Automatic Generation of Audio Data in an Audio Framework

IFI TR 93.42

November 1993

Helmut Schauer, Alireza Darvishi

Department of Computer Science,

University of Zurich, Switzerland

darvishi@ifi.unizh.ch

Matthias Rauterberg

Work and Organizational Psychology Unit,

Swiss Federal Institute of Technology (ETH), Zurich, Switzerland

rauterberg@rzvax.ethz.ch

Massud Motavalli

Swiss Federal Laboratories for Testing, Material, and Research, Dübendorf/ZH, Switzerland

motavalli@empa.ch

Keywords: audio framework, auditive feedback, human - computer interaction, knowledge engineering, knowledge acquisition, non speech sound generation, object oriented programming, software engineering, software ergonomics, user interface management system.

Scientific Information

1. Summary of the Research Program

The objective of this research program is the development of concepts, 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.

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.

In the first place, a framework concept for the description of sounds is presented, in which sounds can be represented as auditive signal patterns along several descriptive dimensions of various objects interacting together in a certain environment. Based on 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 with respect to their physical condition (status 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 1.) occurs very frequently in everyday life, 2.) the interacting objects can be easily and well described by their material characteristics and 3.) the knowledge of solid-state physics can be utilised.

On the theoretical level, first of all a better overview and a better understanding of the capabilities, restrictions and problems of the existing instruments (tools) for the automatic generation of audio data can be anticipated. Out of the new concepts to be developed, essential impetuses for the following fields can result.

* CAD - architecture: new possibilities for the representation of sound insulating rooms, walls, materials etc. In general, the acoustics of buildings can be simulated and interactively modified.

* Materials research: new possibilities for the interactive representation of hidden and thus only inadequately visually representable materials characteristics.

* Multimedia: new possibilities for the interactive representation of complex sound generating events and processes.

* User interfaces for people with impaired vision: simulations with the utilisation of audio data will in future also have their application in the training of people with impaired senses, in particular of people with damaged vision.

2. Research Plan

2.1. The State of Research

Many computers have a sound-generator which produces a simple beep-sound to indicate the errors. For a long time the only audio-information for user-interfaces contemporary modern workstations have signal-processors and analogue-digital converters and therefore sounds can be used in software.

2.1.1. Definition of Basic Terms

The terms which are used later in this chapter are defined as follows:

Audio signal pattern:	description of all perceptible audio signals.
Speech:	the description of all audio signals which have describable grammar structures.
Music:	complex audio signal pattern which has rhythmic describable structures.
Tone:	simple audio signal pattern with rhythmic describable structure.
Everyday-sounds:	audio signal patterns which have not been sufficiently researched to give a description and creation structure.

We call the combined group of everyday-sounds and tones "sounds". When we mean the assignment of a sound to an event we put the event-identifier in inverted commas (for example: event = impact, sound="impact").

2.1.2. Comparison of Auditory and Visual Signal Pattern

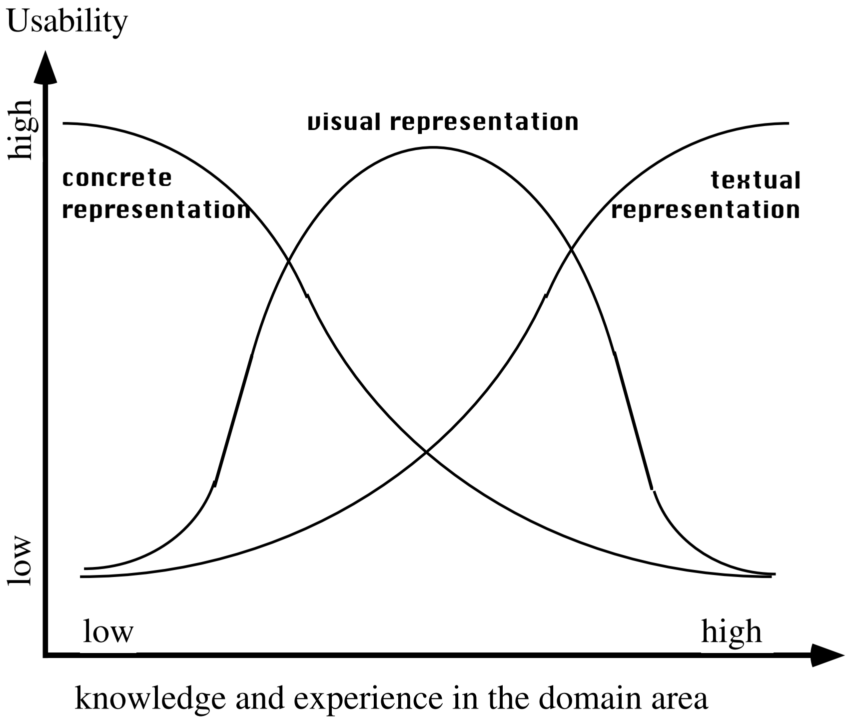

Figure 1 When to use different representation media [Marmolin-92:47].

The textual representation of information is of most use when the user is familiar with the domain area and can demonstrate much experience and knowledge in that domain area. In comparison, more concrete(visual and auditory) representations of information which the user can query are of most use when the domain area is new and unknown. By comparing audio signal patterns with visual signal patterns the different advantages of each can be shown (see Table 1).

Table 1 Comparison of auditory and visual signal patterns in time-space continuum [Gaver-89:71].

	Time Dimension	**Space Dimension**
auditory	Auditory signal patterns exists **in** time. • They are well suited for changing events, • accessible only for a limited time span.	Auditory signal patters exists **over** space. • They can be "processed" and do not require visual contact, • only a limited number can be processed at the same time.
visual	Visual signal patterns exist **over** time. • They are well suited for static objects, • can be sampled over time.	Visual signal patterns exist **in** space. • They require active eye contact, • messages can be spatially distributed.

Sounds and music can be utilised to improve the user's understanding of visual predecessors or can stand alone as independent sources of information. (For example: sounds as diagnostic support applied with the direction of a process simulation [Gaver-91]).

The parallel use of different media and the resulting parallel distribution of information, for example by simultaneously showing a predecessor through a concrete representation and its explanation through audio distribution, leads to a denser sharing of information. In this case, the user can dedicate his attention solely to the visual information, which has parallel audio support. This reduces the need to change textual or other visual delivery and prevents the overflow of visual information [Gaver-91]. The redundancy of information represented visually and auditory, as long as the representation of the information is realistically formed, is sensed not as disturbing, but instead it demands and increases information reception. It is important that with simultaneous information representation, that the information is harmonised together and that the different media are well synchronised.

[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.

[Marmolin-92] Marmolin, H. (1992) Multimedia from the perspective of psychology. in L. Kjelldahl (ed.) Multimedia: systems, interaction, and applications. Berlin, Heidelberg: Springer.

2.1.3. Everyday Sound Perception

The perception of auditory signal patterns in everyday life can come in very different forms: a car driving by on the street, a dripping faucet, the confusion of voices from a crowd of people, opera music, a plane flying by, the buzz of a travel alarm clock, the beeping of a wristwatch, etc.

All of these auditory signal patterns are divided into 4 categories: speech, music, sound and noise; sometimes noises and sounds are heard and grouped together. All of these categories are described sufficiently in the physical world through the mixing and superposition of different pitches, frequencies, volumes and sound duration. One of the essential differences between these categories, however, lies in their semantics:

Speech serves primarily to convey information, while music and noise can have a pleasant or an unpleasant influence on the emotions. For musicians and other people who are intimate with this area, music and noise have a comparable semantic and informative character as speech does for the normal citizen. Aside from music and noise, the listener is interested next in the possibility of undisturbed, context-free perception.

In contrast to music and noise, everyday sounds have a self-standing characteristic; they are extremely context sensitive and event related [Gaver-86]. Through the physical interaction of different everyday objects in 3-D space the sounds of everyday life are created and through propagation they become audible through the air. In comparison with music and noises, the semantically relevant dimension of sound lies not with the characteristic quality of the auditory signal pattern itself, but rather with the quality of the sound producing event as it respects the concerned object [Mountford-90].

This difference leads us to the conclusion that sounds are interpreted differently than music based upon their quality. When listening to music we are primarily interested in the effect of music on us; while when hearing everyday sounds we are interested in the quality of the sound producing object and the accompanying circumstances (i.e.. surrounding conditions, events, etc.). Of course, music can be heard from the perspective of every day use; in this case the listener pays attention to the nature and tune of the instrument in use, to the tempo, to the acoustic, to the place of performance, etc. This method of listening to music is dependent upon the listener's knowledge of this domain field; only someone who is experienced with music will be able to extract all of the various information from a piece of music.

The average adult is, for the interpretation of everyday sounds, an expert with a large degree of knowledge from experience. This knowledge allows one to evaluate everyday sounds according to the following criteria for relevant information:

1. Information about the physical occurrence: we hear, if the fallen glass clinks or breaks.
2. Information about unseen structures: when knocking on the wall, we can hear if it is hollow.
3. Information about dynamic changes: when filling a wine glass, we can hear when it is full and runs over.
4. Information about abnormal conditions: we hear, when the car engine ceases to function properly and runs irregularly.
5. Information about occurrences outside of the visual field: the sound of footsteps behind us "tells" us if someone is approaching.

Listening to everyday sounds is based upon the perception of events and not upon the perception of sounds in and of themselves. This fact becomes clear in the following example:

Illustrative example:

A pen dropped upon a piece of paper from a height of about 15 cm created a different sound than when it is dropped upon the hard surface of a desk. An altogether different sound is created when a rubber eraser is dropped upon the paper or, respectively, on the desk.

The sound created in each case of the previous example is neither a characteristic of any of the participating objects (pen, rubber eraser, sheet of paper, desk surface) nor a characteristic of the occurrence "dropped" itself. The four different sounds in the examples are, in terms of an observation that holds true to the reality of the situation solely determined by their respective interaction and environmental conditions. Everyday sounds are therefore due to a lack of better descriptive possibilities, often described through the underlying occurrence (see Table 2).

Table 2 The nine possible combinations for sound-creating object interactions (Surroundings := air, propagation medium := "air")

Object type	Ol: solid	Ol: liquid	Ol: gas
Os: solid	"bells sounding" clapper ->bell	"stirring" wish->cream	"archery" arrow->air
O_S:liquid	"pouring" coffee->cup	"spray" stream->pond	"raining" drop->air
Os: gas	"rustling" wind->leaves	"boiling water" air bubbles->water	"exploding" balloon->air

[Legend: Object (small, moving fast) "O_S" interacts "->" with Object (large, still, slow) "O_l"]

Every sound is also a result of one or more interactions between two or more objects in a definite place and in definite surroundings and can be defined as the following:

Sound = f(Interaction(s), objects, surroundings, place, time)

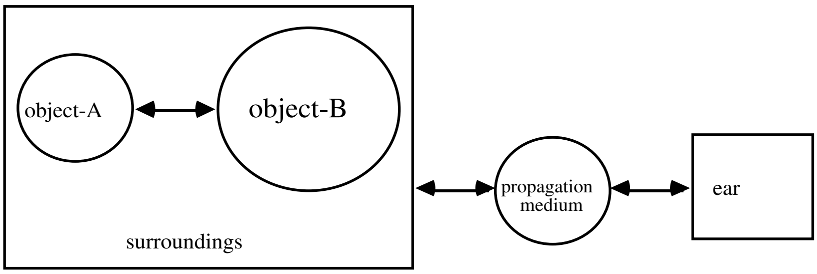

Figure 2 Representation of all primary shared components with the production of everyday sounds

Every interaction possesses attributes which have an influence on the produced sound. At the same time the shared objects can participate in the production of sound from different aggregate conditions, materials, and even their configuration. The configurations of these materials possess attributes that also can have an influence on the produced sound (Fig. 2).

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

[Mountford-90] Mountford, S. & Gaver, W. (1990) Talking and listening to computers. in B. Laurel & S. Mountford (eds.) The Art of Human-Computer Interface Design. (pp. 319-334), Reading MA: Addison-Wesley.

2.1.4. Existing Research Projects and Results

With the manufacture and adaptation in the world of interactive computer systems (i.e.. multimedia applications) one must decide between the following three types of substitution:

1. *Propagation*: here the quality of the propagation channel is important (i.e. bandwidth, etc.);
2. *Intake, saving, and reproduction*: this is actually time synchronous with visual processing to add extra significance;
3. *Automatic Productio*n: in this area the realistic generation of context sensitive, auditory signal patterns has the most significance.

We limit ourselves in this research proposal to the third area: the automatic production of sounds. The existing research estimates for audio estimates with interactive systems will be listed.

The computer manufacturer, Apple, researches the substitution of audio media for improvement of graphical user interface and presents the so-called "Auditory Icons" such as "Earcons" [Gaver-86] [Gaver-89] [Gaver-90] [Blattner-89].

The Bell Communications Research Centre in collaboration with the Massachusetts Institute of Technology developed an "Audio-Windowsystem" as an analogue of the "Visual Windowsystem" which realised one, two, and three dimensional representations of auditory source [Ludwig-90] [Wenzel-91].

The Olivetti Research Center was developing a so-called "Audio-Server" an analogy to the "Window-Server" with a user interface based on the Client-Server Model [Binding-90].

None of the three projects mentioned above, however, allow the description of sounds on a highly abstract semantic level.

[Binding-90] Binding, C., Schmandt, C., Lantz, K. & Arons, B. (1990) Workstation audio and window-based graphics: similarities and differences. in G. Cockton (ed.) Engineering for Human-Computer Interaction. (pp. 233-247), Amsterdam: North-Holland.

[Blattner-89] Blattner, M. et al. (1989) Earcons and Icons: their structure and common design principles. Human Computer Interaction, 4:11-44.

[Gaver-86] Gaver, W.(1986) Auditory Icons: Using Sound in Computer Interfaces. Human Computer Interaction, 2:167-177.

[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 et al. (eds.) Human Computer Interaction INTERACT'90 (pp. 735-740). Amsterdam: Elsevier.

[Gaver-91] Gaver, W.(1991) The SonicFinder: An Interface That Uses Auditory Icons. Human Computer Interaction, 4:67-94.

[Ludwig-90] Ludwig L., Pincever, N. & Cohen, M. (1990) Extending the notion of a window system to audio. Computer 23(8):66-72.

[Smith-90] Smith, S., Bergeron R. & Grinstein, G. (1990) Stereophonic and surface sound for exploratory data analysis. in Proceedings of Computer-Human-Interaction CHI'90. New York: ACM.

[Wenzel-91] Wenzel E., Wightman F. & Kistler D. (1991) Localization with non-individualized virtual acoustic display cues. in S. Robertson, G. Olson & J. Olson (eds.), Reaching through technology CHI'91. (pp. 351-359), Reading MA: Addison-Wesley.

2.2. The State of Own Research

It is mainly the following factors in their respective forms and combinations which influence the load for humans in the man-computer interaction: the organization of the working-place and the working-place environment, the content of the tasks on the screen, the temporal share of the activity on the screen in proportion to the whole activity, the content of the computer-support and the user-friendliness of the interactive system. Our investigations have shown that the user-friendliness of a data-processing system crucially influences its acceptance by its users. Three different aspects should hereby be discriminated: (1) the functionality and the amount of information, (2) the availability (response-times and disturbances) and (3) the operational handling mediated by the user-interface [Rauterberg-91a] [Rauterberg-91b] [Rauterberg-91c] [Rauterberg-92a] [Rauterberg-92b] [Rauterberg-93].

The results of an experiment for the investigation of the effects of audio-feedback showed, that the results of a database-query at the user-interface with an individually selectable acoustic feedback were as good as with a previously adjusted standard interface without any acoustic feedback. However, if the users who considered the acoustic feedback as useful are compared with those who considered it not necessary, significant performance-advantages result for the persons with a positive mind about the acoustic feedback.[Rauterberg-91d]

[Rauterberg-91a] Rauterberg, M. (1991). Benutzungsorientierte Benchmark-Tests: eine Methode zur Benutzerbeteiligung bei Standardsoftware-Entwicklungen. In: D. Ackermann & E. Ulich (eds.), German Chapter of the ACM Berichte Nr. 32 "Software-Ergonomie'91", S. 96-107, Stuttgart: Teubner.

[Rauterberg-91b] Rauterberg, M. (1991) Benutzungsorientierte Benchmark-Tests: Anwendung auf die Gestaltung der Benutzungsoberfläche von ADIMENS GT hin zu GT+. In: M. Rauterberg & E. Ulich (eds.), Posterband zur Software-Ergonomie'91 (S. 151-162). Zürich: ETH, Institut für Arbeitspsychologie.

[Rauterberg-91c] Rauterberg, M. (1991) From passive to active joins: A task oriented desktop interface conception for a relational DataBase Management-System. Proceedings of the Conference "Intelligent Text and Information Handling RIAO 91" Barcelona (Spain) - April 2-5,1991. Vol. 2 (pp. 809-821).

[Rauterberg-91d] Rauterberg M., Baumgartner A., Jenny A. & de Lange M. (1991) Erhöhung der Transparenz durch individualisierte Soundunterstützung bei Datenbankrecherchen. Technical Report IS-1-91. Institut für Arbeitspsychologie. Zürich: ETH.

[Rauterberg-92a] Rauterberg, M. (1992) Designing and testing the usability of a relational data base system with end-user oriented benchmark tasks. In: P. Brödner & W. Karwowski (eds.), Ergonomics of Hybrid Automated System III (pp. 301-306). Amsterdam London Tokyo: Elsevier.

[Rauterberg-93] Rauterberg, M. (1993, in press) A product oriented approach to quantify usability attributes and the interactive quality of user interfaces. In: H. Luczak, A. Cakir & G. Cakir (eds.), Work With Display Units (WWDU). Amsterdam: Elsevier.

2.3. The Research Plan in Detail

2.3.1. Analysis of the Existing Models and Concepts

First of all, the contemporary concepts and models in the field of structural mechanics have to be analysed and investigated for their appropriateness for this research-project (e.g. [Dual-88] and [Koller-83]).

In order to gain the knowledge from sound-specialists, the following "knowledge-engineering" methods are taken into consideration[Welbank-90]:

1. interviews with specialists in the field of film and radio;
2. observation of sound-engineers while they are producing sounds with solid objects (e.g. coconut-halves for the simulation of the clatter of hooves)
3. multidimensional scaling of noises.

[Dual-88] Dual, J. (1989) Experimental Methods in Wave Propagation in Solids and dynamic Wiscometry. Dissertation Nr. 8659, Zürich: ETH.

[Koller-83] Koller, M. G. (1983) Elastischer Stoss von Kugeln auf dicke Platten. Dissertation Nr. 7299, Zürich: ETH.

[Welbank-90] Welbank, M. (1990) An overview of knowledge acquisition methods. Interacting with Computers. 2(1):83-91.

2.3.2. Definition of the Requirement

From the analysis of existing models and concepts as well as empirical investigations of the pre-phase a definition for the requirements for a new concept and a new model should be derived.

2.3.3. Object-Oriented Design

The example of the impact of a falling ball on the surface of a plate shows the physical process of producing sound through the interaction of the objects ball and plate (see Fig. 3).

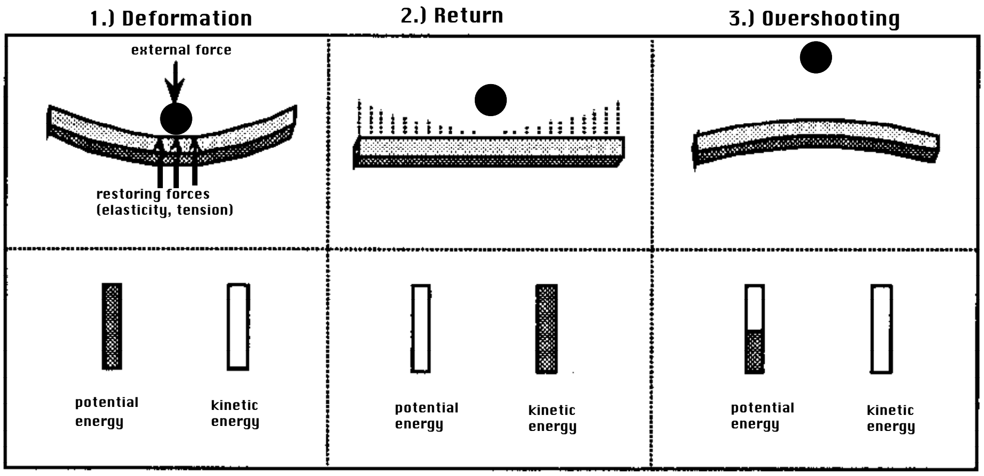

Figure 3 The relation ship between potential and kinetic energy when a solid plat is deformed by external forces [Gaver-93].

The produced sound "ball hits the plate" can be described by the behaviour of the vibration of the surface, which has been produced by the deformation resulting from the fall of the ball. "When an object is deformed by an external force, internal restoring forces cause a build up potential energy (1). When the external force is removed, the object's potential energy is transformed to kinetic energy, and it swings through its original position (2). The object continues to vibrate until the initial input of energy is lost (3)" [Gaver-93]. The following presentation should help us to find a classification to describe the sounds on a very high abstract and semantic level.

The following diagram (Fig. 4) shows a global classification of everyday sounds regarding to the aggregation state of the interacting objects.

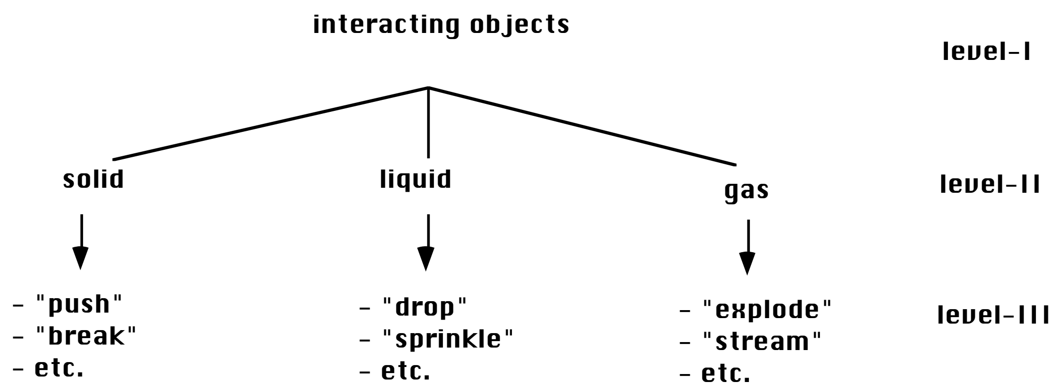

Figure 4 Sound classification of objects regarding to their aggregation state.

The highest level in Figure 4 shows the class of everyday-sounds which are produced via interaction of two or more objects. The second highest level shows the three different material states which produce qualitatively diverse sounds via the interaction of two or more solid-objects produce sounds like "push", "break". Sounds like "drop", "sprinkle" are produced via interaction of different liquid objects, and sounds like "explode", "stream" are produced via aero-dynamic interactions in the air. All of these so-called "primitive-sounds" are characteristic f

for their classes. The described sound-classification is very similar to the object oriented soft-ware development. In fact the object oriented concepts provide good and powerful tools in designing the above-mentioned sound-classification. Object-oriented soft-ware concepts are appropriate in many cases for soft-ware-engineering paradigms.

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

2.3.4. Audio-Framework

The term "framework" comes from the object oriented world and is understood as follows:

already programmed-frames which allow the programmers to supplement the application of specific parts in a program.

An audio-framework should in the term of object oriented terminology provide the basic-sound classes and the possibilities to combine and manipulate them. At the moment there is no such audio-framework. As an example we are going to approach the sound-class "impacts".

Our approach:

Our methodology can be demonstrated by dropping a ball onto a flat plate and later onto a non-flat plate. The four following steps are discussed:

Step one: calculation of contact-force f(t) . It can be shown that the force f(t) can be derived from a differential equation [Koller-83]. The calculated contact-force is a function of material-features (for example density, elasticity-module and mass), geometry of contact-objects (for example the thickness of the ball, radius of the ball) and boundary conditions (for example the initial speed).

Step two: the calculation of deformation-function w(t); from so-called plate differential-equation and via some mathematical operations.

Step three: Fourier-transformation of the function w(t); finally this function can be transformed via the Fourier method: Different frequencies can be produced as sounds [Ackermann-91].

Using non-flat plates can cause complex mechanical-equations. For their solution they must be numerically integrated[Sayir-92].

[Ackermann-91] Ackermann, P. (1991) Computer und Musik, eine Einführung in die digitale Klang- und Musikverarbeitung. Berlin: Springer.

[Dong-62] Dong, S. B., Pister, K. S. & Taylor, R. L. (1962) On the theorie of Laminated on Isotropic shells and plates. Journal of Aerospace Science, 29:969-975.

[Sayir-92] Sayir, M. B., Häusler, K. & Partel, M. (1992) Impact Behavior of corrugated sheets of fibre-reinforced cement. Composites, 23(6).

2.3.5. Developing a Prototype

Modern workstations like "NEXT" provide sound and music kits. They also provide so-called basic operations for manipulation of sound-patterns in the frequency-level. In addition they offer the opportunity to assign a digitizied sound to a certain dialogue-element , for example a button. All of these approaches assign an unchangeable sound to a object or operation. These sounds are very synthetic. It means contexfree and does not fulfil our expectation of the real world. Instead of that we understand a sound as a result of one or more interactions between two or more objects in a certain place and a certain environment. It means that a sound can convey more information about the sound-source, it's place and environment. Therefore it should be calculated in real-time and is context-sensitive.

2.4. The Importance of the Research Project

On the theoretical level a better overview and understanding of the skills and restrictions of existing tools for automatic generation of audio-data are expected. The new concepts which are planned for development will give a basic impulse to the following fields:

CAD: Architecture:

New possibilities of examining the sound-isolation of rooms, walls, materials etc. Generally, the acoustics of buildings could be changed and checked interactively.

Material-research:

New possibilities for hidden and unseen material-features.

Multimedia:

New possibilities for interactive creation of sound producing events and actions.

Simulation: computer support learning

A pedagogic target of computer support learning will clarify the processes. These processes could be physical or chemical processes.

User interfaces for the visually impaired:

Simulations also find their applications in the future for handicapped and especially visually impaired computer users, e. g. blind computer users could recognise the dialogue-objects and movable-pictures in a 3-D space via sounds in addition to the synthesisers.