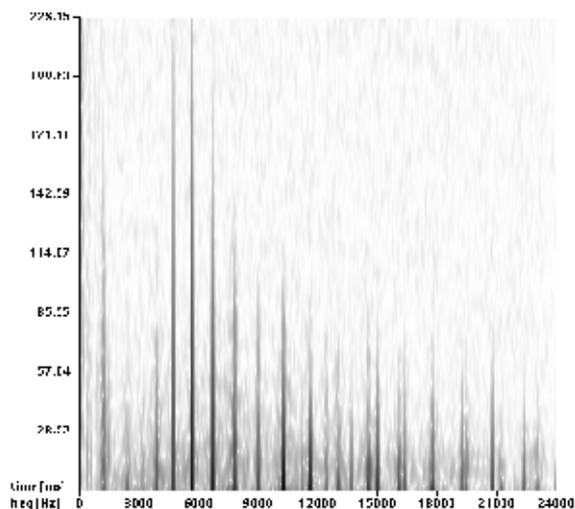
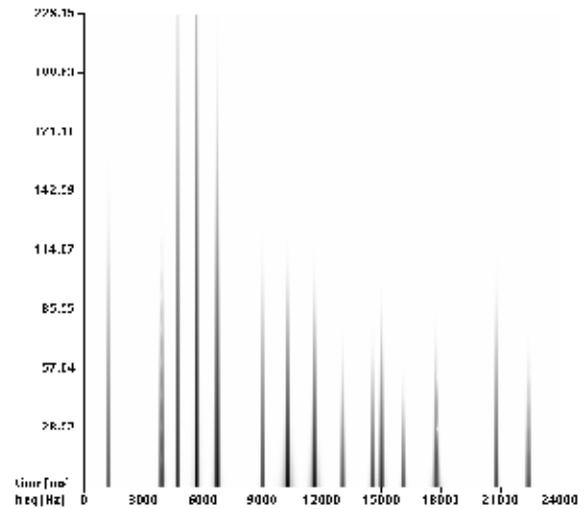

fig.4 Spectrum of the real sound

fig.5 Spectrum of the synthesized sound

A comparison in the second way was also made, using a developed test environment. With this application we have analysed whether subjects can distinguish between the real and synthesized sounds. The subjects were chosen arbitrarily and had no special musical experiences. The results were also very good.

CONCLUSION

Some problems and shortcomes for using GUIS for non-sighted computer users and two different approaches for adaptation of GUIS were presented. Different computer applications concerning use of non speech audio and two approaches for modelling impact sounds were discussed. Recording of impact sounds for further analysis, the analysis process, physical model for synthesizing impact sounds, software generation of impact sounds, some comparisons between model generated and real impact sounds were presented. The purely physical model for synthesizing simple impact sounds seems to be succesful. Further, more investigations are supposed to be done to examine the usability of purely physical models for more complex impact sounds. However, the implemented impact sounds are enough powerful to be used for presentation of objects and interactions in non visual interfaces. A prototype of such interface is planed for the next phase of the current project. The developed real-time synthesis algorithms need low 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 Non

speech 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: North-Holland.

[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).

[Gold-75] Gold, B., Rabiner, L. R. (1975) Theory and Application of Digital Signal Processing, Prentice-Hall, Inc. Englewood Cliffs, New Jersey.

[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.*