

## Acoustic Chase:

Designing an interactive audio environment  
to stimulate human body movement

by Simon Karl Josef Schiessl  
Dipl. Ing. Physikalische Ingenieurwissenschaften (2001)  
Diplommedienberater (1998)  
Technische Universität Berlin

Submitted to the Program in Media Arts and Sciences, School of Architecture  
and Planning in Partial Fulfillment of the Requirements for the Degree of Master of  
Science in Media Arts and Sciences

at the  
Massachusetts Institute of Technology

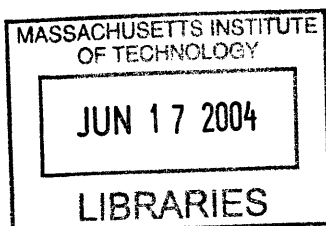
June 2004

© 2004 Massachusetts Institute of Technology  
All rights reserved

.....  
Signature of Author  
May 12, 2004

.....  
Thesis Advisor  
Chris Csikszentmihályi  
Assistant Professor of Media Arts and Sciences

.....  
Accepted by  
Andrew B. Lippman  
Chair, Departmental Committee on Graduate Studies  
Program in Media Arts and Sciences



ROTCH



# **Acoustic Chase:**

## **Designing an Interactive Audio Environment to Stimulate Human Body Movement**

by Simon Karl Josef Schiessl

Submitted to the Program in Media Arts and Sciences, School of Architecture and Planning in Partial Fulfillment of the Requirements for the Degree of Master of Science in Media Arts and Sciences.

### **ABSTRACT**

An immersive audio environment was created that explores how humans react to commands imposed by a machine generating its acoustic stimuli on the basis of tracked body movement. In this environment, different states of human and machine action are understood as a balance of power that moves back and forth between the apparatus and the human being. This system is based on spatial sounds that are designed to stimulate body movements. The physical set-up consists of headphones with attached sensors to pick up the movements of the head. Mathematic models calculate the behavior of the sound, its virtual motion path relative to the person, and how it changes over time.

Thesis Supervisor: Chris Csikszentmihályi  
Assistant Professor of Media Arts and Sciences



**Acoustic Chase:**

Designing an interactive audio environment  
to stimulate human body movement

by Simon Karl Josef Schiessl  
Dipl. Ing. Physikalische Ingenieurwissenschaften (2001)  
Diplommedienberater (1998)  
Technische Universität Berlin

Submitted to the Program in Media Arts and Sciences, School of Architecture  
and Planning in Partial Fulfillment of the Requirements for the Degree of Master of  
Science in Media Arts and Sciences

at the  
Massachusetts Institute of Technology

June 2004

.....|.....  
Thesis Reader  
John Maeda  
Associate Professor of Design and Computation

.....|.....  
Thesis Reader  
Joe Paradiso  
Associate Professor of Media Arts and Science



## Table of Contents

1. Introduction .....	9
2. Technical Platform .....	13
2.1 Motion Sensor.....	14
2.1.1 Overview Tracking Technology .....	14
2.1.2 Inertial Motion Sensing .....	18
2.1.2.1 6 Axis System.....	19
2.1.2.2 3 Axis System.....	21
2.1.3 Signal Processing.....	22
2.1.3.1 Spatial Geometry .....	23
2.1.3.2 Filtering .....	28
2.2 Sound Generation .....	30
2.2.1 3D Audio .....	32
2.2.1.1 Psychoacoustic Principles.....	32
2.2.1.2 Displaying Spatial Audio.....	33
2.2.2 Direct Sound 3D .....	35
2.3 Positioning Sounds .....	37
3. Sound Implementation.....	40
4. Placement Within the Artistic Context.....	41
4.1 Appearance of the Installation .....	41
4.2 Individual and Machine .....	43
4.2.1 Mutual Influence.....	43
4.2.2 Controlling And Being Controlled by Machines.....	45

4.3 The Use of Sound From a Media Perspective .....	46
4.3.1 Simplified And Abstract Sounds .....	46
4.3.2 Aleatoric Processes .....	48
4.3.3 Loops .....	48
5. Evaluation – Opinion of Artists and Curators .....	50
6. Future Work.....	53
6.1 Gesture Recognition .....	53
6.2 Finite States .....	54
7. Conclusion .....	57
8. References .....	58



## 1. Introduction

Technical devices are generally considered to be compliant instruments that are fully controllable tools. From this perspective, humans have total power to operate machines at will. However, technologies often act back on humans, reversing the intended command direction. First of all, there is the inherent danger of malfunctions and accidents. Besides the obvious consequences of an accident, the possibility of dysfunctions alone requires constant vigilance. Driving a car, for example, demands the driver's full attention, and partly subordinates him to the technical device.

Coercions like this can be found on different scales and in varying degrees of directness. The driving example describes a person and a technical object in a confined one-to-one situation. But as technical objects grow to systems, their interactions become manifold and complex. To continue the car example to less oblique effects, the need for fuel has huge political and environmental implications, even causing war.

Interactive art pieces are usually designed to be compliant systems that wait for user input, and are often are designed for strictly limited and controlled output. This can be understood from their evolution from information systems and data terminals, and their tradition of data storage and retrieval. Even with such minimal functionality, interactive artwork goes far beyond traditional painting and sculpture, which naturally produces passive objects for appreciation. By leaving simple interaction schemes behind, technological art has the potential of an inner animus that develops its own intentions and can even escape control. So, human "linkage with the machine is never free from anxiety or the structures of domination."<sup>1</sup> This issue has been reflected in a variety of art pieces, covering topics such as body and

---

<sup>1</sup> R. Frieling, Reality/Mediality Hybrid Processes Between Art and Life  
([http://www.medienkunstnetz.de/themes/overview\\_of\\_media\\_art/performance/23/](http://www.medienkunstnetz.de/themes/overview_of_media_art/performance/23/))

machines,<sup>2</sup> blind spots of technology from a social point of view,<sup>3</sup> or technology purposely used for controlling humans.<sup>4</sup>

Acoustic Chase looks at the give and take of man-made systems by establishing an interactive artwork with the power to actively impel the human being. It generates man-machine interactions intentionally provided with degrees of freedom that can't be fully controlled by people. This example of a machine with "free" will illustrate how technical systems have the potential to leave their role as compliant tools and develop their own agency. It uses sound and motion tracking as the main interaction components, in order to establish a mutual relationship of power in which the human being is partially controlled by a device, which in turn is controlled by the human.

This paper describes the technical platform for that system as well as different types of embedded acoustic stimuli. As the final system is designed to provide a coherent aesthetic experience as a piece of art, its expressiveness and place within this context is also examined.

Acoustic Chase expands on ideas that I explored in "Haptic Opposition," a previous project, which combined touch and vision to create a human machine interaction in which the apparatus physically pushed back against human motion. There, the concept of acting back was taken very literally, and expressed in the form of a handle that was forcefully driven by the machine. The system focused on a physical contact with a real object, and an interaction designed to grow into chaotic fight.

Haptic Opposition had a clear object-like character, being a curious machine that people could touch and physically fight with. It manifested itself as a very real entity in space, and the fight with it became hard work. Acoustic Chase directs the viewer's

---

<sup>2</sup> Cp. Ars Electronica topics "Flesh Factor" (1997), "Genetic Life" (1993), "Life Science" (1999), or "Next Sex" (2000)

<sup>3</sup> Cp. Judith Butler "Cyborg Bodies", Chris Csikszentmihalyi "Afghan Explorer", Natalie Jeremijenko-Bureau of Inverse Technologies "Suicide Box"

<sup>4</sup> Cp. CTRL[Space] – Rhetorics of Surveillance from Bentham to Big Brother, ZKM 2002

attention to the here and now, and abandons the use of expressive media as a window to medial worlds.

The main technical components of Acoustic Chase are headphones, an inertial tracking unit attached to them, and a system for sound generation (see Figure 1). The sound control is designed to increase the acoustic stimuli until specific head motions are tracked.



**Figure 1: headphones with attached IMU**

By switching from haptic feedback to sound, the quality of interaction changes significantly. First, the object-character that is an essential part of Haptic Opposition vanishes. Sound displayed by headphones is closely and directly applied to the ears, resulting in “phantom” locations where it seems to come from. These may be undefined, global directions for regular sound, or from virtual spots in space for 3D spatialized sound. Thus, the physical object is replaced by disembodied sounds, and the user interaction volatilizes from a defined spot, the handle, and diffuses into the whole space around the user. In Acoustic Chase, the clear distinction between the object and the observer dissolves into partial immersion.

Though less direct than tangible force, sound can also be used to push and impel persons. Volume is a basic quality on an audio “push” that compares to the strength of a physical push. Audio signals, verbal and non-verbal, are a familiar method of commanding often used in teaching, interpersonal relationships, or military applications. They are difficult to escape: people can’t easily “hear away” or “close their ears.” However, sound by itself leaves much room for imagination. This characteristic fits nicely with the often oblique and indirect ways humans are subject to diffuse pressures from their environment.

Acoustic Chase does not use sound in a musical way but rather generates a collage of prerecorded snippets. This media aspect adds an additional layer of expression into the system: By selecting the appropriate sound material, Acoustic Chase adapts its core idea of a balanced system of power to different fields: the content is not limited to talking about technology itself, but can draw analogies to other forms of coercion, such as addiction or peer pressure.

Head tracking provides an intuitive interface for moving through the soundscape. By recording the head position, which, of course, tracks the ears as well, the virtual audio scene can react to human motions in a natural way. Physical objects have been replaced by disembodied sounds, so there is no need for any manual interface. Instead, the human subject is defenselessly exposed to the sound. And since the acoustic objects are invisible, they are untouchable as well.

## 2. Technical Platform

The basic elements of the technical platform are the motion tracking, sound generation, and control unit. The motion sensor tracks the movements of the head and therefore the position of the ears. This information will be used to localize the person within the virtual audio environment, and as input variables for the higher-level man-machine interaction. The audio generating unit plays back stored audio samples using mixers and filters, allowing effects such as putting a virtual sound source to a specific geometric spot in the space around the listener. Headphones are used as output to the user from the sound generator. Based on the tracked motion of the person, mathematically described models control the behavior of the sounds.

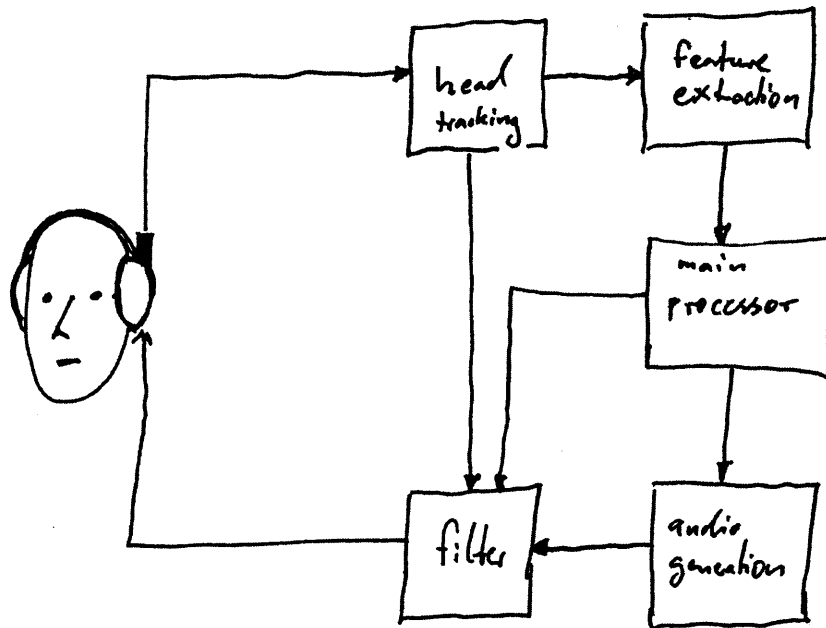


Figure 2: basic elements of the technical platform

## **2.1 Motion Sensor**

This project focuses on generating an audio environment that interacts with the person in two ways: considered as a geometric space, the user's movements inside this environment lead to quasi natural relocation of the virtual sound sources relative to the person's head. On a higher level, the tracked motions influence the animation of the sound sources described by specific behavior patterns. All this is done by tracking the movements of the person's head. This equals the main tracking task in VR environments, and a number of solutions have been developed.

### **2.1.1 Overview Tracking Technology**

Well-established principles are mechanical, acoustic, optical, magnetic, and inertial technologies. Youngblut et al. give an overview of available tracking systems.<sup>5</sup>

Mechanical systems use joint linkages with attached encoders that directly bridge the remote system with a fixed reference frame. The defined geometry in conjunction with high accuracy position encoders provides a very precise tracking result. Their main disadvantage, the subject being tethered by a cumbersome system of rods, is inherent to the principle. Mechanical tracking has been replaced by a variety of contact-free sensing technologies.

Optical systems generally watch the scene through one or more cameras and estimate the orientation and position of the object of interest by image analysis. Many image processors thereby rely on tags that have to be mounted on the tracked

---

<sup>5</sup> C. Youngblut, R. E. Johnson, S. H. Nash, R. A. Wienclaw, and C. A. Will, Review of Virtual Environment Interface Technology, Institute for Defense Analyses – IDA, Paper P-3186, March 1996

object. This approach has gained a lot of interest in research<sup>6</sup>, but has not had a major impact on commercial products so far<sup>7</sup>.

Several methods have been successfully applied to replace the linkages used by a mechanical probe head, and to measure the geometric relationship between the body frame, which is attached to the moving object, and the fixed reference frame.

Acoustic systems gauge the running time between one or several ultrasonic emitters to one or several receivers. This results in a mesh of straight lines of distances, which can be resolved to the relative position and – with enough emitters and receivers – also to orientation, covering a space up to 15 ft. Their main weaknesses are the sensitivity to environmental conditions, like wind and noise interference, and the difficulty to radiate and receive the waves throughout all directions. The ultrasonic head tracker from Logitech, for example, only allows movements within a 100-degree cone and 5 ft along linear axis to ensure sound contact between the source and the receiver.<sup>8</sup>

By generating varying magnetic fields and sensing their relative strength at the moving probe, precise contact-less sensing within a confined space may be accomplished. These may be resolved to relative position and orientation. With a

---

<sup>6</sup> Cp.: M. La Cascia, S. Sclaro, and V. Athitsos, Fast, reliable head tracking under varying illumination: An approach based on registration of texture-mapped 3D models, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(4): pp. 322–336, April 1999.

J. Rekimoto. A vision-based head tracker for fish tank virtual reality: VR without head gear. In *Virtual Reality Annual International Symposium (VRAIS '95)*, pp. 94–100, 1995.

Z. Zivkovic and F. van der Heijden, A stabilized adaptive appearance changes model for 3D head tracking. In *IEEE ICCV Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems (RATFG-RTS'01)*, pp. 175–182, 2001.

L.-P. Morency, A. Rahimi, N. Checka, and T. Darrell. Fast stereo-based head tracking for interactive environments. In *Proceedings of Conference on Automatic Face and Gesture Recognition*, pp. 375–380, 2002.

D.O. Gorodnichy, S. Malik, and G. Roth. Affordable 3D face tracking using projective vision. In *Proceedings of International Conference on Vision Interface (VI'2002)*, pp. 383–390, 2002.

<sup>7</sup> One of the rare available products is the marker based system DynaSight that tracks the head position in front of a camera, limited in space of a person sitting in front of a desktop screen. (<http://www.orin.com/3dtrack/docs/index.htm>)

laserBird is a optical system without camera; it uses scanning laser beams and photo-receptors, mounted onto the target, to track within a cone of approx. 60 degree and a distance up to 3 ft. (<http://www.ascension-tech.com/products/laserbird.pdf>)

<sup>8</sup> <http://www.vrdepot.com/manual-tracker.pdf>

careful antenna design, spaces up to 10ft can be covered.<sup>9</sup> The main drawback with this method is that the magnetic fields are very weak and drop significantly over distance, so that ferromagnetic and metallic objects can cause significant distortions.

Inertial system offer a completely different approach to track motion: instead of measuring the geometric relationship between the fixed and the body frame, they work without an external reference, but maintain a contained fixed frame by integrating the changes in position and angle over time. Being a self-contained system without having the need of any external aids makes this motion tracking principle highly attractive for navigation over big distances and through arbitrary environment.

The main drawback thereby is that, due to the lack of an external reference, long-time stability is very hard to achieve, and measurement errors quickly build up to a significant misalignment between the calculated and the actual position. This effect turned out to be particularly critical for linear movements, as linear inertial sensors are only able to pick up acceleration, which has to be double integrated to give information about displacement. Even worse, the accelerations resulting out of head motion are much smaller than the gravitational field of the earth; and this small signal has to be separated from this significant static acceleration for motion tracking.

This still limits the field of application of inertial sensing for head tracking to angular degrees of freedom only (3D systems).<sup>10</sup> To overcome this problem, hybrid systems add data from additional sensors to maintain long-term stability. For

---

<sup>9</sup> cp. Flock of Birds (<http://www.ascension-tech.com/products/flockofbirds.php>)  
see also tracking systems from Polhemus (<http://www.polhemus.com/FastTrak/fastrak.pdf>)

<sup>10</sup> E. Foxlin, Inertial head-tracking, M.S. Thesis, Dept. of E.E.C.S., MIT, 1993.



example, an electronic compass can constantly adjust the orientation, or acoustic systems provide the Cartesian coordinates, whereas gyroscopes measure rotation.<sup>11</sup>

The key characteristics to evaluate different tracking technologies are resolution, accuracy, and system responsiveness.<sup>12</sup> Besides these basic features, one most consider the size of the sensor part that is attached to the person must be considered, as well as the size of the working volume, the need for preparing the environment with special equipment, and the general robustness and easiness to use. For Acoustic Chase, the latter features gain special importance, as all technical components should be as small as possible, and no cumbersome elements should disturb the user experience or limit the design possibilities of the outer appearance. Furthermore the system must be easy to use, without any need for special preparation or assistance. Also the hardware should be easy to set up and take down, with minimal needs in preparing the environment or calibrating. Accuracy is subordinated here: the acoustic sense is much less precise on localizing objects than the visual sense. While the angular resolution of the human hearing is pretty high, especially in forward direction (around 2°), it is very hard to hear how far a sound source is away.<sup>13</sup> The sensor therefore should have a good angular resolution, but may lack in picking up the linear axis.

Considering the stated priorities for selecting a sensor technology, inertial systems are a good choice, as they meet all requirements: they are small, self-contained without the need for external reference aids, offer a high angular resolution, and are pretty fast. Furthermore, this type of sensor is available for a very

---

<sup>11</sup> E. Foxlin, M. Harrington, and Y. Altschuler, Miniature 6-DOF inertial system for tracking HMDs, SPIE vol. 3362, Proc. AeroSense '98 Conference on Helmet- and Head-Mounted Displays III, Orlando, FL 1998

<sup>12</sup> Resolution refers to the fineness of the measurement; accuracy describes how close the measured value is to the true value; system responsiveness comprises different types of time lags that result out of the measurement, in particular sample rate, and time between the reported and the actual movement of the tracked object.

K. Meyer, H.L. Applewhite, and F.A. Biocca(1992). A Survey of Position-Trackers. Presence: Teleoperators and Virtual Environments. 1 (2) (spring 1992), pp. 173–200.

<sup>13</sup> B. Hartmann: How We Localize Sound, in Physics Today on the Web, 1999 (<http://www.aip.org/pt/nov99/locsound.html>)  
J. Blauert, Spatial Hearing, 2nd ed., J. S. Allen, trans., MIT Press, Cambridge, Mass. 1997

moderate cost. The main challenge in using this technology is to deal with its strong limitations in tracking linear motion. As described below in detail, MEMS inertial units inherently don't offer long-term stability, especially for the linear axis, so they are mostly used for angular tracking only. Acoustic Chase is purely acoustic feedback system, so it has fewer requirements in tracking the linear position. This offers the possibility to base the location tracking on the inertial system and avoid a hybrid solution. Nonetheless, for further experiments, adding a tracking system with higher locational accuracy, like magnetic technologies, will offer advanced possibilities.

### **2.1.2 Inertial Motion Sensing**

The advantage of a self-contained tracking system that moves with the object and doesn't need external reference points is an ideal navigation system for aircrafts, used for the first time in military rockets during WWII. A rotating plate (Gimballed Inertial Platform) that could freely move around the three rotational axes was kept stable in its original reference orientation by fast spinning masses, while the navigated aircraft moved relative to it. This technology has been further developed to a complete Inertial Navigation System by adding accelerometers and other electromechanical elements to stabilize the inertial platform.<sup>14</sup>

In later systems, electronic integrators replaced the stabilized platform that performed as a three-dimensional mechanical integrator, getting rid of complex mechanical bearings. In this strap-down approach the accelerometers and gyroscopes were fixed (“strapped down”) to the chassis.<sup>15</sup>

---

<sup>14</sup> D. MacKenzie, *Inventing Accuracy: A Historical Sociology of Nuclear Missile Guidance*, MIT Press, 1991

<sup>15</sup> D. H. Titterton, J. L. Weston, *Strapdown Inertial Navigation Technology*, IEE Radar, Sonar, Navigation and Avionics, No 5

Recent developments in MEMS technology allow micro-sized sensors to pick up angular rates and linear accelerations.<sup>16</sup> These single chip sensors can be combined to a very small sized strap-down inertial measurement unit, sensing all six rotational and linear degrees of motion. With this enormous downscaling in size, and also in price, inertial measurement has gained attention for use in tracking devices for human motion. The challenge is to achieve high accuracy with sensors that are much less accurate than the ones used for aircraft.<sup>17</sup>

### 2.1.2.1 6 Axis System

With low-cost micro-machined sensors available, three orthogonal angular rate sensors (gyroscopes) and three accelerometers parallel to them can be integrated in small packages, principally providing a full inertial measurement unit.

In the case of head tracking, the three orientational degrees, yaw, pitch, and roll, relate to shaking, nodding, and tilting the head. This defines a natural body frame fixed to the head, with the origin in the middle of the head, the x-axis pointing upwards, the y-axis pointing to the front, and the z-axis to the side, parallel to the ear-axis, as shown in figure 3. With the origin of this frame inside the body, the inertial measurement unit can't coincide with all axes and the origin, but has to be placed with an offset. A typical position would be on top of the head or near one ear, with the axes of the measurement unit oriented parallel to the main head axes. The picked up rates then directly correspond to the rotation about the main head axes. No placement without direct implantation, however, is perfect: the accelerometers

---

<sup>16</sup> A. Lawrence, *Modern Inertial Technology*, Springer-Verlag, New York, 1993

A. Shkel, R. Horowitz, A. Seshia, and R. T. Howe, *Dynamics and Control of Micromachined Gyroscopes*, in *The American Control Conference*, San Diego, CA, June 1999

[http://www.analog.com/Analog\\_Root/sitePage/mainSectionHome/0,2130,level4%253D%25252D1%2526Language%253DEnglish%2526level1%253D212%2526level2%253D%25252D1%2526level3%253D%25252D1,00.html](http://www.analog.com/Analog_Root/sitePage/mainSectionHome/0,2130,level4%253D%25252D1%2526Language%253DEnglish%2526level1%253D212%2526level2%253D%25252D1%2526level3%253D%25252D1,00.html)

<sup>17</sup> E. Foxlin, M. Harrington, and Y. Altschuler, *Miniature 6-DOF inertial system for tracking HMDs*" SPIE vol. 3362, Proc. AeroSense '98 Conference on Helmet- and Head-Mounted Displays III, Orlando, FL 1998

measure an additional component during rotations, as the whole inertial measurement unit gets displaced with rotation angle multiplied by center offset. Depending on the requirements in accuracy, this offset has to be numerically corrected, by transforming the frame of the inertial measurement unit to the head frame.

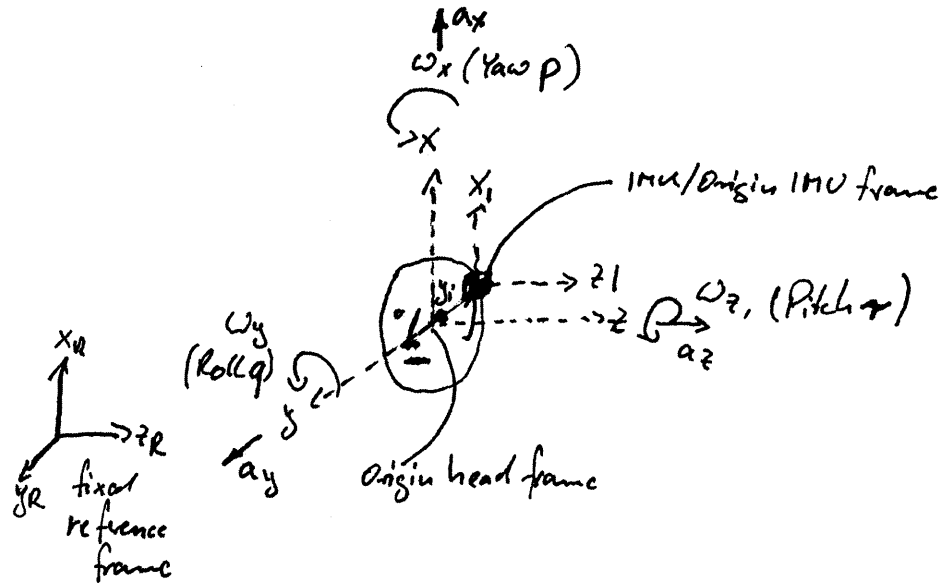


Figure 3: body frame and reference frame

Starting from an initial reference position, constant instantaneous integration of the angular speed, picked up by the gyroscope, provides the momentary position of the moving body. The three accelerometers measure the total acceleration vector, which is the sum of the constant gravitational vector  $g$  and acceleration resulting from motion. Using the orientation information from the gyroscope, the gravitational vector can be converted to the body frame, and subtracted from the measured acceleration. The remaining acceleration data is first integrated to velocity, which in turn is integrated to position offset. Figure 4 illustrates this process.

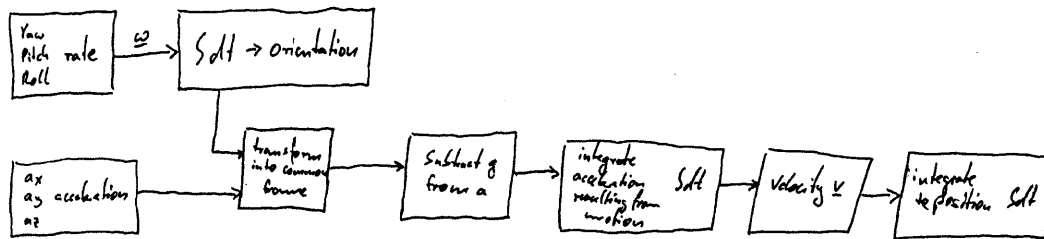


Figure 4: integration of sensor signals

In this idealized flow diagram sensor inaccuracies and drift are not included. MEMS sensors currently are much less stable and precise than the high quality accelerometers and gyroscopes used for aviation (up to a factor  $10^6$ ).<sup>18</sup> This is especially critical, as already small errors in orientation will cause parts of the gravitational vector to be erroneously added to the integrals that provide speed and position. Due to the double integration, bias in measuring acceleration will cause linear increasing speed and, even worse, quadratic growing position offset, making the tracking result unstable. To keep this error below 1 cm/s, the pitch and roll accuracy must be better than  $0.05^\circ$ .<sup>19</sup>

### 2.1.2.2 3 Axis System

With the currently available micro-sized inertial sensors, reliable position tracking has not yet been proven feasible without frequent updates. All commercial systems for inertial head tracking therefore focus on providing orientation degrees only. These 3DOF systems primarily process the angular rates from the gyroscopes, using additional sensors for drift correction. In this setup the acceleration data is only used to find the downward directions and to stabilize pitch and roll. In so-called

<sup>18</sup> Laser gyros (RLG) drift about  $0.001^\circ/\text{hr}$ , whereas a typical MEMS gyro drifts about  $1800^\circ/\text{hr}$

<sup>19</sup>  $\alpha < \text{asin}(0.01/9.8) = 0.058^\circ$

MARG units (Magnetic Angular Rate Gravity) additional magnetic compass sensors eliminate yaw drift.<sup>20</sup>

Acoustic Chase will use an inertial measurement unit equipped with three gyroscopes and three accelerometers, mounted on headphones near the ear. For different applications, a varying subset of the available sensor outputs will be processed. One of them will be a 3DOF configuration, where the x gyroscope picks up the turning of the head (yaw), and the y and z accelerometers work as tilt sensors, delivering the nodding and tilting of the head (pitch and roll) in a range up to  $\pm 90^\circ$ . For content requiring position information, filters will be used, allowing location tracking with low accuracy over short periods of time. Imprecision in position is acceptable for acoustic content, as the aural-sense is much more sensitive to horizontal rotation than to linear differences, and generally much less accurate than human vision. Long-term drift of the sensors is being reset to zero by subtracting a low-pass response from the signal.

### **2.1.3 Signal Processing**

MEMS sensors from the manufacturer Analog Devices were chosen for the inertial measurement unit. The ADX-series (accelerometers and gyroscopes) offers good performance in a small package at low prices. The sensors are fully integrated and need just a few simple external components to operate.<sup>21</sup>

The range of these sensors matches the requirements of head tracking. Their moderate range of linear acceleration is mainly determined by 1g gravity plus acceleration resulting from motion below 1g; so the ADXL202E accelerometer with

---

<sup>20</sup> W. Frey, Application of Inertial Sensors and Flux-Gate Magnetometers to Real-Time Human Body Motion Capture, Master's Thesis, Naval Postgraduate School, Monterey, California, September 1996.

<sup>21</sup> ADXL202E: low-cost  $\pm 2$  g dual-axis accelerometer with duty cycle output  
ADXRS300EB:  $\pm 300^\circ/\text{s}$  Single Chip Rate Gyro

its  $\pm 2g$  range covers all normally occurring acceleration values. Angular velocities are within a peak maximum of  $600^\circ/s$ .<sup>22</sup> By adding an external scale resistor, the measurement range of the ADXRS300 gyroscope, which is normally  $300^\circ/s$ , doubles to this value. The read-out rates for the sensor data must be higher than 30Hz, as head motions can contain frequencies up to 15Hz.

The angular rate is displayed as a linear voltage curve centered around 2.5V (half the supply voltage of 5V). It is digitized with a quantization of 12 bits, which fully exploits the sensor accuracy which is around  $0.5^\circ/s$ .<sup>23</sup>

The acceleration sensors interface via a duty cycle output, whose ratio of high and low time is linear to the measured acceleration. The resolution is 1mg, or 0.05% of the total scale.<sup>24</sup> The timer that picks up the PWM-signal therefore must count with a minimum number of 2,000 steps per cycle.

In the current version of Acoustic Chase, an embedded controller handles the sampling of the sensors, and performs some preliminary filtering of the signals, then passes these results to a Windows machine that handles the remaining processing steps.

### 2.1.3.1 Spatial Geometry

Two main coordinate systems have to be considered for inertial navigation: the reference frame remains fixed in space and describes the position of the tracked head and other objects, such as the sound sources, in a common view. The body frame is attached to the moving object, which here is the human head, with the y-axis

---

<sup>22</sup> E. Foxlin and N. Durlach, "An Inertial Head Orientation Tracker with Automatic Drift Compensation for Use with HMD's," Proceedings from VRST'94, Virtual Reality Software and Technology, Singapore (August 23-26, 1994).

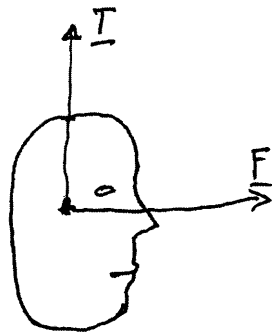
<sup>23</sup> The nonlinearity of the ADXRS300 is stated as 0.1% typ., which is  $0.6^\circ/s$  for the  $600^\circ/s$  range. The noise density is specified as  $0.1^\circ/s \sqrt{\text{Hz}}$ , which is  $0.5^\circ/s$  random rate error for a 30Hz bandwidth. 12 bits divide a range of  $\pm 600^\circ/s$  in 4096 steps of  $0.3^\circ/s$ .

<sup>24</sup> The specified noise floor is  $200\mu g\sqrt{\text{Hz}}$ , which means a resolution of 1mg at 30 Hz bandwidth

pointing in forward direction, x-axis to the top, and z-axis to the side establishing a right-handed coordinate system. The local difference between the center of the inertial measurement unit and the origin of the body frame for the human head of about 3 inches will be neglected for the further discussion, as it only affects the position, which is anyway flawed with significant errors resulting from the measuring principle based on double integration.

To fully describe the orientation of the human head, two vectors are used: the top and the front vectors. The top vector  $\underline{T}$  points straight up through the top of the head, and the front vector  $\underline{F}$  points forward through the listener's face at right angles to the top vector.

By default, the front vector is  $\underline{F}=(0, 1, 0)$ , and the top vector is  $\underline{T}=(1, 0, 0)$ .



**Figure 5: top and front vector describing the head orientation**

The inertial forces are referenced to the body frame, strapped-down to the moving head. Thereby the gyroscopes pick up yaw rate  $\omega_x$  (head shaking), roll rate  $\omega_y$  (tilting), and pitch rate  $\omega_z$  (nodding); the accelerometers also are moved with the head, and pick up accelerations referenced to the body frame. Suitable coordinate transformations allow passing the numerical values for measurands and vectors from one frame to the other without losing consistency.



Signal processing is split in two branches: obtaining the head orientation, and getting its position.

On the assumption that the head is mostly in upward position, or in the top 180°-dome, a stable way to track roll and pitch (q, r) is to use the y and z accelerometer as tilt sensors and obtain the derivation from the horizontal position with following equations:

$$q = \sin^{-1}\left(\frac{a_y}{g}\right) \quad (1)$$

$$r = \sin^{-1}\left(\frac{a_z}{g}\right) \quad (2)$$

Yaw (p) is obtained from the x gyroscope, integrating its measured angular rate in body frame coordinates:

$$p = \int \omega_x dt \quad (3)$$

These angles are transformed to the orientation vectors  $\underline{T}$  and  $\underline{F}$  in the reference frame by starting at the default upward front position for  $\underline{F}$  and  $\underline{T}$  and first rotating  $\underline{F}$  about  $\underline{T}$  with angle p, then rotating  $\underline{F}$  and  $\underline{T}$  about  $\underline{F} \times \underline{T}$  (which is the local z or tilt axis) with q, and after that rotating  $\underline{T}$  about  $\underline{F}$  with r. The resulting vectors  $\underline{T}$  and  $\underline{F}$  define the actual position of the tracked head.

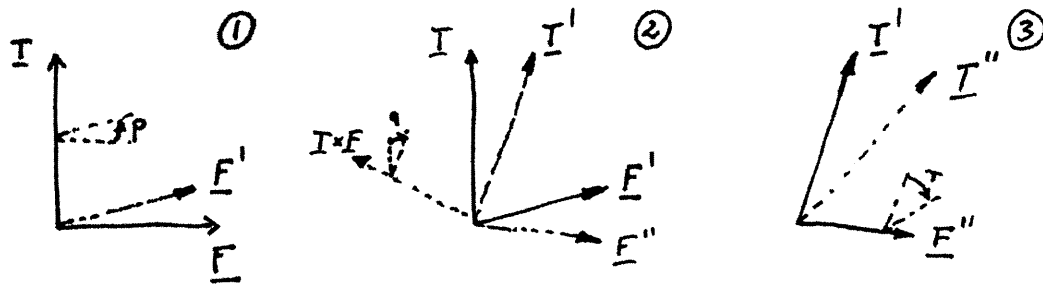


Figure 6: 3 consecutive rotations about  $\underline{T}$ ,  $\underline{F} \times \underline{T}$ , and  $\underline{F}$

To obtain head displacement, gravity has to be cancelled out of the acceleration data. As the accelerometers are fixed to the head, the gravitation vector moves relative to them, whenever the head is rotated. Splitting the measured data in components from gravitation and from motion can only be achieved if the momentary position of the gravitation referenced to the body frame is known with high precision. Assuming that the average rotation of the human head over a longer period of time is zero, the low-passed filtered data rates from the accelerometers will define the gravitation vector. To maintain the correct orientation of the gravitation expressed in body coordinates, this vector is rotated with the angular rates provided by the gyroscopes. Now knowing the direction of gravitation in body coordinates, the acceleration due to motion can be calculated by simply subtracting  $\underline{g}$  from the measured acceleration  $\underline{a}$ .

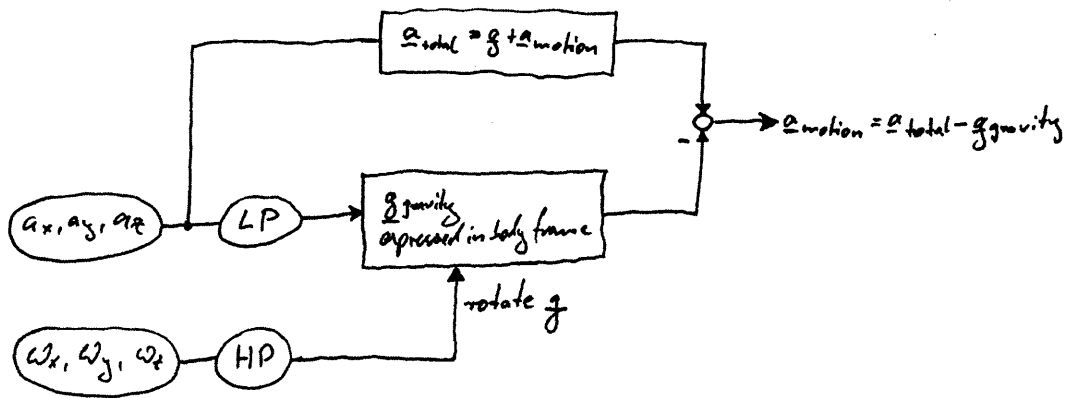


Figure 7: correction of gravity

Velocity, still expressed in the body frame, is obtained by integrating the acceleration over time.

$$v = \int a^* dt \quad (4)$$

Integrating speed to position is done in reference coordinates, by using the head orientation vectors  $\underline{T}$  and  $\underline{F}$ .  $v_y$  denotes the speed in forward direction of the head, so the head moves with  $v_y$  parallel to  $\underline{T}$ . Accordingly it moves with  $v_x$  parallel to  $\underline{T}$ , and with  $v_z$  parallel to  $\underline{T} \times \underline{F}$ .

$$dx = v_x \underline{T} + v_y \underline{F} + v_z \underline{T} \times \underline{F} \quad (5)$$

$$x_{t+\Delta t} = x_t + dx \quad (6)$$

### 2.1.3.2 Filtering

Inertial tracking is very sensitive to measurement errors, as these tend to build up during the involved integration steps. Important error sources are sensor noise, scale factor, offset, bias and drift.<sup>25</sup> For achieving optimal accuracy in tracking, these errors have to be minimized or compensated using a suitable mathematical model with carefully system calibrated constants.<sup>26</sup>

For Acoustic Chase, the precision requirements are much less rigid; first, as the location abilities of the human hearing are limited, second as acoustic content in this environment is designed for performance with somewhat imprecise head tracking.

This means that strong filters can be used to stabilize the output signals of the inertial tracking unit. This will keep the filter implementation less complex than using the Kalman filter, which is often used for high precision inertial measuring.<sup>27</sup> The following assumptions determine the filter process:

The horizontal orientation of the head (yaw  $p$ ) can have long time drift. This corresponds to a slow turning of the acoustic scene in the horizontal plane. As the virtual acoustic objects are not locally referenced to the real scene, the person won't notice the slow rotation of the whole scene around this axis.

Pitch and roll ( $q$ ,  $r$ ) of the head are tracked using accelerometers, which rely on gravitation and therefore have limited bias. By means of a high-pass filter these two angles are slowly dragged to zero position, assuming that the average head position is upright. This mainly compensates different angles of headphones and head that will naturally occur whenever the headphone is set up. This bias compensation,

---

<sup>25</sup> R. Dorobantu, Simulation des Verhaltens einer Low-cost strapdown IMU unter Laborbedingungen, Institut für Astronomische und Physikalische Geodäsie Forschungseinrichtung Satellitengeodäsie TU-München, IAPG/FESG No. 6, 1999

<sup>26</sup> R. M. Rogers, Applied Mathematics in Integrated Navigation Systems, American Institute of Aeronautics and Astronautics, AIAA, Direction Series, 2000

<sup>27</sup> For information on the Kalman filter see G. Welch, G. Bishop, An Introduction to the Kalman Filter, TR-95-041, Department of Computer Science, University of North Carolina at Chapel Hill, updated in 2002

though, is limited to  $\pm 30^\circ$  to not overcompensate when a person tilts his or her head for a really long time, like when lying on the floor.

High-pass filters are also used to stabilize location tracking, the most critical process in inertial measurement. Following the same idea as above, the long-time average velocity of the head is assumed to be zero. This can be justified, as most natural human head movements change frequently over time, like changing orientation, accelerating, or slowing down. Furthermore, the acoustic content can be optimized for reacting upon abrupt motion rather than on slow constant linear displacement. By selecting an adequate corner frequency, both integration steps, one for velocity and one for position, can be kept stable. This allows the use of inertial measurement for linear location measurement limited to track offsets from a momentary origin.

The following diagram shows the first order high-pass filter used here. This filter is synthesized by subtracting a low-pass response from the signal. The time constant for rotation has been chosen as  $\tau_{rot} > 20$  sec, the time constant for linear position as  $\tau_{lin} > 5$ sec.

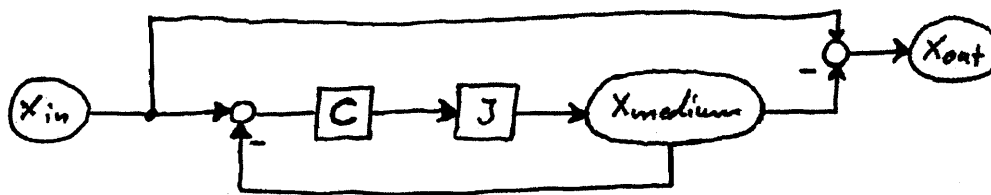


Figure 8: high-pass filter (C is a constant factor; J is an integrator)

## 2.2 Sound Generation

Acoustic Chase uses pre-recorded and digitally stored audio content that is instantaneously accessed, filtered, and mixed to generate real-time feedback from tracked human motions. The sound system therefore flexibly organizes the audio content and alters its physical quality during playback. This chapter describes the signal chain and audio processing techniques.

Basic sound control comprises volume adjustment, mixing of several channels, and varying the playback speed. Adding effect blocks, coupled with a crossbar mixer, allows for a rich variety of sound manipulation, including 3D effects. Figure 9 shows the audio paths from the playback sources passing through different processing blocks to the headphone output.

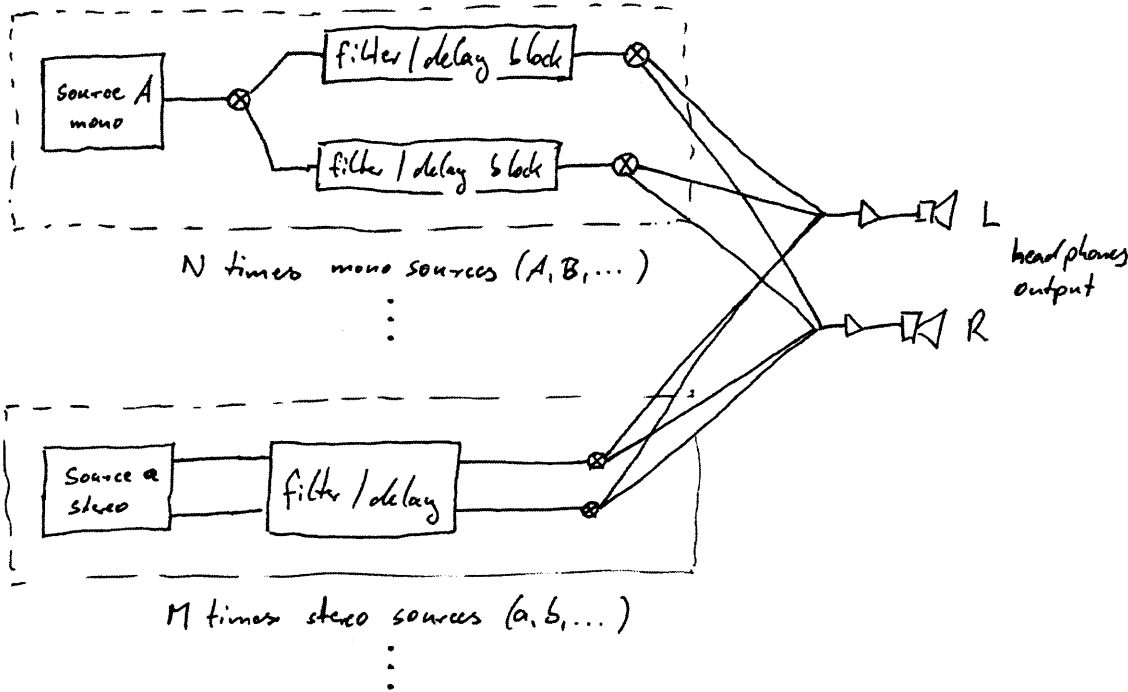


Figure 9: audio paths

The acoustic material is either single channel or dual channel. The mono signals are split up, with both forks passing through individual filters, configured for spatialization as described below. Stereo signals go through effect filters as well, but bypass the 3D stage.

Signal sources, mixers, and filters are accessible by the main controller for instant changes.

## 2.2.1 3D Audio

### 2.2.1.1 Psychoacoustic Principles

3D audio positions sounds around the listener. The possibility to generate localized audio spots that move in space is an important feature for Acoustic Chase.

Spatial hearing is based on three primary mechanisms:<sup>28</sup>

- Interaural Level Difference (ILD)

For sound coming from the side, the head damps the signal getting to the ear that is turned away. This creates a difference in amplitude between both ears, especially for higher frequencies.

- Interaural Time Difference (ITD)

Small differences in arrival times of sound, not coming from center plane rectangular to the axis between both ears, can be decoded to determine the direction. This effect works well for lower frequencies.

- Head Related Transfer Function (HRTF)

Depending on the arrival direction, sound waves are scattered differently by the external ear, the head, and shoulders. This directional filtering with attenuations and boosts over the frequency range, gives further information about the location of the sound for almost all orientations, including the vertical axis.

Additional secondary clues for localizing sounds come from the influence of room acoustics, most importantly echo from the walls, which provide reverberation cues.

---

<sup>28</sup> Cp. J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*, The MIT Press, Cambridge, Massachusetts, 1983.

B.C.J. Moore, *An introduction to the psychology of hearing*, 5th edition. University of Cambridge. 2003



### 2.2.1.2 Displaying Spatial Audio

Acoustic Chase is based on headphones. Thus, the audio signals are directly transferred to the ears without crosstalk between the two channels and with no external reflections. Embedding the sound localization cues in audio streams in such a framework is called binaural synthesis, and many techniques for generating an immersive sound field are based on this. For example, a simple, yet effective method is placing two microphones inside the ears of an “artificial head”, which modifies the recorded sound-waves like a real head would do before they reach the inner ear. Played-back next to the ear by headphones, the original spatial quality is reproduced.

Filters and sound manipulation can imitate all these sound localizing cues, naturally generated by the human or artificial head. A typical system, as shown in Figure 10, adds level differences, inter-aural delays, and transfer functions separately for each ear to give the impression that the sound comes from a specific direction.<sup>29</sup>

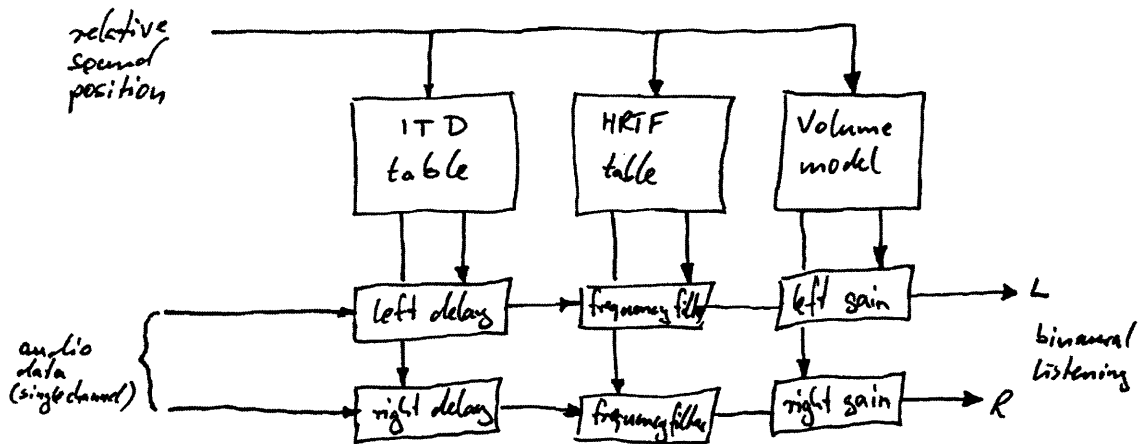


Figure 10: spatialization of an audio signal

<sup>29</sup> J. Huopaniemi, M. Karjalainen, HRTF Filter Design Based on Auditory Criteria, NAM, Helsinki, 1996

The most critical aspect thereby is to design the head-related-transfer functions. As this complex frequency filter results out of anatomic features, such as head and ear shape, HRTFs can differ significantly from person to person. Therefore using non-individualized transfer-functions generally limits the achievable results.<sup>30</sup>

The sophistication of the head-related-transfer functions is a main characteristic distinguishing 3D audio systems, whereas the other two primary features of spatial hearing, ILD and ITD, are easy to realize with delay lines and volume control. Beyond that, simulating the environmental acoustic conditions of a specific scene, like echo from wall reflection and Doppler effect, may add to the realism of the acoustic model.<sup>31</sup> With massive computational power available even on small systems, complex HRTFs and spatial rendering can now be designed, propelling the quality of synthesized audio scenes in 3D environments.<sup>32</sup>

Most commercial systems run on PC platforms and are developed for the game industry. DirectSound3D, for instance, is a standardized API that bundles a growing variety of hard- and software components in one platform.<sup>33</sup>

Specialized single-chip processors optimized for audio applications outside the PC domain, such as home entertainment, are also currently gaining momentum. These stand-alone chips are able to embed all primary 3D cues into audio streams,

---

<sup>30</sup> E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, Localization using nonindividualized head-related transfer functions, *J. Acoust. Soc. Am.*, 94(1), 1993, pp. 111–123.

W. G. Gardner and K. D. Martin, HRTF measurements of a KEMAR, *J. Acoust. Soc. Am.*, 97(6), 1995, pp. 3907–3908.

<sup>31</sup> D. R. Begault, *3-D Sound for Virtual Reality and Multimedia*, Academic Press, Cambridge, MA, 1994

<sup>32</sup> "InTheMix", a 3D audio installation takes advantage of the very sophisticated rendering system AuSim (see note 32). W. L. Chapin, AuSIM Incorporated, et al. (2000): "InTheMix", interactive audio environment, Siggraph 2000 (<http://www.ausim3d.com/InTheMix/history.html>).

Another 3D augmented sound reality application has been developed by the Fachhochschule Hagenberg: M. Haller, D. Dobler, P. Stampfl, Augmenting the Reality with 3D Sound Sources, ACM SIGGRAPH 2002

<sup>33</sup> Alternative systems are OpenAL and AuSIM.

OpenAL is a software interface to audio hardware, specialized to generate simple 3D audio, using distance-related attenuation, Doppler effect, and environmental effects such as reflection, obstruction, transmission, and reverberation. (<http://www.openal.org>)

AuSIM3D physically models the propagation from the sound sources to the listener with a high degree in details to achieve a convincing experience (<http://www.ausim3d.com>)

allowing for the design of miniature systems based on embedded controllers, without the overhead of a complete workstation.<sup>34</sup>

### 2.2.2 Direct Sound 3D

Acoustic Chase uses Direct Sound, the Microsoft API for sound control. This software package offers basic 3D audio, including head-related-transfer functions. Several independent localized sound sources can be displayed in parallel to the standard audio channels and a variety of acoustic effects are available.<sup>35</sup>

Direct Sound is open for expansion, both in lower layers, as in supporting hardware acceleration on sound cards, as well as in higher layers, like software modules for advanced 3D audio. EAX, for example, optimizes the Direct Sound functions and offers enhanced audio attributes.<sup>36</sup> Thus starting with standard Direct Sound provides the opportunity to significantly improve the acoustic rendering quality by adding 3<sup>rd</sup> party tools, with only minimal changes in code.

The 3D effect in Direct Sound is controlled by positioning the listener and one or more sound sources in space. The orientation is defined by a top vector, pointing straight up through the listener's head, and a front vector, pointing straight forward through the listener's face. Position and orientation of the sound sources are described in a similar way. Additionally, the virtual sounds offer a "cone feature" for non-uniform sound emission, with volume dropping outside the cone. By providing velocity information, Doppler effects are realized as well. Table 1 and 2 give an overview of the involved parameters.<sup>37</sup>

---

<sup>34</sup> See TAS3103 Digital Audio Processor (<http://focus.ti.com/docs/prod/folders/print/tas3103.html>)  
The SHARC Melody chip series from Analog Devices offers similar features  
([http://www.analog.com/Analog\\_Root/productPage/productHome/0,2121,SSTMELODYSHARC,00.html](http://www.analog.com/Analog_Root/productPage/productHome/0,2121,SSTMELODYSHARC,00.html))

<sup>35</sup> see <http://www.microsoft.com/windows/directx/default.aspx>

<sup>36</sup> see <http://eax.creative.com>

<sup>37</sup> Tables taken from the Microsoft DXSDK9 documentation

## IDirectSound3DListener8 Interface

### Global sound properties

Method	Description
<u>GetDistanceFactor</u>	Retrieves the distance factor, which is the number of meters in a vector unit.
<u>GetDopplerFactor</u>	Retrieves the multiplier for the Doppler effect.
<u>GetRolloffFactor</u>	Retrieves the rolloff factor, which determines the rate of attenuation over distance.
<u>SetDistanceFactor</u>	Sets the distance factor.
<u>SetDopplerFactor</u>	Sets the multiplier for the Doppler effect.
<u>SetRolloffFactor</u>	Sets the rolloff factor.

### Listener properties

Method	Description
<u>GetOrientation</u>	Retrieves the orientation of the listener's head.
<u>GetPosition</u>	Retrieves the listener's position.
<u>GetVelocity</u>	Retrieves the listener's velocity.
<u>SetOrientation</u>	Sets the orientation of the listener's head.
<u>SetPosition</u>	Sets the listener's position.
<u>SetVelocity</u>	Sets the listener's velocity.

Table 1: IDirectSound3DListener8 interface

## IDirectSound3DBuffer8 Interface

### Minimum and maximum distance

Method	Description
<u>GetMaxDistance</u>	Retrieves the maximum distance, which is the distance from the listener beyond which sounds in this buffer are no longer attenuated.
<u>GetMinDistance</u>	Retrieves retrieves the minimum distance, which is the distance from the listener at which sounds in this buffer begin to be attenuated.
<u>SetMaxDistance</u>	Sets the maximum distance.
<u>SetMinDistance</u>	Sets the minimum distance.

### Position

Method	Description
<u>GetPosition</u>	Retrieves the position of the sound source.
<u>SetPosition</u>	Sets the position of the sound source.

### Sound projection cone

Method	Description
<u>GetConeAngles</u>	Retrieves the inside and outside angles of the sound projection cone.
<u>GetConeOrientation</u>	Retrieves the orientation of the sound projection cone.
<u>GetConeOutsideVolume</u>	Retrieves the volume of the sound outside the outside angle of the sound projection cone.
<u>SetConeAngles</u>	Sets the inside and outside angles of the sound projection cone.
<u>SetConeOrientation</u>	Sets the orientation of the sound projection cone.
<u>SetConeOutsideVolume</u>	Sets the volume of the sound outside the outside angle of the sound projection cone.

### Velocity

Method	Description
<u>GetVelocity</u>	Retrieves the velocity of the sound source.
<u>SetVelocity</u>	Sets the velocity of the sound source.

Table 2: IDirectSound3DBuffer8 interface

## 2.3 Positioning Sounds

In Acoustic Chase 3D sounds are modeled as moving particles based on Newtonian mechanics. This physical model describes a mass point traveling in three-dimensional space affected by force fields. The resulting motions are generally smooth and intuitively understood by the listener.<sup>38</sup>

Newton's equations describe the change in velocity of a particle over time in reaction to impulses from other objects outside, as well as the particle's impetus. The basic equations add all acting forces  $\underline{F}_j$  to an acceleration vector  $\underline{a}$ , which gradually changes the velocity  $\underline{v}$  in an integration step, which in turn is integrated to a continuous position path  $\underline{x}(t)$  over time:

$$\underline{F}_{tot} = \sum_{j=1}^N \underline{F}_j \quad (7)$$

$$\underline{a} = \frac{1}{m} \underline{F}_{tot} \quad (8)$$

$$\underline{v}_{i+1} = \underline{v}_i + \underline{a} dt \quad (9)$$

$$\underline{x}_{i+1} = \underline{x}_i + \underline{v} dt \quad (10)$$

The key to get a broad bandwidth of motions out of these equations is the actual implementation of how the acting forces  $\underline{F}_j$  are determined as a function of time, particle's position, and other parameters.

For Acoustic Chase the motion of a 3D sound is governed by two force fields, both originating at the listener's position: first, a far-reaching field that either attracts

---

<sup>38</sup> This method is widely used to generate realistic animation of virtual objects. Cp. B . Arnald, G . Dumon, G . Hégro, N. Magnenat-Thalman, D. Thalman, Animation Control with Dynamics, in: Magnenat-Thalman N, Thalman D (eds) State-of-the-Art in Computer Animation, Springer, Tokyo, 1989, pp.113-124.

D. Thalman, Robotics Methods for Task-level and Behavioral Animation, in: D. Thalman (ed) Scientific Visualization and Graphics Simulation, John Wiley, Chichester, UK, 1990, pp.129-147.

or repels the virtual object in relation to the listener; second, a well of fast rising locally confined outward forces that push the moving object away when it comes too close, ensuring that it doesn't get glued to the listener's center position. This resembles a virtual listener's body that can't be intruded. Figure 11 shows the radial field resulting from the overlay of the two discussed forces. Parameter  $\xi$  gradually changes the far-reaching field from attraction to repulsion.

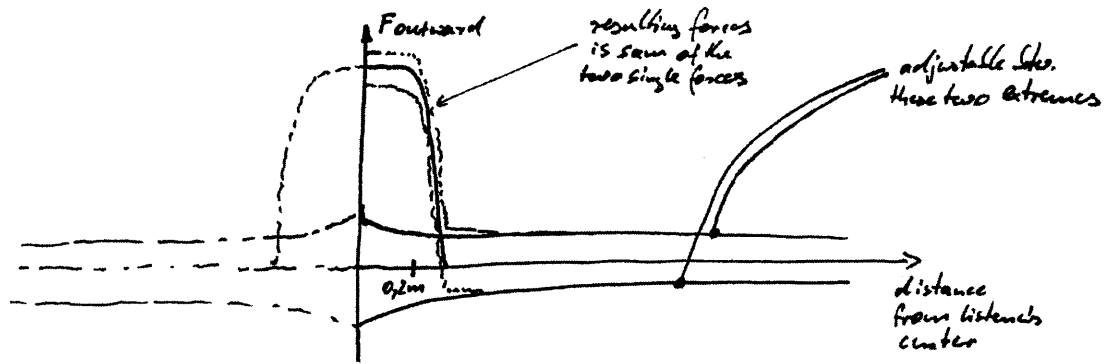


Figure 11: radial force field

To keep the velocity of the moving object within a certain range, two additional forces  $\underline{F}_j$  are introduced: friction  $\underline{F}_F$  and self-drive  $\underline{F}_D$ . A growing friction coefficient decelerates the objects, when its speed exceeds a certain limit.

$$\underline{F}_F = -\lambda \underline{v} \quad \text{with} \quad \begin{cases} \lambda = \lambda_1 & \text{if } |\underline{v}| < v_{crit.} \\ \lambda = \lambda_2 & \text{if } |\underline{v}| \geq v_{crit.} \\ \lambda_1 < \lambda_2 \end{cases} \quad (11)$$

Added random forces propel the sound object and give it own drive. Parameters  $b$  and  $c$  adjust the rate and amplitude of these pushes.

$$\begin{aligned} \underline{F}_D &= \underline{x} \text{ rand}(c) + \underline{y} \text{ rand}(c) + \underline{z} \text{ rand}(c) && \text{if } b > \text{rand}(1) \\ \underline{F}_D &= 0 && \text{else} \end{aligned} \quad (12)$$

### **3. Sound Implementation**

Live recordings from an amusement arcade in Coney Island/NY were chosen for the sound implementation. This sound moves as a virtual object around the listener, and approaches in irregular loops circling faster and faster the closer it gets. Whenever the person moves or shakes his head, the sound is left behind or repelled. But the chasing object keeps track and approaches again.

My hypothesis was that the soundscape of the lively gambling hall offers some features that fit with the Acoustic Chase concept. When the sound is heard from far away in low volume, it gives an impression of an attractive and joyful place. But this turns into a hectic atmosphere when the sound is close. Higher and higher volume and hectic motions around the listener develop into an aggressive atmosphere that besieges the listener.

The captured sound fragments from several gambling machines form a dense mix of melodic elements, artificial voices, and many other noises. This rich sounding collage overwhelms the listener at first, and it takes a while to make some sense of the heard sound. Although it is a concrete sample, the soundscape has abstract qualities. It is very hard to build a coherent imagined picture for all the different noises. This should keep the mind of the listener on site and not immediately transport it to a different world as it would happen with a more suggestive recording.

Gambling also talks about circulating money, winning and losing, and addiction. This adds connotations to Acoustic Chase that point to dimensions beyond the actual acoustic experience.



## **4. Placement Within the Artistic Context**

Acoustic Chase is a sound installation in the field of interactive art based on technology. This category covers a wide range of artistic practice. Still, as a main characteristic, sound is generally not produced by human players following a score, but result from processes automatically executed by technical devices. These pieces rely on a strong conceptual component, treating sound not as instrumental music, but as acoustic material, editing and mixing recorded or synthesized sound fragments. The degree of actual interaction between human and artwork can span from passive objects to highly interactive “complex machines, where the user does not so much individually control the work, but cooperates, obstructs and directs.”<sup>39</sup>

By examining the appearance, the interaction concept, and the use of sound, as well as establishing relations to other pieces, this chapter will survey the space of communicative dimensions in which Acoustic Chase is located.

### **4.1 Appearance of the Installation**

The main physical part of Acoustic Chase that the user interacts with is the set of headphones, which keeps the visual appearance minimal. Once they are set up and the actual experience begins, the headphones stay out of the field of view, contributing to an “invisible installation”.

The combination of immersive audio with free sight puts the participant in an intermediate state between being part of both a virtual environment and a real space. The unchanged visual information blends with autonomous, disembodied sounds that react to the human movements, come closer and move away, or maintain their locations when the person moves through space. They represent the influences and constraints people are exposed to. They are quite there, but not seeable or tangible.

---

<sup>39</sup> [http://www.aec.at/en/archives/prix\\_archive/prixJuryStatement.asp?iProjectID=2560](http://www.aec.at/en/archives/prix_archive/prixJuryStatement.asp?iProjectID=2560)

This theme of imaginative sound locations or objects is used in many sound installations. Hiding the image of the sound source offers the potential to build up an imagination that is more powerful or multifaceted than an actual depiction. This can have different connotations: At R. Horn's object "The Turtle Sighing Tree"<sup>40</sup> soft sounds can be heard at the end of the long metal branches of the tree. The sound sources are hidden, but it seems that they are objects, located inside the tree, at a romantic place, not accessible for humans. J. Cardiff's "Forty-Part Motet"<sup>41</sup> takes a different approach to disembodied sounds: 40 speakers play back a chorus of 40 separately recorded voices, substituting each individual singer by a loudspeaker column. A magic moment arises from the discrepancy of the seen object, a black case on a tripod, and the natural beauty of the human voice singing the choral. The spatial distribution of the speakers, placed like the singers at the recording session, add to the impression that the square boxes are a clear representation of human beings, resulting in a tension between visual sense and the hearing. Both pieces are very sculptural, with the sounds being tied to fixed objects in space. The user determines how to approach them; there is always a clear distance maintained between the human being and the object.

In Acoustic Chase the clear separation of a fixed object in space and individual changes. The sound is all around the listener, with spatialized sound spots moving dynamically and through self-control. The person no longer sets the distance by walking back and forth, but a variety of (head) motions determine the actions of the acoustic environment.

The sounds, as the representation of the artwork, not only leave a passive state, they are also designed to confront and incite the listener. They do this by loud volume, approaching and cycling around the human head, and appropriate content, like spoken commands or suggestive content.

---

<sup>40</sup> R. Horn, *The Turtle Sighing Tree*, 1994

<sup>41</sup> J. Cardiff, *Forty-Part Motet*, 2001

While hearing has a dominant role here, the visual impressions are still important, and work as an anchor to keep the person on site. Locating the installation inside an empty, neutral room provides an adequate balance between hearing and seeing. By removing all distracting objects from the field of sight, the focus of perception shifts from the dominating visual sense to the secondary senses, increasing the awareness of them. “The absence of sight immediately stimulates an intimate, introspective listening experience that can be very unique.”<sup>42</sup>

## **4.2 Individual and Machine**

### **4.2.1 Mutual Influence**

Acoustic Chase is an active system that is designed to play with the person. An important element thereby is the complexity of the interaction – that is, how the acoustic scene is rendered in response to the tracked movements. Beside the actual sound content, this will significantly define the user experience. A key feature of Acoustic Chase is that it doesn't stay passive, but provokes the user in a situation of mutual influence.

The following three examples of interactive systems illustrate different aspects of user participation:

A very passive form of interaction is used for “Immersive Audio” by C. Moeller. This 3D audio installation enables a visitor to dive into virtual space, filled with sound objects that can be seen and heard. The objects remain passive and the person is a visitor with no influence.

“Very Nervous System” from D. Rokeby is a discrete interactive sound installation that triggers sounds by detecting the physical presence of the user inside specific image areas of a video camera that watches the scene from above. Each state

---

<sup>42</sup> G. K. Montgomery, Outer Ear/Inner Eye (<http://www.generatorsoundart.org/GSA-16.html>)

change of the system, that is changing which sound to play, is triggered by a defined event, which is activity in a specific cell in the video image. This approach introduces hidden information layers that appear one after another, released by the respective trigger events. As each change in system output switches between predefined states, and usually requires user action, the overall actions of the machine remain stable in clear boundaries. Nonetheless, Rokeby sees his installation already beyond a simple “control system. [...] The changing states of the installation are a result of the collaboration of these two elements (installation and person). The work only exists in this state of mutual influence”.<sup>43</sup>

A different quality of interaction is achieved with a continuous feedback system without the conservative discrete steps that limit the possible configurations of the generated output. By increasing the number of trigger points and decreasing reaction time, the discrete interaction can be transformed to a continuous system. The artwork liquefies to a dynamic process, whose states are no longer constrained by a small discrete set of manifestations, but form a continuum over a potentially large scale. These mechanisms go beyond simple trigger effect rules, but define an open interaction. Depending on the complexity, chaotic elements and system responses can occur, and the artist no longer completely overlooks what states the artwork will ever enter. In “Untitled Ball” by D. Jolliffe the continuous system is defined by the distance of the viewer, the physical properties of the wooden ball, and the characteristics of the control system. Because of its stateless feature, the movements of the ball are chaotic and unpredictable. Still, the feedback system uses mechanical principles that can be understood intuitively, leading to comprehensible movements.

---

<sup>43</sup> D. Rokeby, Lecture for "Info Art", Kwangju Biennale (<http://homepage.mac.com/davidrokeby/install.html>)

#### 4.2.2 Controlling And Being Controlled by Machines

Acoustic Chase explores how sound signals can be designed to gain some degree of control over humans. The headphones establish a very close contact between device and person, which can be seen as an “intimate interface”, directly connecting to the human input port, the ears. Whenever the headphones are put up, the machine gets direct access to the body that can’t be blocked any more. There is also a connotation of electrodes used for electroshock therapy, or helmets used for brainwashing in movies.<sup>44</sup>

The use of sound further supports this theme of a device partly intruding the individual: Many types of sound are not so clearly localizable. Compared to vision, hearing doesn’t establish such a clear distance to the sensed object. Hearing is sensitive to all directions and can’t be shut completely, always allowing sound to get through. In the art installation ACCESS Internet user can “shoot” sounds with an audio spotlight to visually tracked persons in public spaces.<sup>45</sup>

Sound also offers a great range in tones, melodies, and volume, which can act on the subconscious and influence feelings. A loud fire horn, for example, immediately triggers a surge of adrenalin, whereas a soft voice can develop power of persuasion by constantly talking over a long time. The computer HAL in 2001 uses a vocal interface for addressing the people on board;<sup>46</sup> they can’t escape – the machine’s smooth voice follows them everywhere, and slowly endears itself to the astronauts.

On the other hand, the power of sound is, of course, limited. People can withstand acoustic commands and ignore them. In Acoustic Chase the intensity of the acoustic stimuli increases the longer the person ignores the commands. The weak

---

<sup>44</sup> E.g. in the music video of C. Cunningham “Come On My Selector”, 1999

<sup>45</sup> M. Sester, Access, presented at Siggraph 2003

<sup>46</sup> D. G. Stork (Ed.), HAL's Legacy: 2001's Computer as Dream and Reality, MIT Press, 1997

coupling via sound establishes a mutual system that constantly rebalances instead of a clear command direction in either direction.

Stelarc, an artist interested in coupling the body with technology, uses a much more direct way to give machines power over humans: In “Ping Body”<sup>47</sup> and “Fractal Flesh”,<sup>48</sup> the human being gets subject to computer control by triggering of the muscles via applied electric impulses. In this extreme “human-machine symbiosis” the machine gains direct control over the body with a clear tendency of total mastery by the apparatus.

Connecting with the machine via electrodes is a complex process with a ritual connotation, in which the human slowly subordinates himself wire by wire, resulting in a bond not easy to leave. In Acoustic Chase, getting in contact with the machine is lightweight and just a matter of seconds: by putting the headphones on, people enter the interaction; by putting them down, they are out again.

## **4.3 The Use of Sound From a Media Perspective**

### **4.3.1 Simplified And Abstract Sounds**

Technical media has radically changed the ways in which sound was perceived, generated, and thought about. “Technologies of electric media were integrated into the creative techniques, [...] gaining control and the technically determined feasibility of what was previously unachievable. The storage, transmission and synthesis of sound as well as intermedia transformation and virtualization.”<sup>49</sup>

Recording devices capture any kind of sounds, which, now disembodied, have been

---

<sup>47</sup> <http://www.stelarc.va.com.au/pingbody/ping.html>

<sup>48</sup> <http://www.stelarc.va.com.au/fractal/>

<sup>49</sup> G. Föllmer, Audio Art ([http://www.medienkunstnetz.de/themes/overview\\_of\\_media\\_art/audio/21/](http://www.medienkunstnetz.de/themes/overview_of_media_art/audio/21/). Rev. 2004-04-30)

added to the material for artistic production. Sounds became part of the aesthetic vocabulary – the “art of noise”<sup>50</sup> became a leitmotif of modern times.

In music “there was a tendency through the whole twentieth century, from the Futurists on, to use noises, anything that produced sound, as a musical instrument.”<sup>51</sup> With availability of the tape recorder, the movement *Musique Concrète* worked with recorded samples as sound objects, which transform to music objects when assembled to a collage.<sup>52</sup> Recording technology introduced a new way to work with non-instrumental sounds and provided a basis for composing that is “no longer dependent upon preconceived sound abstractions, but now using fragments of sound existing concretely and considered as sound objects defined and whole...”<sup>53</sup> The early use of tape recorder for “organization of sound” (Cage) by manipulations of the tape, like cutting and splicing, anticipated the cut and paste of digital production, and remains a common principle to many contemporary artworks.

In accepting almost any kind of object as a potential element for composing, the established concept of linear procession, as described by a sequence of notes in scores, was also challenged, and two principles of arranging musical objects became important: indeterminism and loops.

Similar developments took place in all Western art and reflected a more and more complex industrial and urban world, where well-arranged entities have been replaced by a confusing system of an immense number of processes and activities that define life. There is little left from the once granted principle of cause and effect – events seem to happen simultaneously, randomly, and chaotic. In response, new forms of art have been introduced, which reflect on our contemporary situation.

---

<sup>50</sup> proclaimed in 1913 by the Futurist painter Luigi Russolo.

<sup>51</sup> R. Kostelanetz, John Cage and Richard Kostelanetz: A Conversation About Radio, *The Musical Quarterly* 72.2 (1986), pp. 216–227.

<sup>52</sup> “objets sonores” and “objets musicaux”; cp. P. Schaeffer, *Traité des Objets Musicaux*, Seuil, Paris, 1966.

<sup>53</sup> Schaeffer, quoted after C. Joel, *The great opening up of music to all sounds. Electric sound : the past and promise of electronic music*. Upper Saddle River, NJ, Prentice Hall, 1997

### 4.3.2 Aleatoric Processes

Digital systems are an ideal tool for handling media objects, and allow a wide range of processes on them. Implementing random control mechanisms takes the idea of fragmentation one step further toward the rearranging of output in real time. The artist thereby defines the rules and the possible media content, but transfers the generation of the final output to a machine, and in the case of aleatoric pieces even to random factors. “Christina Kubisch’s Clock Tower Project, drawing from this tradition, also relies on chance: the position and intensity of the sun, mediated by a computer program, determine the sequencing of tones in the compositions; a passing cloud changes everything.”<sup>54</sup>

The 3D sound fragments, circling in chaotic ways around the listener, are the aleatoric elements in *Acoustic Chase*. They point to the disperse fragments of reality that we are exposed to in our lives.

### 4.3.3 Loops

Loops on the other hand, express not so much the fragmentation of the world, but have a meditative or rhythmic connotation. In the installation “He Weeps For You,”<sup>55</sup> Bill Viola creates a secluded space that reminds of ancient times, or of the beginning of everything. By focusing the attention on a water drop, which regularly falls on an amplified drum, he gives us “experience of continuity, constancy, and the connections between micro- and macrostructures. He produces a space of experience based on total perception. In so doing, he addresses “archetypal” notions like the

---

<sup>54</sup> [http://www.massmoca.org/visual\\_arts/sound\\_art.html](http://www.massmoca.org/visual_arts/sound_art.html)

<sup>55</sup> <http://www.cnca.gob.mx/viola/2.html>



inexorable cycle of renewal, and produces a situation of perception that is directed towards primeval forms and patterns for conceiving human life.”<sup>56</sup>

In Acoustic Chase, the motif of the loop expresses the restlessly following sound teaser that approaches the listener in circles. These geometric loops are irregular. They start slowly as long as the acoustic chaser is far away, but as the sound comes closer, it gains speed and flutters wildly around the person’s head. For the time the listener stays in the acoustic environment an escape is impossible – the sound will always follow in enduring circles, chasing the human being.

---

<sup>56</sup> H. Helfert, Technological Constructions of Space-Time Aspects of Perception  
([http://www.medienkunstnetz.de/themes/overview\\_of\\_media\\_art/perception/1/](http://www.medienkunstnetz.de/themes/overview_of_media_art/perception/1/). Rev. 2004-04-30)

## 5. Evaluation – Opinion of Artists and Curators

The technical platform is working and provides an acoustic environment, reacting to body movements, as described in Chapter 3. The rotational position tracking is stable with a yaw drift less than 5°/s. The pitch and roll tracking has a good long-time stability because of the usage of accelerometers that reference to the gravity vector. The position sensing continually reset to zero with a small time constant of 5 sec. This allows tracking of rapid movements only. This is compatible with the current implementation of the sound movement, as it generates fast motions of the acoustic object, with velocities much higher than the “phantom” speeds resulting from drift and low-pass resetting of the sensor signals.

Acoustic Chase was evaluated by senior professional artists and curators in a formal critique session.<sup>57</sup>

The participants felt the confrontation through the acoustic stimuli and moved away when the sound approached closely. The current system realized the concept of motion tracking and spatial sounds that approach and incite the human being to action.

From an artistic point of view, the use of the casino sound in combination with the described interaction principle established a coherent experience that can transport the hectic and up-heated situation that is typical for a gambling hall.

During the review session, a number of suggestions were made regarding methods to improve the artistic quality of the piece, add to the user experience, and enhance its expressive power. In particular, the audio content, the sound quality, the interaction concept, and the appearance of Acoustic Chase were discussed:

---

<sup>57</sup> This panel consisted of Bill Arning (curator of the List Visual Arts Center), Chris Csikszentihalyi (artist and professor of Media Arts and Sciences), John Maeda (artist and professor of Design and Computation), Joan Jonas (artist and professor of Visual Arts), Winnie Wong (Associate Director of Art Interactive).

The gambling hall-sound that Acoustic Chase works with right now is very dense and bundles a lot of acoustic events in one stream. This can make it difficult for the listener to locate the sound within the acoustic environment and establish a relationship beyond the pure volume and hecticness of the sample. Two directions might improve this situation:

One is to concentrate on “cleaner” sounds that result from one source only. This will rely on the concise strength of a carefully selected sound instead of accumulating a multitude of noises playing simultaneously to an overwhelmingly dense layering. A simple example illustrates in which direction this could develop; the buzzing of a bee. Many people develop strong emotions of anxiety when they hear an aggressive bee circling closer and closer.

Another way to dissolve the very dense cluster of the current sample is to distribute the different involved sounds to separated objects inside the acoustic environment instead of bonding them in one channel. These objects then move independently and give richer impressions with more spatial details.

The quality of the acoustic spatialization still lacks quality, especially in the forward backward direction and in displaying the distance of the sound. Better sound rendering will improve the effect of the acoustic chase a lot.

The interaction of Acoustic Chase doesn't become clear right from the start, and it takes a certain openness of the user and usually a short approaching phase to understand and get involved in the interaction process. One thing that can be improved here is to tighten the linking of the sound actions to the body movements and especially shorten the reaction time upon human motions. This will make it easier to understand the system's reactions. It will also reduce the feeling that the simulated environment is very “floaty” – that is, lacking fixed references to the user.

For future versions of Acoustic Chase, it might also be desirable to make the interaction concept more complex and add elements of evolution and discovery into it. This will maintain the interest of the listener for a longer period of time and

expand the expressive possibilities of Acoustic Chase beyond a fixed system of one specific sound interacting in one specific way.

The visual sense, which currently is not actively served by Acoustic Chase, could be used in future versions to supplement the heard content. For example adding a specifically designed photography or video projection can provide references for the acoustic content and expand its meaning.

## **6. Future Work**

These results and the suggestions of the critique panel motivate further research and exploration. Some main lines along which future work can be aligned regard interaction principle, acoustic content, and technical platform.

The current implementation of a sound approaching in circles is a starting point for more detailed and complex interaction patterns. One interesting direction would be to add software modules that analyze the body movements more carefully, such as gesture recognition, and supply the basis for an interaction scheme that goes beyond direct approaching and repelling.

This also will allow implementing acoustic content with a storyline that evolves during interaction, making the experience more interesting and diversified.

Another promising interaction concept for Acoustic Chase is to implement verbal instructions, spoken by the machine. These direct commands try to animate the participant, and they get louder and more demanding if the person resists. Thereby one specific command may be repeated several times. When the person follows each of them, the machine's voice changes to a mellifluous tone. If the human being and the machine find a common speed, a calm and meditative atmosphere arises for a little while, before the commands change.

This concept could be realized by commands referring to head gestures only. Below, two ideas are provided, how such a system could be implemented in the behavior system.

### **6.1 Gesture Recognition**

A high-level signal analysis module provides the recognition of head gestures. Here we focus on gestures that can be described by only using angular degrees, which covers the common human head gestures:

- Rotation around yaw axis:

- turn left
- turn right
- head shaking ('no')
- Rotation around pitch axis:
  - look up
  - look down
  - head nodding ('yes')
- Rotation around roll axis:
  - tilt head to the right
  - tilt head to the left
  - tilt head left and right ('maybe')

A variety of methods for recognizing head gestures have been presented in literature, most notably using Hidden Markov Models,<sup>58</sup> neuronal networks,<sup>59</sup> and Fuzzy Logic.<sup>60</sup>

## 6.2 Finite States

Finite state machines are a powerful method for realizing complex behaviors that run through several consecutive stages. Such systems consist of different states and transitions between them. Beginning at a start state, certain events trigger transitions

---

<sup>58</sup> C. Morimoto, Y. Yacoob, and L. Davis, Recognition of Head Gestures using Hidden Markov Models in International Conference on Pattern Recognition, 1996, pp. 461-465

<sup>59</sup> J. Tang, R. Nakatsu, A Head Gesture Recognition Algorithm, Proceeding of International Conference on multi-modal Interface, OCT. 14 – 16, 2000, Beijing China.

<sup>60</sup> T. Frantti, S. Kallio, Fuzzy logic aided gesture recognition, KBCS-2002 International Conference on Knowledge Based Computer Systems, 2003, Hongkong, China.

to one of the possible next states; from there other transitions lead to further states, and so on.<sup>61</sup>

Figure 12 shows an example of a state system that plays a sound file saying “yes”, anytime a head nod occurs. Transitions lead from all states to the idle state. These transitions are triggered by inactivity for a period of some seconds. The system remains in the idle state until either an upward or a downward motion is detected. An upward motion leads to an up state. From there the only transition besides the default idle path is to the down state, triggered by a downward motion. When this motion occurs, the sound file is triggered, and the down state is entered. An upward motion leads back to the up position.

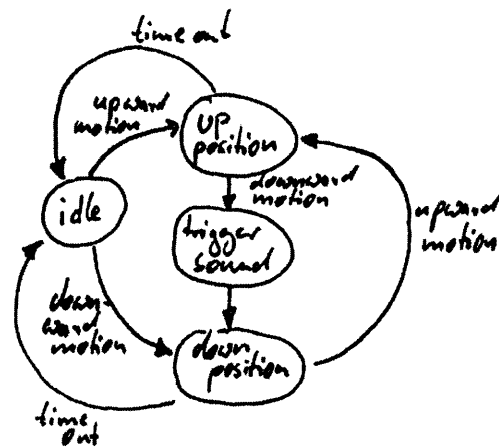


Figure 12: simple finite state system

Another example shows a linear sequence of states that represent spoken instructions to be followed by the person. When a state is entered, an audio sequence is played, telling the listener what he or she has to do. If the pattern analysis system detects the required action, the next state is entered, which asks to do another movement. Any other action triggers the reentering into the current state, which will

<sup>61</sup> R. C. Martin, UML Tutorial: Finite State Machines, Engineering Notebook Column, C++ Report, June 98

repeat the demanded instruction. When the person remains inactive for a specific time, the system also loops the present state.

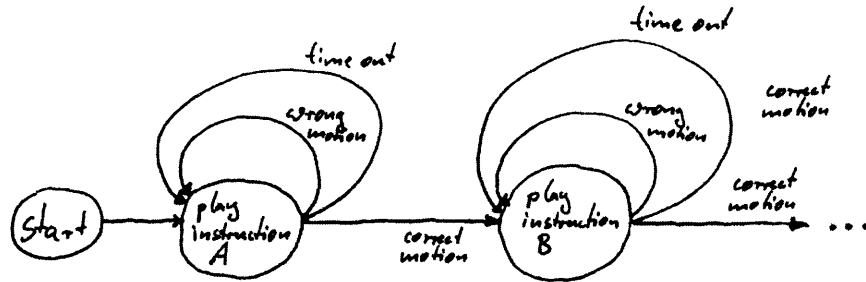


Figure 13: chain of linear states



## **7. Conclusion**

It was a short and soft sound from far away that turned out to be one of the eeriest noises I ever should hear in New York. At first I nearly didn't notice it, some crackling and a deep swoosh. As it continued for 10 seconds or so, it drew my attention, and I was wondering what it might be.

I was sitting near a window at my office downtown. I went there very early on this September morning, before the attack happened.

Shortly after that sound, my girlfriend called me to tell that the first tower came down. This immediately turned my small sound experience, that at first seemed unimportant, into a message of immense suffering.

I ran out on the street to get away from the scene, and dived into a sea of noise from police sirens, shouting people, and stuck traffic.

## 8. References

- B. Arnald, G. Dumon, G. Hégro, N. Magnenat-Thalmann, D. Thalmann, Animation Control with Dynamics, in: Magnenat-Thalmann N, Thalmann D (eds) State-of-the-Art in Computer Animation, Springer, Tokyo, 1989, pp.113–124.
- D. R. Begault, 3-D Sound for Virtual Reality and Multimedia, Academic Press, Cambridge, MA, 1994
- J. Blauert, Spatial Hearing, 2nd ed., J. S. Allen, trans., MIT Press, Cambridge, Mass. 1997
- M. La Cascia, S. Sclaro, and V. Athitsos. Fast, reliable head tracking under varying illumination: An approach based on registration of texture-mapped 3D models, IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(4): pp. 322–336, April 1999.
- W. L. Chapin, AuSIM Incorporated, et al. (2000): "InTheMix", interactive audio environment, Siggraph 2000 (<http://www.ausim3d.com/InTheMix/history.html>)
- R. Dorobantu, Simulation des Verhaltens einer Low-cost strapdown IMU unter Laborbedingungen, Institut für Astronomische und Physikalische Geodäsie Forschungseinrichtung Sattelitengeodäsie TU-München, IAPG/FESG No. 6, 1999
- G. Föllmer, Audio Art ([http://www.medienkunstnetz.de/themes/overview\\_of\\_media\\_art/audio/21/](http://www.medienkunstnetz.de/themes/overview_of_media_art/audio/21/). Rev. 2004-04-30)
- E. Foxlin, Inertial head-tracking, M.S. Thesis, Dept. of E.E.C.S., MIT, 1993.
- E. Foxlin and N. Durlach, An Inertial Head Orientation Tracker with Automatic Drift Compensation Doe Use with HMD's, Proceedings from VRST'94, Virtual Reality Software and Technology, Singapore (August 23–26, 1994).
- E. Foxlin, M. Harrington, and Y. Altschuler, Miniature 6-DOF inertial system for tracking HMDs, SPIE vol. 3362, Proc. AeroSense '98 Conference on Helmet- and Head-Mounted Displays III, Orlando, FL 1998
- T. Frantti, S. Kallio, Fuzzy logic aided gesture recognition, KBCS-2002 International Conference on Knowledge Based Computer Systems, 2003, Hongkong, China.
- W. Frey, Application of Inertial Sensors and Flux-Gate Magnetometers to Real-Time Human Body Motion Capture, Master's Thesis, Naval Postgraduate School, Monterey, California, September 1996.
- R. Frieling, Reality/Mediality Hybrid Processes Between Art and Life ([http://www.medienkunstnetz.de/themes/overview\\_of\\_media\\_art/performance/23/](http://www.medienkunstnetz.de/themes/overview_of_media_art/performance/23/). Rev. 2004-04-29)
- W. G. Gardner and K. D. Martin, HRTF measurements of a KEMAR, J. Acoust. Soc. Am., 97(6), 1995, pp. 3907–3908.

- D.O. Gorodnichy, S. Malik, and G. Roth. Affordable 3D face tracking using projective vision. In Proceedings of International Conference on Vision Interface (VI'2002), pp. 383–390, 2002.
- M. Haller, D. Dobler, P. Stampfl, Augmenting the Reality with 3D Sound Sources, ACM SIGGRAPH 2002
- B. Hartmann: How We Localize Sound, in Physics Today on the Web, 1999 (<http://www.aip.org/pt/nov99/locsound.html>)
- H. Helfert, Technological Constructions of Space-Time Aspects of Perception ([http://www.medienkunstnetz.de/themes/overview\\_of\\_media\\_art/perception/1/](http://www.medienkunstnetz.de/themes/overview_of_media_art/perception/1/). Rev. 2004-04-30)
- J. Huopaniemi, M. Karjalainen: HRTF Filter Design Based on Auditory Criteria, NAM, Helsinki, 1996
- C. Joel, The great opening up of music to all sounds. Electric sound: the past and promise of electronic music. Upper Saddle River, NJ, Prentice Hall, 1997
- R. Kostelanetz, J. Cage and R. Kostelanetz: A Conversation About Radio, The Musical Quarterly 72.2 (1986), pp. 216–227.
- D. MacKenzie, Inventing Accuracy: A Historical Sociology of Nuclear Missile Guidance, MIT Press, 1991
- R. C. Martin, UML Tutorial: Finite State Machines, Engineering Notebook Column, C++ Report, June 98
- K. Meyer, H.L. Applewhite, and F.A. Biocca(1992). A Survey of Position-Trackers. Presence: Teleoperators and Virtual Environments. 1 (2) (spring 1992), pp. 173–200.
- B.C.J. Moore, An introduction to the psychology of hearing, 5th edition. University of Cambridge. 2003
- L.-P. Morency, A. Rahimi, N. Checka, and T. Darrell. Fast stereo-based head tracking for interactive environments. In Proceedings of Conference on Automatic Face and Gesture Recognition, pp. 375–380, 2002.
- C. Morimoto, Y. Yacoob, and L. Davis, Recognition of Head Gestures using Hidden Markov Models in International Conference on Pattern Recognition, 1996, pp. 461–465
- J. Rekimoto. A vision-based head tracker for fish tank virtual reality: VR without head gear. In Virtual Reality Annual International Symposium (VRAIS '95), pp. 94–100, 1995.
- R. M. Rogers, Applied Mathematics in Integrated Navigation Systems, American Institute of Aeronautics and Astronautics, AIAA, Direction Series, 2000

- A. Shkel and R. T. Howe, Polysilicon Surface Micromachined Rate Integrating Gyroscopes. UC-Berkeley Office of Technology and Licensing. Case Number B99-077
- A. Shkel, R. Horowitz, A. Seshia, and R. T. Howe, Dynamics and Control of Micromachined Gyroscopes, in The American Control Conference, San Diego, CA, June 1999
- M. Sester, Access, presented at Siggraph 2003
- D. G. Stork (Ed.), HAL's Legacy: 2001's Computer as Dream and Reality, MIT Press, 1997
- J. Tang, R. Nakatsu, A Head Gesture Recognition Algorithm, Proceeding of International Conference on multi-modal Interface, OCT. 14-16, 2000, Beijing China.
- D. Thalmann, Robotics Methods for Task-level and Behavioral Animation, in: D. Thalmann (ed) Scientific Visualization and Graphics Simulation, John Wiley, Chichester, UK, 1990, pp.129-147.
- D. H. Titterton, J. L. Weston, Strapdown Inertial Navigation Technology (IEE Radar, Sonar, Navigation and Avionics, No 5)
- G. Welch, G. Bishop, An Introduction to the Kalman Filter, TR-95-041, Department of Computer Science, University of North Carolina at Chapel Hill, updated in 2002
- E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, Localization using nonindividualized head-related transfer functions, J. Acoust. Soc. Am., 94(1), 1993, pp. 111–123.
- C. Youngblut, R. E. Johnson, S. H. Nash, R. A. Wienclaw, and C. A. Will, Review of Virtual Environment Interface Technology , Institute for Defense Analyses – IDA, Paper P-3186, March 1996
- Z. Zivkovic and F. van der Heijden, A stabilized adaptive appearance changes model for 3D head tracking. In IEEE ICCV Workshop on Recognition,