

3-D Audio in Mobile Communication Devices: Methods for Mobile Head-Tracking

Christoph Pörschmann

Institute of Communications Engineering
Cologne University of Applied Sciences
Betzdorfer Str. 2, 50679 Köln, Germany

phone +49 221 8275 2495, email: Christoph.Poerschmann@fh-koeln.de
www: www.fh-koeln.de

Abstract

Future generations of mobile communication devices will serve more and more as multimedia platforms capable of reproducing high quality audio. In order to achieve a 3-D sound perception the reproduction quality of audio via headphones can be significantly increased by applying binaural technology. To be independent of individual head-related transfer functions (HRTFs) and to guarantee a good performance for all listeners, an adaptation of the synthesized sound field to the listener's head movements is required.

In this article several methods of head-tracking for mobile communication devices are presented and compared. A system for testing the identified methods is set up and experiments are performed to evaluate the pros and cons of each method. The implementation of such a device in a 3-D audio system is described and applications making use of such a system are identified and discussed.

Keywords: Spatial Audio, Virtual Environment, Head Tracking

Digital Peer Publishing Licence

Any party may pass on this Work by electronic means and make it available for download under the terms and conditions of the current version of the Digital Peer Publishing Licence (DPPL). The text of the licence may be accessed and retrieved via Internet at <http://www.dipp.nrw.de/>.

1 Introduction

Modern mobile communication devices are no longer used as simple mobile phones but are enhanced to serve as mobile multimedia platforms for multiple applications. Therefore, additional features to telephony become more important. Many devices currently available on the market already feature audio players for MP3, permitting reproduction of high quality audio. Furthermore, the devices allow the download of music files via specific portals, which is considered an important market for the provider. Other manufacturers equip their mobile phones with an FM-Radio so that they can as well be used for stereo music reproduction. In the future the distribution of video via download or broadcast (e.g. DVB-H) will gain importance as well. For an overview over the actual broadcasting services please refer to [WW06].

Due to the relatively low audio quality of the loudspeakers integrated in mobile phones the music reproduction focuses on presentations over headphones. However, audio reproduction via headphones is often identified as being artificial and unnatural because sound sources are localized inside the head. To overcome such disadvantages binaural technology can be applied. During the last decades headphone-based virtual auditory environments have become a field of intense scientific research [WWF88, Beg94, Bla97]. Several psychoacoustic experiments showed that virtual auditory sound-source presentation can yield perceptions that come close to real conditions [WK89, Bro95].

For binaural music reproduction in a mobile communication device both channels of a stereo audio signal have to be filtered in a way that the sound pressure in

the listener's ear canals is synthesized to be equivalent to the intended situation, for example to a loudspeaker listening scenario. The head-related transfer function (HRTF) of the corresponding direction represents the acoustic transfer from a sound source in the environment to the listener's ears for one specific incidence direction. The impact of the geometry of the listener's torso, head and ears are considered in the HRTF as interaural level and time differences as well as a spectral filtering. Furthermore, room reflections are taken into account which can either be gained from measurements in a real room or by adding a simulated binaural room impulse response of a typical listening environment. Such a technology has been extensively described in the literature [Bla97], and several products applying this technology are available on the market (e.g. Lake Dolby Headphones, AKG Hearo Audio-sphere 999. Please refer to [Beg94] for an overview on the development of spatial 3-D audio and to [BLSS00] for a detailed description of techniques for measuring HRTFs.

However, several problems can be observed when applying this technology. Especially front-back confusions occur in sound-source localization when non-individual impulse responses are being used and head movements which enhance the localization of sound sources are not being considered [BWLA00]. In order to make the listener be immersed in the virtual auditory scene an adaptation of the sound field to the listener's head movements is required. Studies [Wen96a, Wen96b] demonstrated that enabling head motions dramatically improved the localization accuracy of virtual sources synthesized from non-individualized HRTFs. In particular, average rates of front-back confusions decreased from about 28% for static localization to about 7% when head motion was enabled. A comparison of advantages of static and interactive systems can be found in [Pel01].

Several systems have been set up which allow the sound field reproduction via headphones to be adapted to the listener's head movements. Only some of them are summarized here.

Already in 1988 the so-called Convolvotron was implemented [WWF88]. This system allows the auralization of up to four sound sources and reflections. The head position is measured by means of a Polhemus headtracking device and the sound field is adapted to the listener's head movements by convolving the audio signals with the appropriate head related impulse response (HRIR).

In a co-operation between the Studer Professional Audio AG and the German Institut für Rundfunktechnik (IRT) a system called BRS (Binaural Room Scanning) was developed [MFT⁺99]. By means of a head-tracking device (e.g. Polhemus Fastrak) the head orientation in the horizontal plane is measured and the sound field is adapted to these head movements by convolving the audio signal with the appropriate stored binaural impulse response. An auralization of up to 5 sound sources and a length of the room impulse responses of up to 85 ms allow not only the auralization of the direct sound component but also of the early reflections.

As a commercial product applying this technology Beyerdynamic has developed in co-operation with SonicEmotion a system called "Headzone" which uses headphones with an integrated head-tracking device in order to adapt the presented sound source direction to the head rotations. The system utilizes an ultrasonic head-tracking device. As fixed reference position the system's base station is used [AG06].

In [BLSS00] a system capable of synthesizing up to four virtual sound sources and up to 63 reflections of one sound source in real time is described. The so-called SCATIS system allows the consideration of reflections up to the 3rd order. The position and the orientation of the listener are tracked by a Polhemus Fastrak device and also the sound-radiating objects can be moved freely in the virtual environment. The sound-field parameters depend on the actual position of the sound sources and the receiver in the virtual room. The calculation of the sound field is based on the rules of geometric acoustics. The mirror-image method is applied and each reflection in the virtual environment is represented by a mirror image sound source [KSS68, LB92]. The auralization hardware is based on 80 Motorola DSPs 56002. An enhancement of the SCATIS system (IKA-Convolver) allows up to eight sources to be convolved with long binaural room impulse responses selected depending on the head orientation [Nov05].

Research was also conducted in adapting 3-D audio to mobile communication devices, for example by Sensaura [Sci02] or by QSound [QSo06]. Actually there are several mobile phones on the market which incorporate 3-D sound. A reference design for a UMTS/EDGE for running Linux has been equipped by the developing companies Infineon Technologies, Samsung Electronics, Emuzed und Trolltech with the 3-D audio solution from QSound [Inn05]. This makes

it possible for a manufacturer of mobile communication device to get access to 3-D sound engines easily. Principally, mobile communication devices are perfectly suited as platform for the described technology. Current and future mobile communication devices will run multiple applications reproducing headphone-based high-quality audio. Furthermore, mobile devices offer the signal processing capabilities required for binaural real-time processing of the audio signals. However, the signal processing and memory requirements of the algorithms need to be scaled, the influence of time-lags needs to be considered and finally adequate sensors and algorithms need to be found which allow the measurement of the listener's head position and orientation.

At this point the lack of adequate head-tracking devices hinders the integration of a head-tracked 3-D audio system into mobile communication devices. [MAB92] investigated the different technologies for position tracking devices. However, as many of these devices are not suited for mobile applications this paper describes research initiated in order to develop head orientation measuring techniques suitable for such applications requiring robust, cheap and small sensors.

The first section focuses on the requirements for head-tracked 3-D audio regarding the orientation measurement techniques from a psychoacoustic starting point. They form the basis for the following sections, where different sensor technologies for measuring the head orientation are investigated and compared. In a next step the set-up of a mobile head-tracking device and its integration into a 3-D audio system are described. Finally, potential applications of mobile devices making use of head-tracked 3-D audio are identified.

2 Requirements for head-tracked 3-D audio

This section describes the psychoacoustic requirements for the head orientation measurement regarding different aspects of psychoacoustic perception. At first, the limits of precision of the human localization capabilities are discussed, and it is determined which degrees of freedom need to be tracked for the desired types of applications. Subsequently, critical limits concerning the system latency and frame rate are derived in order to guarantee an adequate adaptation of the sound field to the listener's head movements. Fi-

nally, the maximal velocities of the head movements are determined in order to select sensors with sufficient capabilities.

2.1 Localization

Human sound source localization can be performed very exact in the horizontal plane [Bla97] because it is not only based on monaural but mainly on binaural cues (Interaural Time Differences, Interaural Level Differences). The determination of the elevation is based on monaural cues only, mainly on the spectral filtering caused by the geometry of the outer ear with head and torso and is thus less exact. The localization blur, which can be defined as the amount of a sound source's displacement that cannot be perceived as a change in the position of an auditory event, is minimal in the horizontal plane for frontal directions and has been measured in different experiments to be about 1° . For a comparison of the experiments in literature refer to [Bla97]. The localization blur in the median plane is significantly higher. [Bla97] summarizes different measurements which show results between 9° and more than 20° depending on the set-up and the incidence direction. For an adequate adaptation of the sound field to the listener's head movements the sensor resolution should be at least in the range of the localization resolution.

However, this criterion is not the only relevant one for a head-tracking sensor, as the human absolute localization errors are much higher than the localization blur. Thus a slight permanent error in the head orientation measurement has less influence on the human perception than one varying in time.

An important question regarding the distinction between different sensors is how many degrees of freedom need to be tracked by the device. Principally, in order to allow a subject to move freely in the virtual world a consideration of all six degrees of freedom, three translational and three rotational are required.

[MFT⁺00] investigated the localization performance of head-tracked audio in the median plane. The elevation of eleven loudspeakers in the median plane was determined by the subjects both enabling and disabling the vertical head-tracking. The investigations showed that the consideration of the vertical head movements did not significantly improve the localization performance. In the experiment, the frontal sources appeared to be elevated, and this elevation persisted when using the additional vertical head-tracking.

2.2 Distance perception

Another aspect of the spatial perception of sound sources is distance estimation, which is based on different factors. According to [Bla97, Col63, Law73], in free space sound is attenuated by 6 dB for each doubling of distance. At distances below 4 m spectral cues influence the distance perception due to the curvature of the wavefronts. At distances larger than 15 m dissipation has effect on the air propagation. High frequencies are affected more than low frequencies by this effect which leads to a changed timbre of the sound depending on the distance. Furthermore, the ratio of the sound pressure levels affects the distance perception as well [MK75].

To consider translational movements appropriately the movements in all three translational degrees of freedom need to be measured. However, for mobile applications it has to be questioned whether an adaptation of the sound field to the listener's translational movements and thus an altered distance to the sound sources is desirable. When a person uses his mobile 3-D audio application, for example while walking, he does not want to perceive the sound sources approaching or moving away. In such situations it makes no difference whether the translational movements are considered or not. Hence, in many situations a consideration of translational head movements is not required or desired and does not have to be regarded in a first system design.

2.3 System latency

The system latency necessary for an undisturbed perception of the virtual environment can be described as the time difference between the initiation of a head movement and the change of the audio signals at the headphones due to the recalculation of the sound field. An increase of the system latency causes a degradation of the sound presentation's responsiveness. It is obvious that several factors influence the system latency: The time delay caused by the head-tracking device, the update rate for the recalculation of the sound field, the required time for the digital signal processing including the D/A conversion of the audio signals. An update rate which is not adequately high results in a missing smoothness of the system [Pe101].

[Wen96c] determined the upper latency limit by measuring the "Minimum Audible Movement Angle" (MAMA, [Per82, PT88]). She found that a maximal total system latency of less than 58 ms at an angular

velocity of 360°/s is not perceived by the listener. This result matches tests performed with the SCATIS Virtual Auditory Environment Generator at the Institute of Communication Acoustics in Bochum. In their investigations [BLSS00] no perceptible disturbances at latencies of less than 60 ms and frame rates of 60 Hz were recognized.

Investigations on the dynamic aspects of auditory virtual environments carried out by [San96] showed that update rates below 20 Hz and latencies of 96 ms and above have a significant degrading effect on both time and accuracy of sound source localization under free-field conditions.

[Wen96a, Wen96b, Wen01] examined that head movements which listeners use to aid localization may be as fast as about 175°/s (in particular, left-right yaw) for short time periods (e.g., about 1200 ms). Our own measurements which were performed with one of the gyroscope sensors described later confirmed the results [Len02]. Here, maxima of between 150 and 200°/s were observed in a series of measurements.

2.4 Summary

It can be concluded that the measurement of head movements is required to increase out-of-head localization and to prevent front-back confusions. It seems sufficient at least for low-cost applications to track only horizontal head rotations and to guarantee an update rate of about 60 Hz and a system latency of less than 58 ms. Concerning the latency, it should be noted that not all the delay is induced by the head orientation measurement device, but that the auditory signal processing and the D/A converting of the audio signals unavoidably cause a delay as well.

3 Sensor technologies

After having identified the psychoacoustic requirements for a mobile head-tracking device, in this section different sensor technologies which allow the orientation of a listener to be measured are presented. In particular the techniques for head orientation measurements are discussed which are suitable for mobile applications. As the sensors shall later be integrated into a set of headphones they need to fit in size and dimensions for this task. Sensor techniques that require sensors which are too large, too expensive or which consume too much electrical power can hardly be used for mobile applications and are not further investigated.

Furthermore, sensors which require an external reference to be placed somewhere in the environment are not regarded here, even though such a reference could be theoretically integrated in a mobile communication device. Such systems which, for example apply a sender creating an electromagnetic, optical, or acoustical field (e.g. ultrasound) and a receiver detecting the different components of this field would either require major changes in the concept and design of the mobile communication device or be quite impractical in use. For each identified sensor technology a measurement system is set up in order to evaluate the sensor's performance and to compare advantages and disadvantages of each technology.

3.1 Acceleration sensors

Acceleration sensors generally consist of a reference mass suspended by compliant beams which are being anchored to a fixed frame. Generally, the motion of the mass can be detected in all three degrees of freedom. However, typically such sensors are designed in a way that motions are only measured in one or optimal two directions.

A variety of transducer mechanisms with differing specifications is available on the market, since they are used in a wide range of applications. A good overview can be found in [YAN98]. Only some of the types are briefly described here.

Piezoresistive sensors incorporate piezoresistors in their suspension beam and are based on the idea that relative movements between the frame and the reference mass alter the length of the suspension beam. This way the resistivity of the piezoresistors embedded in the suspension beam is changed.

Capacitive devices are based on the principle that a capacitance between the reference mass and a fixed conductive electrode is changed by external accelerations. This capacitance is then measured using electronic circuitry.

Tunneling devices use a constant tunneling current between one tunneling tip (attached to a movable microstructure) and its counterelectrode to sense displacement. The first resonant accelerometers were fabricated using quartz micromachining. Silicon resonant accelerometers are generally based on transferring the reference mass's inertial force to axial force on the resonant beams and hence shifting their frequency. So-called thermal devices are based on thermal transduction. One type uses the principle that the tempera-

ture flux from a heater to a heat-sink plate is inversely proportional to their separation. Hence, by measuring the temperature using thermopiles, the change in separation between the plates which depends on acceleration can be calculated.

Combining two acceleration sensors which are positioned in parallel at a defined distance l allows the rotation of one axis of the object to be measured. Each acceleration sensor measures the momentary acceleration:

$$\vec{a}(t) = \frac{d^2}{dt^2} \vec{s}(t) \quad (1)$$

The angular acceleration can be calculated from the output signals $a_1(t)$ and $a_2(t)$ of the acceleration sensors which are positioned in parallel:

$$\ddot{\varphi}(t) = \frac{d^2}{dt^2} \varphi(t) = \frac{a_1(t) - a_2(t)}{l} \quad (2)$$

A twofold integration over time leads to the instantaneous angle. Starting from an inertial situation, $\dot{\varphi}(0)$ and $\varphi(0)$ can be set to zero as border conditions.

Several problems arise when using acceleration sensors for measuring head rotation:

- Calibration of the sensors needs to be performed
- Changed distance between the sensor e.g. due to bending of the headphones strongly affects the performance
- Acceleration due to both gravity and translational movements are much larger than acceleration due to rotation.
- Double integration of the difference of the sensor signals is required: Even small measurement errors have great effect on the calculated angle.

For evaluating the performance of acceleration sensors for head-tracking in a test system, two ADXL 202E acceleration sensors from Analog Devices were used for the realization of the test system. For the measurements a PCI-Base 300 PC-card from BMC with on board 16 bit A/D converters was used which allows to sample the analog output values of the sensors at a sufficiently high rate. Principally the card can be used at sampling rates of up to 100 000 measurements/s.

Experiments performed by [Len02] showed that head orientation measurement with this sensor was hardly possible due to the problems listed above. Adequate calibration of both sensors in order to receive constant values for the orientation could not be realized. So in

all measurements the errors of the calculated head orientation became unacceptably high already after a few milliseconds. With sensors calibrated in a sufficiently exact way and very high measurement accuracy adequate performance can be reached in principle, but for applications in low-cost consumer products such a solution is critical.

3.2 Magnetic field sensors

Magnetic field sensors are some of the oldest, having been used already for over 2000 years for navigation and direction finding.

Nowadays, these are still the main domains of their application, but new applications have emerged.

The magnetic field of the earth which functions as an external reference can be roughly described as a field created by a bar magnet which runs from the south pole to the magnetic north pole. Thus, the direction of the magnetic flux and the intensity of the magnetic field lines depend strongly on the location on the earth. Close to the equator the lines run parallel to the surface but near the poles the lines are nearly vertical to the surface.

Typically magnetic field sensors measure two horizontal components of the magnetic field, $H_x(t)$ and $H_y(t)$. The actual angle can then be calculated according to

$$\varphi(t) = \arctan \frac{H_y(t)}{H_x(t)}. \quad (3)$$

However, depending on the sign of H_x and H_y , an angle of 180° has to be added to the calculated angle. This well-known adjustment is incorporated into the "arg" function.

The equation given above can only be applied when the sensor is positioned horizontally. Certainly this limitation cannot be accepted for head-tracking applications. To yield reliable results, errors caused by the non-horizontal positioning of the sensor need to be compensated for. Therefore, an additional tilt sensor which measures roll θ and pitch ϕ is required. But knowing roll and pitch is only half the solution, in addition the system must rely on the magnetic field of all three axes (X, Y, Z). Applying equation 4 the azimuth direction of the sensor can be calculated.

$$\varphi = \arctan \frac{H_{yh}}{H_{xh}} \quad (4)$$

with

$$H_{xh} = H_x \cos(\phi) + H_y \sin(\theta) \sin(\phi) - H_z \sin(\phi) \cos(\theta)$$

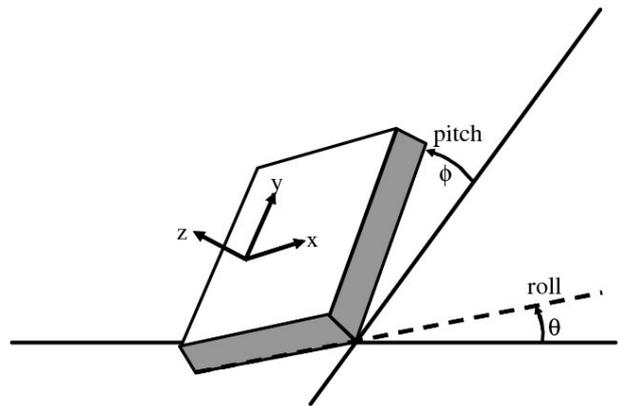


Figure 1: Compass tilt referenced to horizontal plane

$$H_{yh} = H_y \cos(\theta) + H_x \sin(\theta)$$

Several methods for measuring the earth's magnetic field can be applied. Only a very brief description is given here. An overview over different types and applications for magnetic field sensors can be found, for example, in [CBSS98]. Mainly three sensor types are applied for earth-field measurements: fluxgate magnetometers, magnetoinductive and anisotropic magnetoresistive measurement techniques.

Fluxgate sensors are based on the principle that an alternating magnetic field is created in one coil at a certain frequency (e.g. 10 kHz) and this magnetic field is measured with a second coil. Both coils are wrapped around a common high-permeability ferromagnetic core. The measured signal is affected by any change in the permeability and the influence of the earth's magnetic field can be determined either from the change in the core's permeability or from the changes in the core's saturation.

A sensor applying the magnetoinductive principle consists of a single winding coil on a ferromagnetic core that changes permeability within the earth's field. The sense coil is the inductance element of an L/R relaxation oscillator which changes its frequency depending on the magnetic field.

Magnetoresistive sensors consist of a nickel-iron thin film which changes its resistance by 2-3 % in the presence of a magnetic field. Often four of those sensors are connected in an electric circuit (Wheatstone bridge). This way both direction and orientation of one dimension of the magnetic field can be measured. Such sensors can be manufactured on silicon wafers and included in integrated circuits. Thus, they are very cost-effective to produce.

For a more detailed description of the technologies please refer to [CBSS98] and [Len90]. Investigations on the application of low-cost compass systems can be found in [Car00].

In the test system an anisotropic magnetoresistive sensor (Philips KMZ 52) was used. The sensor allows two vertical components of the earth's magnetic field to be measured. Although no tilt compensation was performed the sensor could be used to investigate robustness and accuracy of the technology.

The sensor measurements were performed by [Len02]. The sensor was positioned horizontally on a plate; its output was measured by means of an oscilloscope. The results show an average error of about 2.3° and a maximum of 5.2° . These errors were mainly caused by the remaining tilt of the sensor due to the non-orthogonality of the sensor coils. This coincides with the specifications provided by the manufacturer in which an error due to non-orthogonality of the sensor of about 2° is noted.

An advantage of magnetic field sensors for head-tracking devices is that they are not susceptible to any kind of drift because the earth's magnetic field serves as an external reference. But as for acceleration sensors, several sources of errors can be identified. Magnetic field sensors are principally susceptible to distortions caused by the environment (e.g. metallic surfaces, electrical machines). Furthermore, in order to reliably determine the orientation, all three vectors of the magnetic field and additionally the tilt need to be measured. If the influence of the vertical component is not compensated for, as described above, the system becomes very vulnerable to head movements which are not performed in the horizontal plane. This is due to the fact that the vertical component of the earth's magnetic field can - depending on the position - be significantly larger than the horizontal component.

3.3 Gyroscope sensors

Almost all reported micromachined gyroscopes use vibrating mechanical elements to sense rotation. They have no rotating parts that require bearings, and hence they can be easily miniaturized and batch-fabricated using micro-machining techniques. All vibratory gyroscopes are based on the transfer of energy between two vibration modes of a structure caused by coriolis acceleration. Coriolis acceleration arises in a rotating frame of reference and is proportional to the rate of rotation.

In the following only a very short overview on the different kinds of gyroscope sensors is given. The reader is again referred to [YAN98] for a more detailed description.

Today optical gyroscopes are the most accurate sensors on the market [Law93]. They are based on the principle that a laser ray is reflected many times within an enclosure. If the enclosure rotates, the duration between the laser emittance and reception is altered. In a ring laser gyroscope, the laser beam is guided by mirrors inside the enclosure, in a fiber optic gyroscope by a coil of optical-fiber. However, these types of gyroscopes are much too expensive to be used for low-cost consumer applications. Much cheaper and still adequate for the applications discussed here are different kinds of silicon micromachined vibratory gyroscopes. Several principles of vibratory gyroscopes have been produced, among them tuning forks [HCM⁺95] and vibrating beams [MS94]. A classical example of such sensors with tuning forks is briefly explained here:

The two tines of the sensor are excited in the x-direction (see Figure 2) by electrostatic, piezoelectric or electromagnetic force at the resonance frequency of the tines. If the sensor is then rotated around the z-axis a movement of the tines in the y-direction is induced. This movement is measured either capacitively, piezoresistively or piezoelectrically.

To determine the head movement in the listener's hor-

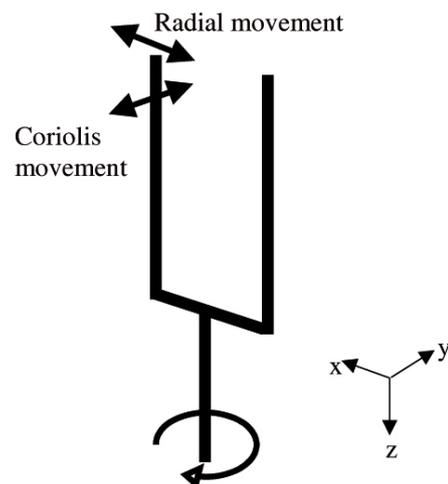


Figure 2: Gyroscope sensor principle. The two tines of the sensor are oscillating in the x-direction at their resonance frequency. Due to the Coriolis force an oscillation in the y-direction is induced when the sensor rotates around the z-axis.

horizontal plane only one gyroscope sensor is required. The actual head orientation can be calculated by

$$\varphi(t) = \varphi_0 + \int_{\tau=0}^t \varphi(\tau) \frac{d\varphi}{dt} \quad (5)$$

φ_0 denotes the head orientation angle at the start of the measurement and should in typical applications be set to zero.

Performance tests were performed by [Len02] with two different sensors (Analog Device ADXRS 300, muRata ENC 03J). The sensor data was fed into a PC by means of the PCI-Base 300 PC-card from BMC. After calibration of the sensor's zero position, the head orientation was determined according to the equations given above.

The dynamic range of both sensors is $300^\circ/\text{s}$. According to the maximal head movement velocity determined above this is sufficiently high for the proposed application. Due to a missing integrated temperature compensation in the muRata ENC 03J sensor, much better performance was obtained with the ADXRS 300. Therefore, in the following this sensor is regarded. For the ADXRS 300 the stability is specified as $0.03^\circ/\text{s}$. The influence of temperature based on deviations caused by a return after temperature excursion is max. $0.1^\circ/\text{s}$ according to the data sheet.

In own additional measurements, drifts of less than $0.1^\circ/\text{s}$ were observed for these sensors, which is sufficiently low for the applications discussed in this paper.

3.4 Comparison of the sensor technologies

Sensors of all three types were tested for their suitability for head orientation measurement. The investigated types were the ADXL 202E acceleration sensor from Analog Devices, the KMZ 52 magnetic field sensor from Philips, and the gyroscope sensors ENC 03J from muRata and ADXRS 300 from Analog Devices. All sensors have analog outputs and thus very low latencies caused by the sensors themselves as any kind of A/D converting and coding of the signals is not required. However, for later applications it has to be guaranteed that the A/D conversion of the signals and the processing in the device are performed sufficiently fast. For achieving more stable results, orientation values measured by the sensor can be averaged. But as averaging increases, the latency of the system a trade-off has to be found between a high precision and a low latency of the sensor.

While any kind of reliable head orientation measurement with the acceleration sensors was impossible due

to the known principal problems listed above, the other sensors showed good results.

The measurement results of the magnetic field sensor were not exact when the rotation did not take place exactly in the horizontal plane. For such a sensor in a head-tracking device the influence of the vertical component needs to be compensated. The compensation causes increased costs for an additional tilt sensor.

The head orientation measurement with the gyroscope sensor was found to be much simpler, as no problems are caused by tilt. However, since those sensors perform only relative measurements of the orientation a slow drift (e.g. due to temperature changes) is observed which needs to be compensated. The lowest drift was observed for the ADXRS 300 because this sensor already features temperature compensation. The investigations and measurements show that gyroscope sensors are an adequate solution for head-tracked 3-D audio systems for mobile applications.

4 System implementation

Based on the results described above a head-tracking device based on a gyroscope sensor was integrated into a headphone. In this section the realization of the head-tracking device, an exemplary integration into an existing 3 D audio system and finally some listening tests will be described:

4.1 Headtracker realization

The sensor was positioned in a way that its measurement axis matches the rotation axis of the listener's head movements. Figure 3 shows the sensor integrated into a pair of headphones (Koss SportaPro). To fix the sensor in its position the free space around it was later clad with casting resin.

4.2 Integration into 3-D audio system

A PC-based 3-D audio system which was developed at the Institute of Communication Acoustics in Bochum, Germany on the basis of the SCATIS demonstrator [BLSS00] was used for the integration.

The IKA-Convolver allows up to eight sound sources to be convolved with long binaural room impulse responses adapted depending on the head orientation [Nov05]. The capabilities of the system are briefly described here. Up to 8 binaural impulse responses



Figure 3: The sensor was integrated into a Koss headphone and (later) clad with casting resin.

can be convolved with incoming audio signals in real-time. Figure 4 shows the general system architecture. The system originally operated with a Polhemus head-tracking device. It allows the measurement of the listener's head to be made at defined frequencies (max. 120 Hz with one sensor, max 60 Hz with 2 sensors) in both the horizontal and median plane. However, in the 3-D audio system applied here only the horizontal orientation data is evaluated for comparison reasons. The system allows the change of the impulse response to be processed in real-time according to the horizontal movements at an update rate of 60 updates per seconds.

The length of the impulse responses is limited only by the calculation power required for the convolution and by the memory resources of the hardware. In the realized system impulse responses of about 2800 taps at a sampling rate of 44.1 kHz are used which correspond to a time length of the impulse responses of about 63 ms. This allows the system not only to reproduce the direct sound but as well to add early reflections. The reflection pattern was chosen according to [Pel00] who derived patterns for the early reflections allowing an exact localization of the sound sources and a good room impression combined with a pleasant timbre of the sound.

The input signals are A/D converted and then fed into a PC via a RME 9636 sound card which is equipped with ASIO drivers. Such drivers have the advantage

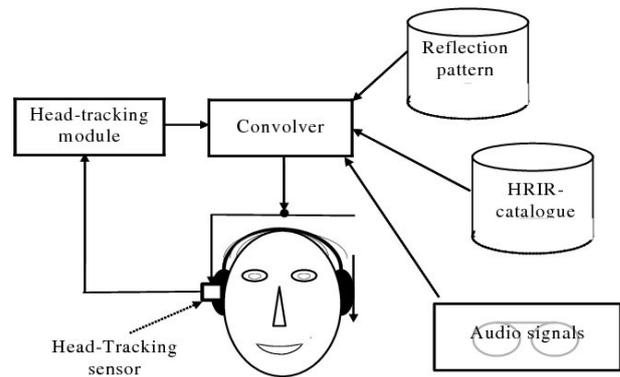


Figure 4: General system architecture of the 3-D audio system. From the sensor signals as input the head-tracking module determines the actual head orientation. The convolver calculates the output signal by convolving the different input channels of the audio signal with the appropriate binaural HRIR for each reflection stored in the reflection pattern database.

that they can be driven at a very low latency. The convolved signals are summed up and auralized via the sound card and the D/A converter. An A/D and D/A converter from RME (AD-I 8 Pro) are used for this purpose. The throughput latency of the audio signals (without any signal processing) can be set to less than 1 ms depending on the sound card's buffer length.

In principle, any set of head related impulse responses (HRIRs) can be used with the system. In the implementation non-individualized HRIRs were used which were measured with an artificial head (head and torso simulator) at the Institute of Communication Acoustics.

The gyroscope sensor (ADXRS 300) was implemented as an additional head-tracking device. This means that the user can switch back to the conventional head-tracker (Polhemus Fastrak) and directly compare the performance of the two head-tracking systems. In this implementation the sensor data was handled in the same way as described above. The analog sensor values were acquired by means of the PCI-Base 300 PC-card and then processed with the 3-D audio system software.

4.3 Return movement

In order to average out errors caused by a sensor drift, a slow return movement is considered. This way, the front direction of the sound source presentation is shifted slowly to the direction the subject is heading to.

Apart from equalizing the measurement errors such a return movement serves to shift the sound sources to the frontal position after the listener has permanently changed his orientation, (e.g. walking around a curve, sitting down after turning around).

4.4 Calibration

In order to calibrate the sensor, different methods can be applied. The easiest one is to store calibration values in the program. Of course any change in the sensor's behavior (e.g. due to temperature shift, aging) influences the shift. A more exact solution is to calibrate the sensor during the set-up of the system or on demand. However, during the calibration phase the sensor has to be kept motionless.

A more advanced solution is to continuously calibrate the sensor during operation. Supposing that the user is not turning permanently round his own axis, the drift of the sensor can be estimated by summing up all the listener's movements over a longer period, and the sensor signal can be compensated appropriately. A similar result can be gained by filtering the sensor signal using a high-pass filter with a very low cut-off frequency ($< 0.001 \text{ Hz}$).

4.5 Listening test results

Several informal listening experiments were performed in order to evaluate the performance of the sensor. As the 3-D audio system was already equipped with a Polhemus Fastrak head-tracking device, a direct benchmarking of both devices was possible.

In the listening experiments it has been guaranteed that the latency did not exceed the critical values as discussed in 2.3. The total system latency was below 40 ms and the update rate of the head-tracking devices was set to 60 Hz.

In the listening tests no performance differences between both head tracking devices were observed, apart from the fact that the device equipped with the gyroscope sensor showed a slight drift of about $0.1^\circ/\text{s}$ which was compensated by the return movement described above.

Comparing the dynamic auralization to a static one without head-tracking, all subjects perceived the environment as significantly improved. This is in agreement with the findings of [Wen96a] and [Wen96b].

Most of the subjects described the timbre of the Koss headphones as natural even though no equalization of the transfer function of the headphones has been

performed. According to previous investigations of the author for high quality headphones no enhancements by equalizing the headphones' transfer functions were obtained. This is mainly caused by the fact that the variance of the transfer functions measured for multiple persons significantly exceeded the amount of which the transfer functions were to be equalized.

5 Applications

Several applications in the field of mobile communication devices exist which could benefit from the described technology. Virtual acoustics in combination with head-tracking can be used in order to enhance existing applications, and is even capable of creating new ones. For example, the quality of music reproduction can be increased significantly as out-of-head-localization and spatialization of the sound sources can be provided. Apart from synthesizing new auditory scenes, the technology also allows the reproduction of current available music formats initially created for loudspeaker reproduction (e.g. Dolby Digital) by placing "virtual" loudspeakers at corresponding positions in a virtual room. In this case the listener gets the impression of music reproduction with very good listening conditions (e.g. studio room). As the signal processing power of future devices will be sufficient for dealing with such tasks, no additional signal processing hardware is required. Mobile head-tracked 3-D audio can be achieved already by a modified headset equipped with a head orientation measurement sensor connected to the device.

One major application of future virtual environments will be immersive teleconferencing systems [WS97]. Such systems are an integral element of a market that is expected to emerge [DR98]. Different components of teleconferencing systems, for example vision or document sharing, are important, but special attention has to be paid to audition, as it dominates interpersonal communication. Regarding the perception of the auditory scene, enhancements can be observed for an immersive teleconferencing system. The separation of concurrent speakers is increased if the speakers are localized at different positions. This allows a participant in an immersive teleconference to concentrate on one speaker even if other speakers are active at the same time [Bod92, Bla97]. As a result, moderating the discussion, which is commonly the case in teleconferences, is no longer required. Simulating reflections and reverberation in the virtual room causes an

increase in the perception of presence and allows the auditory scene to be perceived as if all the participants were in the same room.

Finally, another possible field of applications is immersive games. The performance of headphone-based game consoles which make use of extensive sound effects can be enhanced by giving the player the possibility of localizing sound sources in all directions and in this way increasing the player's immersion into the auditory scene.

A more detailed description of possible applications can be found in [Pör02].

6 Cross-Links to other modalities benefiting from head-tracking

In addition to 3-D audio reproduction, the head orientation device can as well be used to enhance other applications. For example, if the display of the mobile communication device is controlled by the head-tracking device a large virtual screen can be realized. The user can, by turning his head, shift the picture in different directions (motion scrolling). This way the video picture is adapted to the head movements of the listener. This can, for example, be advantageous when looking at a large map. By turning the head and holding the mobile device in front of the user can shift the actual segment of the map.

Furthermore, location based services can be enhanced by head-tracking. This way it is possible to present information to the user based to the actual head orientation. For example, the user can be provided with information about an object in the environment when he faces it. When turning around and looking at a restaurant, information about the menu can be offered.

Such a task can easily be realized with a magnetic field sensor, as it determines absolute values for the orientation. When using a relative sensor (e.g. gyroscope) only a relative direction can be measured directly. To determine the absolute orientation, information from a positioning system (e.g. GPS) can be applied.

Other applications which can potentially be enhanced by head orientation measurement include interactive games as well as new user interfaces.

7 Conclusion

Possibilities for increasing the quality of mobile virtual 3-D audio by applying head-tracking were inves-

tigated. Requirements based on psychoacoustic research were defined in order to guarantee an undisturbed perception of head-tracked audio. Based on these requirements three different sensor technologies for the detection of head movements were compared. It was shown that gyroscope sensors are superior to magnetic sensors and acceleration sensors for the desired field of applications.

A head-tracking device based on a gyroscope sensor which meets the requirements for undisturbed head-tracked 3-D audio and is capable of measuring the head orientation in the horizontal plane was realized and integrated into a pair of headphones. A good performance of the head-orientation sensor was obtained with an existing 3-D audio system in informal listening tests. Several audio applications were identified which can make use of such head-tracking device. Examples are music reproduction, gaming, and teleconferencing. Several other fields of applications can benefit (e.g. gesture control, vision, location-based services) from the head orientation measurement in order to enhance those applications.

8 Acknowledgement

The work presented in this paper was performed at the Siemens department of Technology and Innovation, ICM MP P in Bocholt. The author would like to thank his colleagues for their advice and their encouragement, and Dr. Roland Aubauer for the initiation of the work.

The author also sincerely thanks the anonymous reviewers for their suggestions to improve this article.

The author wants to express special thanks to the Institute of Communication Acoustics for the co-operation concerning the test system which has been used in this project. Finally, the author would like to thank Michael Lenz who contributed with his diploma thesis to the work presented here.

References

- [AG06] Tom Ammermann and Hartmut Gleselmann, *Akustisches Holodeck*, c't Magazin für Computertechnik **22** (2006), 238–242, ISSN 0724-8679.
- [Beg94] Durand R. Begault, *3-D Sound for virtual reality and multimedia*, AP Profes-

- sional, Cambridge MA, 1994, ISBN 0-12-084735-3.
- [Bla97] Jens Blauert, *Spatial hearing - The psychophysics of human sound localization*, MIT Press, Cambridge, MA, 1997, ISBN 0-262-02413-6.
- [BLSS00] Jens Blauert, Hilmar Lehnert, Jörg Sahrhage, and Holger Strauss, *An interactive virtual-environment generator for psychoacoustic research. I: Architecture and implementation*, *Acustica* **86** (2000), 94–102, ISSN 0001-7884.
- [Bod92] Markus Bodden, *Binaurale Signalverarbeitung: Modellierung der Richtungserkennung und des Cocktail-Party-Effektes*, VDI-Verlag GmbH, Düsseldorf, 1992.
- [Bro95] Adelbert W. Bronkhorst, *Localization of real and virtual sound sources*, *J. Acoust. Soc. Am.* **98** (1995), 2542–2553, ISSN 0001-4966.
- [BWLA00] Durand R. Begault, Elizabeth M. Wenzel, Alexandra S. Lee, and Mark R. Anderson, *Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source*, 5134, 2000.
- [Car00] Michael J. Caruso, *Applications of magnetic sensors for low cost compass systems*, Proceedings of IEEE Positioning, Location, and Navigation Symposium (PLANS) (2000), 177–184.
- [CBSS98] Michael J. Caruso, Tamara Bratland, Carl H. Smith, and Robert Schneider, *A new perspective on magnetic field sensing*, Sensors Expo Proceedings (1998), 195–213, ISBN 0-7803-5872-4.
- [Col63] Paul D. Coleman, *An analysis of cues to auditory depth perception in free-space*, *Psychological Bulletin* **60** (1963), no. 3, 302–315, ISSN 0033-2909.
- [DR98] Amitava Dutta-Roy, *Virtual meetings with desktop conferencing*, IEEE Spec-
trum **35** (1998), no. 7, 47–56, ISSN 0018-9235.
- [HCM⁺95] M. Hashimoto, C. Cabuz, K. Minami, K., and M. Esashi, *Silicon resonant angular rate sensor using electromagnetic excitation and capacitive detection*, *J. Micro-mech. Microeng.* **5** (1995), no. 3, 219–225, ISSN 0960-1317.
- [Inn05] Innovationsreport, *Das weltweit erste referenzdesign für ein linux-fähiges umts/edge-dual-mode-mobiltelefon*, www.innovations-report.de, February 2005, Last visited October 9th, 2007.
- [KSS68] Asbjørn Krokstad, Svein Strøm, and Svein Sørdsdal, *Calculating the acoustical room response by the use of a ray-tracing technique*, *J. Sound Vib.* **8** (1968), no. 1, 118–125, ISSN 0022-460X.
- [Law73] P. Laws, *Entfernungshören und das Problem der Im-Kopf-Lokalisiertheit von Höreignissen*, *Acustica* **29** (1973), no. 5, 243–259, ISSN 0003-682X.
- [Law93] Anthony Lawrence, *Modern inertial technology: navigation, guidance and control*, Springer, New York, 1993, ISBN 0-387-97868-2.
- [LB92] Hilmar Lehnert and Jens Blauert, *Principles of binaural room simulation*, *Applied Acoustics* **36** (1992), no. 3-4, 259–291.
- [Len90] James E. Lenz, *A review of magnetic sensors*, Proceedings of the IEEE **78** (1990), no. 6, 973–989.
- [Len02] Michael Lenz, *Untersuchung von Bewegungssensorprinzipien zur Detektion der Kopforientierung bei 3D Audio Anwendungen*, Unpublished thesis, TU Dresden, 2002.
- [MAB92] Kenneth Meyer, Hugh L. Applewhite, and Frank A. Biocca, *A survey of position trackers*, *Sensors Actuators* **1** (1992), no. 2, 173–200, ISSN 1054-7460.
- [MFT⁺99] Philip Mackensen, Uwe Felderhoff, Günther Theile, Ulrich Horbach, and

- Renato S. Pellegrini, *Binaural room scanning - a new tool for acoustic and psychoacoustic research*, The Journal of the Acoustical Society of America **105** (1999), no. 2, 1343–1344, ISSN 0001-4966. [PT88]
- David R. Perrott and Juliana Tucker, *Minimum audible movement angle as a function of signal frequency and the velocity of the source*, J. Acoust. Soc. Am. **83** (1988), no. 4, 1522–1527, ISSN 0001-4966.
- [MFT⁺00] Philip Mackensen, Markus Fruhmann, Mathias Thanner, Günther Theile, Ulrich Horbach, and Attila Karamustafaoglu, *Head-tracker based auralization systems: Additional Consideration of vertical head movements*, Preprint 5135, 2000. [QSo06] QSound, *Qsound*, www.qsound.com, December 2006, Last visited October 9th, 2007.
- [MK75] Donald H. Mershon and L.E. King, *Intensity and reverberation as factors in the auditory perception of egocentric distance*, Perception and Psychophysics **6** (1975), no. 18, 409–415, ISSN 0031-5117. [San96] Jesper Sandvad, *Dynamic aspects of auditory virtual environments*, Preprint 4226, 1996.
- [MS94] K. Maenaka and T. Shiozawa, *A study of silicon angular rate sensors using anisotropic etching technology*, Sensors Actuators **43** (1994), 72–77, ISSN 0924-4247. [Sci02] Scipher, *Technology for tomorrow's world*, Annual Report 2002, 2002.
- [Nov05] Pedro Novo, *Auditory Virtual Environments*, in: Jens Blauert (Ed.) *Communication Acoustics*, Springer Verlag, Berlin, 2005, ISBN 3-540-22162-X. [Wen96a] Elizabeth M. Wenzel, *What perception implies about implementation of interactive virtual acoustic environments*, Preprint 4353, 1996.
- [Pel00] Renato S. Pellegrini, *Perception-based room-rendering for auditory scenes*, Preprint 5229, 2000. [Wen96b] Elizabeth M. Wenzel, *Effectiveness of interaural delays alone as cues during dynamic sound localization of virtual sources*, Journal of the Acoustical Society of America **100** (1996), no. 4, 2608, ISSN 0001-4966.
- [Pel01] Renato S. Pellegrini, *Quality assessment of auditory virtual environments*, Proceedings of the 2001 International Conference on Auditory Displays, Espoo, Finland (2001), 161–168. [Wen96c] Elizabeth M. Wenzel, *Analysis of the Role of update-rate and system latency in interactive virtual acoustic environments*, Preprint 4633, 1996.
- [Per82] David R. Perrott, *Studies in the perception of auditory motion*, in: R. W. Gatehouse (Ed.) *Localization of Sound: Theory and Applications*, Amphora Press, CND-Groton, 1982, ISBN 0-940728-03-6. [Wen01] Elizabeth M. Wenzel, *Effect of increasing system latency on Localization of virtual sounds with short and long duration*, Proceedings of the 2001 International Conference on Auditory Displays, Espoo, Finland (2001), 185–190.
- [Pör02] Christoph Pörschmann, *3-D Audio in mobilen Kommunikationsendgeräten*, Fortschritte der Akustik - DAGA 2002, 2002, ISBN 3-9804568-6-2, pp. 732–733. [WK89] Frederic L. Wightman and Doris J. Kistler, *Headphone simulation of free-field listening II: psychophysical validation*, J. Acoust. Soc. Amer. **85** (1989), no. 2, 868–878, ISSN 0001-4966.
- [WS97] Gordon R. Walker and P. J. Sheppard, *Telepresence - The future of telephony*, BT Technol Journal **15** (1997), no. 4, 11–18, ISSN 1358-3948.

- [WW06] Chris Weck and Edgar Wilson, *Broadcasting to Handhelds: an overview of systems and services*, Technischer Bericht, EBU (European Broadcasting Union) **305** (2006), ISSN 1358-3948.
- [WWF88] Elizabeth M. Wenzel, Frederic L. Wightman, and Scott H. Foster, *A virtual display system for conveying three-dimensional acoustic information*, Proc. Hum. Factors Soc. 32 **32** (1988), 86–90.
- [YAN98] Navid Yazdi, Farrokh Ayazi, and F. Khalil Najafi, *Micromachined inertial sensors*, Proceedings of the IEEE **86** (1998), no. 8, 1640–1659, ISSN 0018-9219.

Citation

Christoph Pörschmann, <i>3-D Audio in Mobile Communication Devices: Methods for Mobile Head-Tracking</i> , Journal of Virtual Reality and Broadcasting, 4(2007), no. 13, October 2007, urn:nbn:de:0009-6-11833, ISSN 1860-2037.
