RWTH AACHEN UNIVERSITY

CORONA

Realizing an Interactive Experience in Visually Untouchable Rooms Using Continuous Virtual Audio Spaces

Diploma Thesis at the
Media Computing Group
Prof. Dr. Jan Borchers
Computer Science Department
RWTH Aachen University

by
Florian Heller

Thesis advisor:
Prof. Dr. Jan Borchers

Second examiner:
Prof. Dr. Gerhard Lakemeyer

Registration date: Jun 13th, 2008
Submission date: Dec 1st, 2008

I hereby declare that I have created this work completely on my own and used no other sources or tools than the ones listed, and that I have marked any citations accordingly.

Hiermit versichere ich, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie Zitate kenntlich gemacht habe.

Florian Heller

Contents

List of Figures

List of Tables

Abstract

What kind of interactive experience can be implemented in a room that cannot be modified in its visual appearance? This question was the starting point for CORONA, an interactive exhibit in the coronation hall in the city hall of Aachen. The current usage scenario and the fact that architectural modifications are not possible lead to audio as the only feasible medium for such a project. We want to create a personalized immersive virtual audio space for the visitor. The vision behind CORONA is to revive the ancient festivities that took place in this important historical building and let the visitor take part of these.
This thesis gives an overview of current state-of-the-art audio experiences and describes the design and implementation of the technical backbone of CORONA. In the first part we present an analysis of available technologies in the areas of indoor location tracking and spatial audio rendering. The second part describes the design and evaluation of our system.

Überblick

Wie realisiert man ein interaktives Exponat in einem denkmalgeschützten Raum? Diese Frage stand am Anfang von CORONA, ein Teilprojekt der Route Charlemagne, mit dem Ziel, ein interaktives Multimedia-Erlebnis im Krönungssaal des Aachener Rathauses herzustellen. Die derzeitige Verwendung und die architektonischen Gegebenheiten lassen nur Klang als Medium für solch ein Exponat zu. Die Vision von CORONA ist es, jeden Besucher in eine persönliche, virtuelle Klangwelt zu versetzen um ihn an der reichhaltigen Geschichte dieses Gebäudes teilhaben zu lassen.

Diese Diplomarbeit gibt zuerst einen Überblick über den Stand der Technik auf dem Gebiet der interaktiven Klangwelten und der Systemkomponenten die benötigt werden, um solch ein Exponat zu realisieren. Der erste Teil stellt eine Analyse von Systemen zur Positionsbestimmung in Räumen und räumlicher Klangsythese vor. Der zweite Teil beschreibt die Entwicklung und Evaluierung der technischen Basis von CORONA.

Acknowledgements

First of all, I would like to thank Prof. Dr. Jan Borchers for giving me the opportunity to work on this important project. I would also like to thank Malte Weiß for letting me work on his idea of CORONA. I hope that I was able to meet your expectations. It was and will be a pleasure to work with you.

I look back on nine month full of work and fun with my room mates from office 2214, especially Anke, Julie, Mareike, Yvonne, Alex, David, Dieter, Jan-Peter, Leo, Moritz, Nori, Thomas, and all those who I forgot. Thank you for all the interesting discussions, for the insights and ideas, and for all the funny moments. Thanks to Ines, Sascha and Sergej for not getting out of cookies and to Alli, Janine, Marian, Bodo, Michael, René, Stefan and my brother Tobias for the support and friendship. Thanks to all those who reviewed this thesis. I think you will recognize some improvements.

Finally, I want to thank my family and especially my parents for the love and support I got! They always believed in me and brought me to the point where I am now.

Thank you!

Conventions

Throughout this thesis we use the following conventions.

Text conventions

Definitions of technical terms or short excursus are set off in coloured boxes.

Excursus:
Excursus are detailed discussions of a particular point in a book, usually in an appendix, or digressions in a written text.

Excursus: *Excursus*

The whole thesis is written in American English.

Chapter 1

Introduction

The idea behind CORONA is to create an interactive experience in a very special location: the coronation hall in the city hall of Aachen as shown in figure 1.1.

CORONA is part of the Route Charlemagne[1] whose aim is a presentation of Aachen as an important european city by depicting the themes 'History', 'Science', 'Europe', 'Religion', 'Power', 'Economy', and 'Media'. Within the Route Charlemagne the city hall plays an important role, as it is the first of serveral planned stations that will be realized.

CORONA is part of the Route Charlemagne

The city hall has it's roots in the 8th century [Helg and Linden, 2006] and since then, went through a lot of sovereigns and modifications. Despite the historic nature of this building, it is still in use for administrative purposes. The mayors office and the council chamber are located on the ground floor whereas the coronation hall on the first floor is, like in the past 1200 years, used for representative festivities. An effect of the usage of the coronation hall is that there are only two exhibits placed in this room. A statue of Charlemagne and the so called 'Reichskleinodien' which are replicas of Charlemages insignia, i.e., the crown, the scepter, and the orb. The other interesting artefacts are five historic frescos and the room's architecture. This results in a lot of information invisible to the visitors. All the coronation festivites and feasts that took place have no physical

[1] www.route-charlemagne.eu

Figure 1.1: The Coronation Hall

representation. The question is how to revive the past for the visitors without conflicting with the normal usage scenario? Any new exhibits placed in the hall would have to be (re-)movable, structural modifications are difficult and undesired, and visually obtrusive systems would destroy the monumental impression of the hall. Additionally, a calm atmosphere should be maintained as it is an exhibition of museum quality.

The lack of physical representations makes it difficult to use audio guides like those available in other museums, even though they are a very common way to provide visitors with personalized additional information. The creation of a new experience is only possible with headphones, as an installation of speakers would conflict with the aforementioned restrictions as well. Headphones also facilitate the creation of a personalized realistic auditory space.

To revive such events as a coronation feast, we want to create a completely virtual audio space which can be experienced through movement in the hall. Listening to the regnant society and the maidservants talk over the same events from different perspectives can reveal some interesting insights to daily life in the eighth century. Such a virtual audio space introduces some new metaphors, like navigation by ear. As there are no visual cues revealing the position of the sound sources, the visitor will orient himself just by listening and follow the whispering cues. As not all sound sources are hearable from the start, the exploration of the audio space, i.e., going to the next perceptible source is done solely by ear.

The experience is to immerse into ancient festivities only by listening

Chapter 2 gives an overview of existing audio guides and the technologies required to create a new interactive audio exhibit. In section 2.1 we present some currently available commercial audio guides for museums and their possibilities. Section 2.2 describes three state-of-the-art research projects that go beyond everything commercially available.

The implementation of such an experience requires two main components: a location tracking system and a spatial audio rendering engine.The simplest approach would be to use GPS as it is a globally available technology for localization, but it does not work in indoor environments which makes it useless for our scenario. Therefore, we analyzed and evaluated several indoor location tracking technologies and present the results in section 2.3. Section 2.4 explains the challenges of spatial audio rendering and shows what is possible in this area with a focus on mobile devices.

We need two subsystems: location tracking and spatial audio rendering

Chapter 3 describes the technical base of CORONA. A specialized location tracking system combined with an electronic compass provide the motion information required by the spatial audio rendering engine that runs on a mobile device and presents a personalized audio stream to the visitor.

The location tracking is evaluated and discussed in chapter 4. Chapter 5 summarizes the goals and contributions of this thesis and gives an overview of future work that needs to be done in this area.

Chapter 2

Related Work

CORONA is thought as a revolutionary audio experience in a museum context, so we looked for related technologies in this field. As the existing guides are not sufficient for our intended purpose, we looked for indoor location tracking systems and spatial audio rendering engines to implement our idea.

2.1 Audio Guides

Audio guides are a common way to provide additional information

Audio guides have been used for more than 50 years to provide additional content to visitors in exhibitions and museums. Starting with reel-to-reel players in 1957[1], the concept of a personalized audio tour has become a common feature in museums. As Bederson [1995] already remarked, *"part of the reason many people go to museums is to socialize, to be with friends and to discuss the exhibit as they experience it. Taped tour guides conflict with these goals because the tapes are linear, preplanned, and go at their own pace"*. The appearance of random-access digital audio players allowed to overcome these restrictions, giving the visitor back the control over what to hear and when.

A keypad-based guide, consisting of a keypad on the device and dedicated numbers on the exhibits, is the most

[1] www.acoustiguide.com

simple variant of a self-paced tour guide. The visitor passes through the exhibition at his own pace and dials the number of exhibits in which he is interested. In this way, no possibly interesting item is omitted in the tour. Keypad-based guides like the AntennaAudio[2] X-plorer™ are a widespread method to provide detailed information to a high number of exhibits in different languages.

Keypad-based guides give the user the control of what to here and when

However, there is a fairly high coherence between the position of the visitor and the information that is interesting at this location, e.g., when a visitor stands in front of a painting, it is very likely that he wants to hear the auditive information available for this painting. The zone, usually located around exhibits, where the guide should react on entrance is called area of interest.
Bederson [1995] combined a modified MiniDisc™ player and infrared beacons to implement such a trigger-based location aware museum guide. A trigger-based audio guide uses beacons placed nearby the exhibit instead of numbers. These triggers can work with infrared light (e.g., Dataton Pickup, coolMuseum), audio watermarks [Gebbensleben et al., 2006], electromagnetic signals like RFID or wireless LAN (WLAN) (e.g., guidePORT, iGuide Wireless) to determine when the visitor reaches a certain area of interest. Nevertheless, the concept of areas of interest also opens up some interesting usability questions:

Trigger-based guides exploit the connection between physical location and the information needed at this point in space

- What happens if a visitor enters an area of interest? Does the sound-sample immediately start or do we require an additional step?
- What happens if a visitor leaves the area while the sample is still playing? Does the playback stop?

The Dataton Pickup™ does not automatically start playback, but requires the user to point-and-click in the direction of the trigger [3]. This additional step leaves the visitor in control of what he wants to listen to and when, on the other hand this also requires a visible cue for the visitor to see where he can point to.

[2] www.antennaaudio.com
[3] www.dataton.com/pickup

Sennheiser[4] also offers the guidePORT™ receiver in a version with an additional keyboard to provide a fallback solution if the visitor feels uncomfortable with the location aware approach.

To synchronize audio playback in critical scenarios like video presentations, an FM-radio receiver can be integrated into the mobile device. The trigger transmits the channel on which the adequate audiostream is available and the audiostream on a separate FM-channel.

FM-Receivers can be integrated for synchronized audio playback

Even if the location information is only discrete, a decent level of interactivity can be achieved as shown in the AudioCave project [Bischof, 2006] or by Lyons et al. [2000]. Both systems realize an interactive game experience with trigger-based approaches.

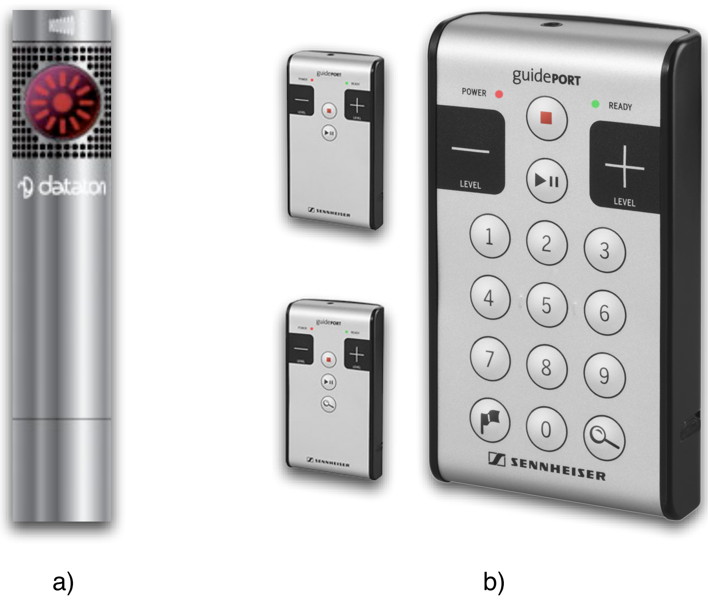

Figure 2.1: Some available Audioguides: a) Dataton Pickup™ b) Sennheiser guidePORT™ is availabe with and without keypad to provide a fallback if the visitor feels uncomfortable with the location aware technology

[4] www.guideport.com

Keypad-based, as well as trigger-based, systems are unsuitable for outdoor applications like city tours. The challenging question is where to place the numbers, such that they still can be found by the visitors. The same problem arises for trigger-based variants, even though it can be attenuated by using triggers with a higher range.

Fully location aware systems, like the Mauerguide[5], build a user-centric information space as shown in figure 2.2. The Mauerguide presents the user a map with points of interest (POI) next to the position determined with GPS. Besides the navigation function, additional multimedia (i.e., videos, photos and audio comments) content corresponding to these POI is stored on the device.

2.2 Virtual Audio Research Projects

The area of next generation audio guides is very broad. New developments in audio rendering and tracking technology allow to build more and more realistic sceneries. On the other hand, these new systems also allow a more deeper analysis of the interaction with virtual audio. Realistic spatial audio rendering generates immersive experiences with a lot of possible use cases. The direction of the following research project is, on the one hand, the interaction with virtual audio spaces, and, on the other hand the development of highly immersive audio experiences.

2.2.1 ec(h)o

In addition to the usual singular audio space, the ec(h)o-project [Wakkary and Hatala, 2006] introduced two audio spaces accessible through a playful approach. Instead of a PDA the visitor gets a small colored cube with 10 cm edge length and headphones. Using a combination of RFID-based and optical tracking, a dynamic soundscape is generated, suggesting an atmosphere relative to the nearest exhibits. Coming closer to an exhibit, the system presents

[5]www.mauerguide.com

Figure 2.2: The User-Centered Information Space of the Mauerguide: the current position is shown by the yellow hairlines and the red icons represent some points of interest.
Photo: Werner Bern, Copyright Antenna Audio 2008.

three audio prefaces in sequence on the left, both, and right channel respectively. Turning the cube starts the audio sequence announced on the respective channel, holding it upright starts the sequence announced in stereo. The orientation of the cube is tracked with cameras mounted under the ceiling. As the tracking only determines the position in space without the head-orientation, no directed auditive information can be presented.

2.2.2 Demor

Demor is a first-person shooter for blind teenagers

Demor[6] was developed at Utrecht School of the Arts as a first-person shooter for blind teenagers. A laptop in the backpack renders a virtual auditory space depending on the data received from a GPS receiver and a digital compass. A modified joystick serves as physical input device. The only orientation information available in the game is presented though headphones. The auditory map can be adapted to the physical conditions of the playing field to provide a realistic impression.

2.2.3 LISTEN

LISTEN augments the environment through a dynamic soundscape

The most interesting project for us is LISTEN [Eckel, 2001]. In LISTEN, the physical environment is augmented through a dynamic soundscape which users experience over motion-tracked wireless headphones [Terrenghi and Zimmermann, 2004]. Instead of using simple binary tracking, the IEMW group at Vienna University of Technology developed a position and orientation tracking system that returns the absolute position in (x, y, z)-coordinates as well as the orientation (azimuth) [Warusfel and Eckel, 2002]. The exposition hall is divided into several areas of interest connected to the visual objects within the environment. These larger zones are again divided into object zone and near field as shown in figure 2.3. The near field is a very small area in direct proximity of the exhibit and serves to present information on very small details of a painting for example. Inside the object zone, a sophisticated auditory rendering process generates a dynamic soundscape related to the exhibit, depending on the visitors position and head orientation. The continuous tracking in combination with the spatial audiorendering allows realistic auditory scenes to be rendered, e.g., speaking paintings.
LISTEN was deployed in the Kunstmuseum in Bonn/Germany, but only eight people could participate at the same time because the motion tracking was to complex to handle more visitors.

[6] demor.hku.nl

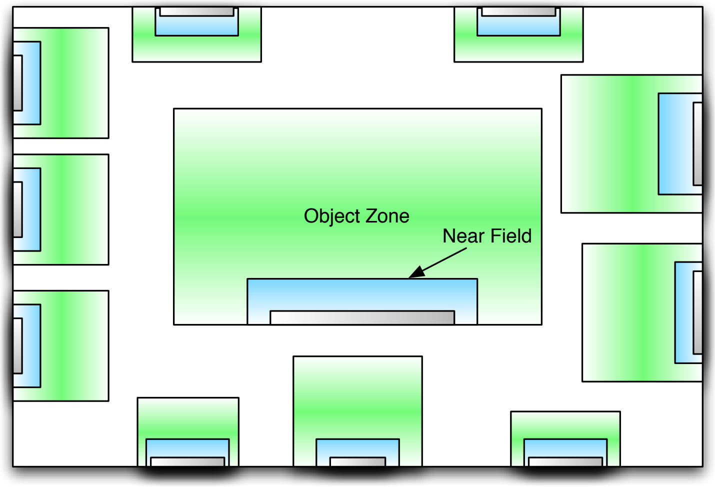

Figure 2.3: The LISTEN Space Model (adapted from Terrenghi and Zimmermann [2004]).

2.3 Position Tracking

Indoor tracking is a field of current research and development as it is an important waypoint to new location aware applications, e.g., industrial, self piloting robots in large warehouses or localization of medical equipment in hospitals. These different types of applications have different requirements concerning the accuracy; if you are looking for a wheelchair in a hospital, a positioning error of two meters is perfectly acceptable whereas this accuracy would be much too coarse for a robot driving along three meter large corridors.

Indoor tracking makes new location aware applications possible

The requirements on the tracking system are very manifold

There are two ways to localize a mobile device (the so-called tag) using several fixed devices (the so-called anchor points or base stations). The anchor-point-based approach needs a centralized infrastructure that combines the informations from all available base stations. This means that the location information of all tracked tags is calculated at a central point which is a possible privacy issue. This is not the case for the tag-based approach where the anchor points just help the tag localize itself. The process of calculating the position from at least three distance values is called trilateration whereas triangulation calculates the position from the incoming angle os some signals. The evaluation in the following sections is done under the conditions explained in chapter 1 i.e., visual unobtrusiveness and no structural alterations.

2.3.1 Wireless LAN

As Wi-Fi is an integrated feature of many new mobile devices, like smartphones or PDAs, there is intensive research on how to use it for localization purposes. Especially systems working with standard hardware give you the localization "for free", even modified access points as fixed infrastructure are acceptable. In the following sections we will describe the pros and cons of the different existing approaches.

Strongest Base Station

A very simple approach is to compare the received signal strengths with a map of access points (AP). Assuming that the strongest AP ist closest, the position can be roughly estimated indoors with an accuracy of 10 m to 30 m, depending on the environment.

Received Signal Strength Indicator (RSSI)

Interpolation between signal strength fingerprints gives a rough location estimation

The received signal strength indicator (RSSI) is a one-byte number added to each received packet by the Wi-Fi card's driver. It is a representation of the signal strength as an arbitrary number. During the propagation from the access point to the client in the free field, the energy of the electromagnetic power reduces with distance as stated in the radio propagation model. This relationship between distance and received signal strength is well known for the free space case [Velayos and Karlsson, 2004]:

$$P(d)[dBm] = P(d_0)[dBm] - 10n \log \left(\frac{d}{d_0} \right)$$

where $P(d)$ is the received power at some distance d, $P(d_0)$ is the transmitted power measured at a reference distance d_0, and n indicates the rate at which the path loss increases with distance. In the free space case $n = 2$, but indoors, refraction, reflection, attenuation and obstacles found in the path have to be taken into account. The Spread Spectrum Scene [2001] proposes a value of 3.5 as a reasonable value while others suggest that the number of obstacles or walls should be included in the model. Velayos and Karlsson [2004] conclude that the free space model cannot be adapted to fit indoor Wi-Fi requirements. Despite those results, Intel Research and UC Berkeley use signal-strength trialateration in the Place Lab[7] project to reach an accuracy of about 10 meters.

Instead of calculating the distance directly from the RSSI, one can build a RSS map by storing the RSSIs of all available access points in range of a specific point. The position is then determined by comparison of the currently measured values with linear interpolation between the stored points. The problem is that the measured RSSs are only valid for a particular Wi-Fi adapter card. This means that in heterogenous hardware environments a dedicated radiomap has to be calibrated for each model. Additionally, the measured RSSI highly depends on the orientation of the measuring device.

Microsoft Research developed RADAR [Bahl and Padmanabhan, 2000] based on a calibrated radio map to determine the nearest neighbors in signal space. In their experiments, the empirical method requires more effort during installation but outperforms the radio propagation which is more flexible. Both methods are significantly better than the strongest basestation approach.
Ekahau[8] uses the radio map procedure in his real-time location system (RTLS) with a specified accuracy of about one meter in typical office environments. The use of proprietary tags or access points ensures a well-known measuring behavior on tag or access point side.

Since we planned to use Wi-Fi enabled mobile devices for

[7] www.placelab.org
[8] www.ekahau.com

CORONA, we did some experiments with Magic Map[9], a freely available software from Humboldt-Universität zu Berlin, to check if an acceptable accuracy can be acheived. Magic Map combines several approaches to improve accuracy. As mentioned in [MagicMap, 2006], 10 m accuracy is achieved if only the location of the access points is known; thus, the position is determined with the radio propagation model. An accuracy of about five meters is reached if there is at least one reference measurement per room. The combination of these two procedures enhances the accuracy to the order of one meter within the circumference of the access points.

MagicMap Test These promising specifications encouraged us to do some own tests with the software. We did the test *in situ*, as the situation in the coronation hall is not comparable to a normal office scenario. We placed four access point (Apple AirPort Extreme Basestations) on the green marked spots in figure 2.4. The radio power of the access points was reduced to 10% in the AirPort-configuration tool. The measurement of the reference points were made with an IBM Thinkpad X31 with an Atheros based 802.11abg wireless LAN adapter card and MagicMap 0.9.2-beta1.
The tracking results were disappointing. It was only possible to guess tendencies, in which quadrant we were, i.e., an error of more than 10 m. We repeated the test with an Apple MacBook Pro and tried different settings but the results did not become better.
An explanation for these bad results is that the coronation hall significantly differs from the standard office environment as that there are nearly always line-of-sight conditions to the access points. This means that there are no significant differences between the RSS fingerprints.

Time Difference of Arrival

A more promising approach is the measurement of the signal propagation time of wireless LAN packets. Electro-

[9] www.magicmap.de

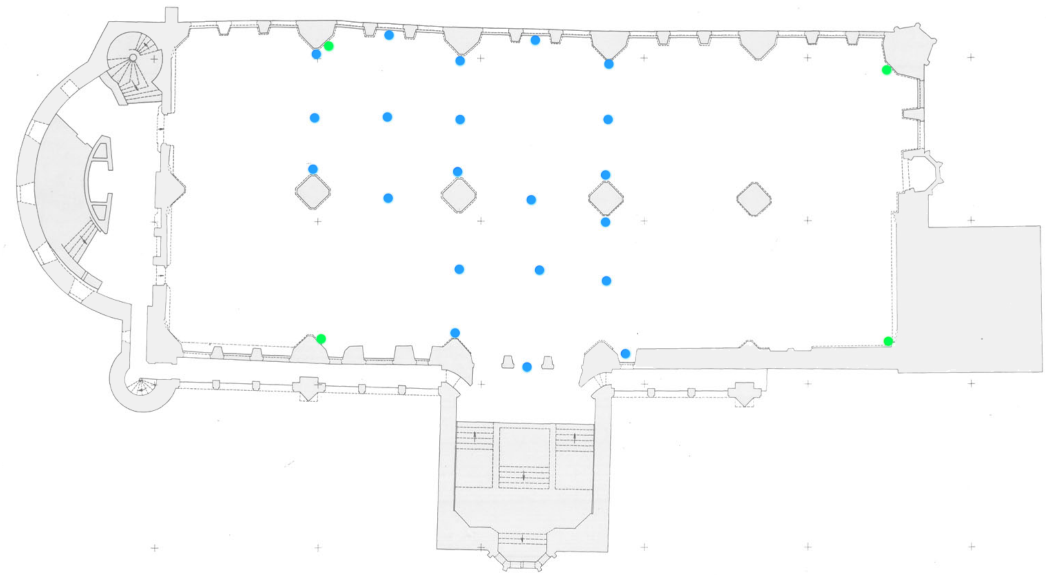

Figure 2.4: Experimental Setup of our Magic Map Test: The access points were situated at the green locations, the blue positions describe the places where reference measurements have been taken.

magnetic waves propagate with the speed of light ($c = 299.792.458\frac{m}{s}$ or $\approx 3 \cdot 10^8 \frac{m}{s}$), so if we know when the packet left the sender and when this packet arrived at the receiver, we can convert this into a distance information. However, very precise timing informations are needed as a resolution of 1 ns already yields to a distance accuracy of 30 cm.

A radio signal covers a distance of 30 cm in 1 ns

Using the timestamps set by the operating system (OS) is inapplicable because the Wi-Fi card communicates a received packet to the system by interrupts. The OS sets its timestamp in the moment the interrupt is attended which depends on the overall system load and can lead to delays of several microseconds [Velayos and Karlsson, 2004].

The timestamps set by the operating system are not accurate enough

Setting the timestamps in the Wi-Fi hardware is much more reliable. The ATH5K[10] driver for Atheros-chip-based cards includes a reception (RX) timestamp but no transmission (TX) timestamp in the packets. Using these timestamps requires synchronized clocks down to the order of nanoseconds in all implied devices. The IEEE 1588 standard describes precise clock synchronization in LAN environments and hardware prototypes reach an average clock drift of 1.1 ns while software implementations in a Linux PC net-

[10]www.madwifi-project.org

work driver reach a 660 ns clock drift [Kannisto et al., 2005]. Nevertheless, as a timestamp resolution of 1 ms is sufficient to implement the Wi-Fi channel access procedure, it is unlikely that hardware manufacturers include this feature in their products.

Additional hardware is needed for a precise time-of-flight measurement

To overcome the problem of the the clock resolution and timestamping, Izquierdo et al. [2006] built a hardware counter connected to the Wi-Fi adapter with a 44 Mhz clock to measure the round-trip-time (RTT). The Request-To-Send (RTS) and Clear-To-Send (CTS) management frames defined in the 802.11 standard are well suited for this purpose as they are directly handled by the firmware of the Wi-Fi adapter. An RTT cycle is defined as the time elapsed between the end of an RTS frame and the beginning of the corresponding CTS frame. The additional clock reports those RTTs to an application on the client through the parallel port. This additional hardware improves the accuracy of Wi-Fi-based tracking to an accuracy of 2 m to 2.3 m.

Even if there are some promising results in the field of localization with Wi-Fi, none of the existing approaches really matched our requirements. The RSSI variants are not precise nor reliable enough to work in a museum environment. The TDOA approach seems much more reliable but suffers from the need of very precise timing information.

2.3.2 Bluetooth

Bluetooth is also a widely available technology in mobile devices, even more than wireless LAN. Implementing a localization system based on this technology would yield in a large number of possibly location aware devices. Hallberg et al. [2003] developed a small Java application for Bluetooth localization. As the range of a Bluetooth device is specified to a maximum of 10 m a rough estimation of the position can already be determined with RSS fingerprints.

A dense mesh of Bluetooth devices is needed for accurate localization

For a more accurate localization, a very dense mesh of Bluetooth devices is needed as at least three devices with known position have to be seen at the same time. If this is the case, the system implemented by Hallberg et al. [2003] reaches an accuracy of roughly two meters. The most serious drawback of this technology is its slowness. Bluetooth is not intended to build connections as fast as possible, instead it is designed as a robust and easy to use inter-platform communication medium. To know if a Bluetooth device with fixed coordinates, a search for active devices has to be started. The search time is limited to a given time (e.g., 5 s) but has a lower bound because the other devices need some time to answer the request. After the list of available devices is generated, the device has to connect to a set of those to measure the link quality. Every connection takes about 200 ms which results in a localization time of about 6 s for four anchor points.

A newer approach described in Genco et al. [2005] achieves an incredible accuracy of 37.5 cm in a mixed indoor and outdoor environment. Instead of using the very coarse RSSI value that actually only makes it possible to know wether a device is within the range of a given base station or not, they use the Link Quality (LQ) value. LQ is a number between 0 and 255 accessible trough the Host Controller Interface of the Bluetooth stack and quite reliable for distance estimation. These results, however, are mainly achieved in simulations and the authors do not mention any details on the refresh rate.
Nevertheless, these ideas deserve further investigation and an own implementation and tests should be considered for future work.

2.3.3 Ultrasound

Cricket combines RF and ultrasound signals

The Cricket indoor location system [Priyantha, 2005] makes use of the different propagation speeds of electromagnetic and sound waves. It consists of beacons and tags that just differ in the software they run. Each beacon periodically transmits an RF message that contains, amongst other information, the beacons unique identifier and at the start of this RF message a unique ultrasonic pulse. When a tag receives an RF message it stops the time until the ultrasonic message arrives and computes the distance.
For a room-wise tracking a single beacon per room is sufficient, for a more precise tracking, at least three beacons have to be in the range of a tag, i.e., approximately 10 m.

The accuracy is within the range of centimeters and it is also possible to get the orientation of the tag if it is equipped with five ultrasonic receivers.
Cricket is unsuitable for the coronation hall because we would have to build a mesh of beacons with a step-width of about five meters. For architectural reasons, it is not possible to place the beacons at the same height on such a grid.

2.3.4 Radio Frequency Identification

Radio frequency identification (RFID) describes an automatic identification procedure to remotely retrieve information from transponders called tags. There are several different types of tags:

- **Passive tags** do not have their own power supply, but use the power induced by the reader to power up the integrated circuit (or microchip) and transmit the data. The read distances range depends on the type of reader and is roughly 10 m with a far-field reader.

- **Active tags** have an own power supply used for both powering the microchip and broadcasting the response signal. This makes them much more reliable in environments unfavorable for RF signals.

- **Semi-Passive tags**, like active tags, have their own battery, but it only supplies the microchip, not the signal transmission. Semi passive tags provide quicker responses as the power collecting phase of the passive tags is omitted and the power consumption is lower then in active tags.

Localization with RFID comes down to a fingerprinting approach as we only get an information which tag is in range or not. A range of 10 m would result in a very coarse location information whereas the near-field version would lead to a very dense mesh of RFID tags which is not feasible in our usage scenario.

2.3.5 UWB - Ultra Wide Band

UWB allows very precise localization

Ultra wide band radio is a technology initially designed to establish short-range high-bandwidth communication, that is why it is also called wireless USB sometimes. While usually RF-technologies try to concentrate the transmitted power to a very small frequency-band, UWB does the exact opposite. It transmits signals with a fairly low energy-level on a very wide frequency band which makes UWB very robust against interferences with other RF technologies. In Germany, a band from 30 MHz to 10.6 GHz is approved for UWB-usage [Eisenacher, 2006].
UWB is also interesting for localization purposes as it is also very robust against multipath fading, i.e., signals are reflected during transmission, and allows extremely precise tracking down to centimeter level [Gezici et al., 2005].

We took a look at an UWB system from Ubisense[11] running in the Laboratory for Machine Tools and Production Engineering at our university. The environment was quite disadvantageous for RF technology as there was a lot of metallic structure within the area to track, but the results were still acceptable. As UWB tracking is mostly TDOA-based and the problem of propagation speed is present as well, the base stations need synchronized clocks and, therefore, a connection between each other.

2.4 Spatial Audio Rendering

To create an immersive virtual auditory space, we have to take a look at the humans audio localization capability. In 1907, Lord Rayleigh found out that the horizontal localization of sound sources bases on two different physical phenomena [Strutt, 1907]. For frequencies higher than approximately 3000 Hz, the horizontal position is retrieved through the interaural level difference (ILD), i.e., the level difference between both ears. When a sound signal is presented from the side, the listener's head interrupts the path from the source to the far ear. The far ear is effectively shadowed,

The localization of sound sources in the horizontal plane is mainly achieved with two different methods, interaural level difference and interaural phase difference

[11] www.ubisense.com

and the sound pressure level is different on both ears. However, with frequencies below 1000 Hz the wavelength can be several times larger than the head which makes it impossible to measure level differences. Therefore, the localization is done with the interaural phase difference (IPD). The human auditory nerve, in contrast to other primates, does not detect interaural phase differences for frequencies above 1400 Hz [Middlebrooks and Green, 1991]. The front/rear ambiguity is resolved by the monoaural and binaural cues that arise from the scattering process from the listeners body, head and ears. These cues are encoded in the so called hear related transfer functions (HRTFs). Every person has its own individual set of HRTFs depending on the ear shape and geometry that is responsible for the vertical localization and the front/rear distinction [Zotkin et al., 2004]. However, ILD and IPD are the most important cues. Even if the source is presented directly in front of the listener such that error is lowest (2°), some stimuli are localized to the rear and vice versa.

To create a realistic immersive audio experience exploiting personalized sets of HRTFs, very high effort is needed [Zotkin et al., 2004]. However using a generalized set of HRTFs or rendering the scene by interpolating between some pre-recorded samples [Algazi and Duda, 2005] the required calculations can be reduced to cope with the limited processing power available on mobile devices. Standards like MPEG Surround [Breebaart et al., 2006] or openAL[12] make several assumptions to reduce the calculation effort but achieve a decent degree of realism and with this achieve an experience that feels plausible.

OpenAL is a cross platform, three dimensional audio API designed for gaming applications. The programmer defines some general parameters concerning the environment, defines sound sources and their properties, and places them in a coordinate system. For the listener, a position and orientation is defined and openAL renders the auditory scene according to the given parameters.

[12] www.openal.org

2.5 Summary

None of the commercially available museum guides satisfied our requirements because the trigger based technology does not provide a fine-grained tracking and the fully location aware systems revert to GPS which does not work indoors. LISTEN is quite similar to what we want to realize, but the system design limits the number of users to eight and a very complex technical backbone is required.
From the tracking side the most promising technology, Wi-Fi, failed during our tests, which brought us to use a system from Nanotron Technologies further described in chapter 3.
For the mobile audio rendering, we decided to use the openAL implementation available on Apple's iPhone™ .

Chapter 3

Implementation of Motion Tracking and Audio Rendering

Wo laufen sie denn ...

Loriot - German humorist

After an evaluation of several tracking approaches described in chapter 2, we decided to use the Nanotron nanoLOC tracking system. The specified accuracy of 2 m is not as high as for other systems, but it offers several other advantages that are important to us: The architecture is extremely flexible as we have full source code access to the firmware running on the tracking modules, and the hardware works completely autonomously, i.e., there is no wiring required between the anchor points.

3.1 The nanoLOC System

The nanoLOC system basically consists of the multi-purpose microchip NA5TR1 which includes data communication and ranging features. This chip can be combined with other hardware to a very specific application as in our case, or to a more flexible general purpose system like the nanoLOC developer boards.

3.1.1 Chirp Spread Spectrum

The nanoLOC technology bases on a technical principle called Chirp Spread Spectrum (CSS) which brings the advantages of UWB (cf. section 2.3.5) to the popular license-free 2.4 GHz ISM-band. It combines the advantages of the three possible modulations for an electromagnetic signal – amplitude, frequency, and phase modulation – to a single method called Multi Dimensional Multiple Access (MDMA).
To transmit data over the air, MDMA uses chirp pulses which are linear frequency modulated signals with constant amplitude. A chirp signal totally fills a given bandwidth for a defined duration, starting with a low frequency and ending with a high one. CSS is very robust against

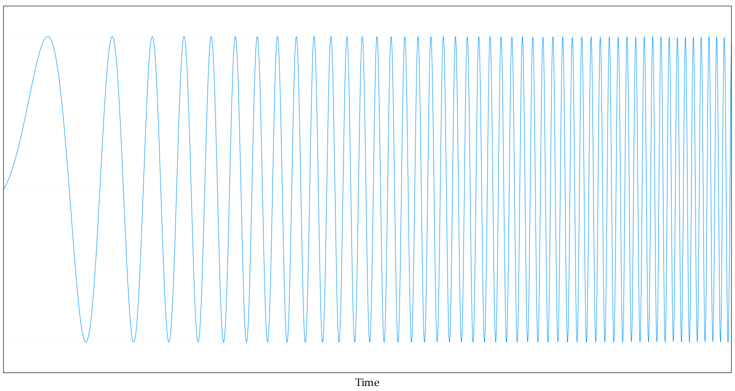

Figure 3.1: A Linear Frequency Modulated Pulse (Up-Chirp): starting with a low and ending with a high signal frequency at a constant amplitude.

narrow- and wideband disturbances which ensures stable working conditions in the highly used 2.4 GHz ISM band where other communication systems induce massive disturbances. This robustness further allows to reduce the transmission power which comes in handy for mobile applications and also reduces the human exposure to electromagnetic signals. Another advantage of CSS is its robust-

CSS is very robust against both interference and multipath fading ness against multipath fading. The electromagnetic signal from the transmitter is reflected by walls and other obstacles and reaches the receiver with several echoes. However, some frequencies get amplified while others get attenuated which results in a communication loss for narrowband transmission systems. The use of the full 80 MHz bandwidth in CSS results in a balance between attenuated and amplified signals [Nanotron Technologies GmbH, 2007a].

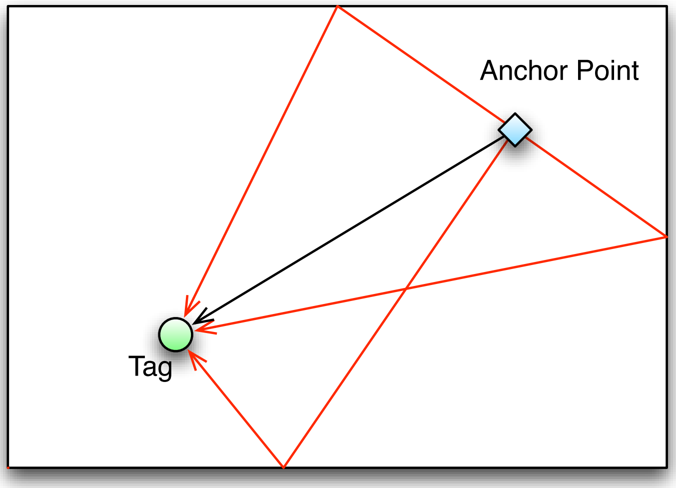

Figure 3.2: The Multipath Problem: in addition to the direct signal, the tag also receives echoes of the original signal which make the TOA measurement difficult.

3.1.2 Ranging Procedure

As the nanoLOC system also uses signal propagation time measurements to determine the distance, the same problems arise as already described in section 2.3.1. To solve the problem of synchronized clocks, the system uses two way ranging (TWR), i.e., it measures the time for a complete round trip between two stations. The tag sends out a packet to an anchor point and simultaneously starts its stopwatch. The anchor point receives the packet, starts its own stopwatch, processes the packet, and returns an acknowledge-

Two way ranging does not require synchronized clocks

ment to the tag containing the measured processing time t_2. The tag receives the acknowledgement and stops its stopwatch at a time t_1. The signal propagation time between tag and anchor point is equal to:

$$t = \frac{t_1 - t_2}{2}$$

As all measurements are done in the firmware of the chip, they are very precise and independent of any other controller.

Excursus: *Quartz Oscillator Drift*

Quartz Oscillator Drift:
Quartz crystal oscillators are used to provide integrated circuits with precise timing information. A small crystal is excited by an electric current and begins to oscillate, generating a uniform wave. However, the oscillation frequency depends on the temperature of the crystal, which is not critical in standard applications, but in our case this fluctuations affect the distance measurement. There are temperature compensated crystal oscillators, but they are very expensive and not available in the required package size.

Although synchronized clocks are no longer required, the measurements rely on the quality of the clock generating oscillator. The round-trip-time (RTT) is rather large compared to the signal propagation time. The processing, transmission and reception times are in the order of hundreds of microseconds whereas the time of flight is in the order of tens of nanoseconds. The RTT-measurement error due to oscillator drift should not exceed 1 ns. This requires a crystal oscillator with a tolerance of about 10 ppm (parts per million) which is more than the usually deployed ones in real time location systems (RTLS). The clock drift problem can simply be avoided by doing the RTT measurement twice and symmetrically. The first measurement goes from the tag to the anchor point and back to the tag, the second measurement starts at the anchor point, goes to the tag, back to the anchor point and then sends the result to the tag (cf. figure 3.3) [Nanotron Technologies GmbH, 2007b]. This is called Symmetric Double Sided TWR (SDS-TWR).

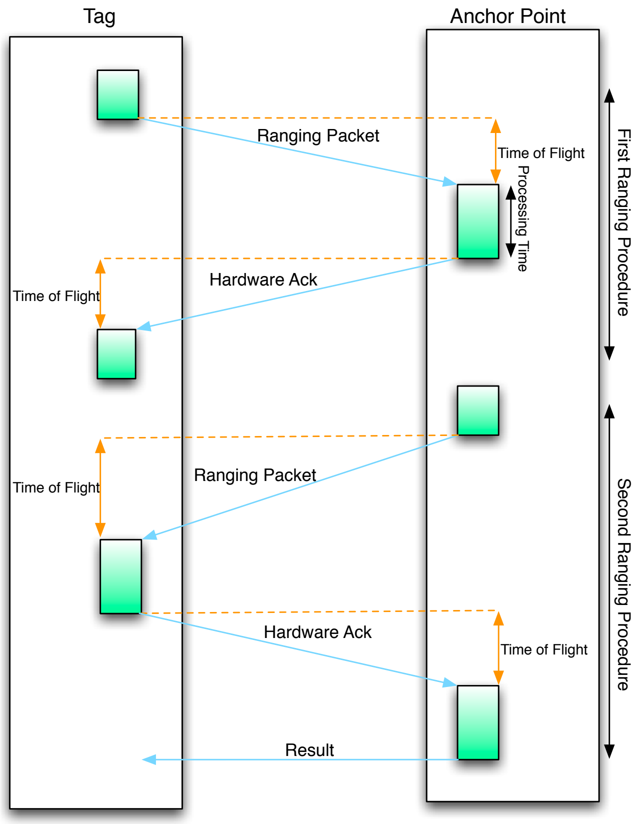

Figure 3.3: Symmetric Double Sided Two Way Ranging (SDS-TWR): the complex ranging procedure does not require synchronized clocks.

3.1.3 Scalability

During the RTT measurement, the anchor point is not able to handle other requests because there is only one counter available to measure the processing time. Since a SDS-TWR cycle takes approximately 10 ms, a complete localization procedure with four anchor points takes approximately 40 ms. Therefore, the number of tags is limited for a given number of anchor points. If we suppose that during a localization cycle all four anchor points are blocked simultaneously, a number of 25 tags can localize itself once per second in an area covered by four anchor points. The assumption is not always true because it is theoretically possi-

The nanoLOC architecture does not scale well ble that four tags synchronize exactly with the anchor point not needed by the others. However, with a higher number of tags, the risk of collisions increases, resulting in a higher number of failed ranging runs. The number of tags T that can simultaneously be localized at a refresh rate R (in Hz) can be expressed as the following equation:

$$T = \frac{\frac{1}{R} \cdot 1000}{P \cdot 10}$$

The problem is to find the correct balance for the three values. On the one hand, we do not want the refresh rate to drop below 1 Hz. On the other hand, we have to handle a decent number of tags, and we have to keep in mind that there are four stone pillars in the coronation hall which cause a loss of connection if they are in the line of sight between the tag and the anchor point. To solve these problems, we decided to divide the coronation hall in ten zones that are covered by four anchor points each (cf. figure 3.4). This way, the problem posed by the pillars is solved and we can track a high number of tags in parallel. As a downside we have to implement a roaming procedure to seamlessly track the visitor throughout the different zones.

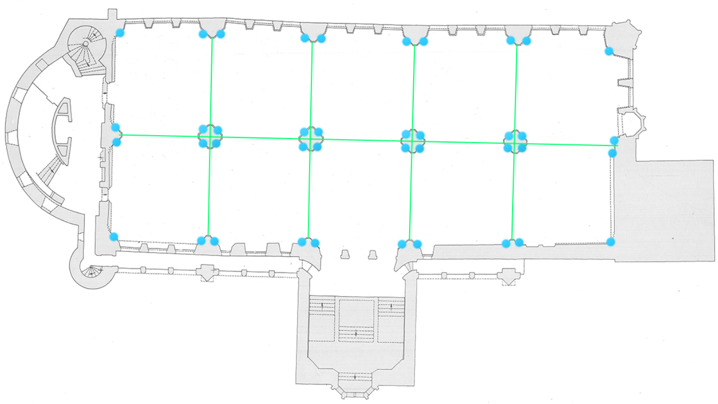

Figure 3.4: The Positions of the Anchor Points: the 10 square areas are covered with four anchor points each.

3.2 Localization Algorithm

The nanoLOC tracking system just returns the measured distance from the tag to given base stations. This information has to be transformed into coordinates in a two-dimensional coordinate system relative to the existing environment. In the optimal case, the measured distance precisely equals the real distance, the circles intersect in exactly one point as shown in figure 3.5. This process is called trilateration.

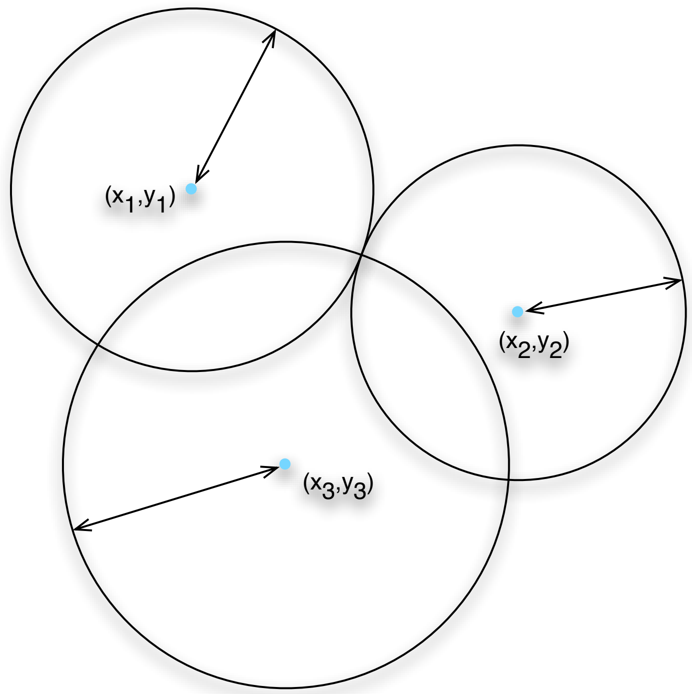

Figure 3.5: Localization with Exact Distances: the three circles intersect in exactly one point.

3.2.1 Mathematical Solution

The circles described by the anchor point i and the measured distance r_i is mathematically defined by the following equation:

$$(x - x_i)^2 + (y - y_i)^2 = r_i^2 \tag{3.1}$$

Now we could set up a nonlinear equations system with these three constraints and solve them simultaneously, but this technique is not feasible because it would yield a nonlinear equation system of high degree.
Instead, we linearize the problem, meaning the reduction of calculating the intersection of straight lines. We use the first constraint as linearizing tool. Adding and subtracting x_1 and y_1 to (3.1) gives us

$$(x - x_1 + x_1 - x_i)^2 + (y - y_1 + y_1 - y_i)^2 = r_i^2 \quad (i = 2, 3) \tag{3.2}$$

Expansion and regrouping leads to

$$(x - x_1)(x_i - x_1) + (y - y_1)(y_i - y_1) = \frac{1}{2}\left[r_1^2 - r_i^2 + d_{i1}^2\right] \tag{3.3}$$

Thus, solving the following linear equation system gives us the coordinates of the intersection point.

$$\underbrace{\begin{pmatrix} x_2 - x_1 & y_2 - y_1 \\ x_3 - x_1 & y_3 - y_1 \end{pmatrix}}_{A} \cdot \underbrace{\begin{pmatrix} x - x1 \\ y - y1 \end{pmatrix}}_{\vec{x}} = \underbrace{\begin{pmatrix} b_{21} \\ b_{31} \end{pmatrix}}_{\vec{b}} \tag{3.4}$$

with $b_{n1} = \frac{1}{2}\left[r_i^2 - r_n^2 + d_{n1}^1\right]$ which means that at least three distances have to be known to be able to calculate the position by trilateration. Equation (3.4) can be extended to more than three distances by multiplying both sides with A^T leading to the well known linear least squares equation

$$A^T A\vec{x} = A^T\vec{b} \tag{3.5}$$

The problem is that measured values are subject to noise and the distance values are likely to be larger than the actual distance. This is clear because in the best case the ranging returns the exact distance, but every obstacle extends the time of flight and thus the measured distance. However, the assumption that all three circles intersect in on point can not be used anymore and the localization algorithm has to cope with these multiple intersection points (cf. figure 3.6). Murphy Jr. [2007] evaluated several mathematical approaches to trilateration with approximate distances for a coal mining company and came to the conclusion that the solution of the linearized equation system (3.4) is not accurate enough for the intended use. In contrast to us, he analyzed the three-dimensional case for an open pit mine where the order of the x and y dimensions were much greater than for the z axis. As the solution of the linearized equation system was only inaccurate for the z axis and in our, two dimensional case, the x and y values are of the same order, resolving the equation (3.5) gives us a simple solution. In the two dimensional case, the product $A^T A$ is

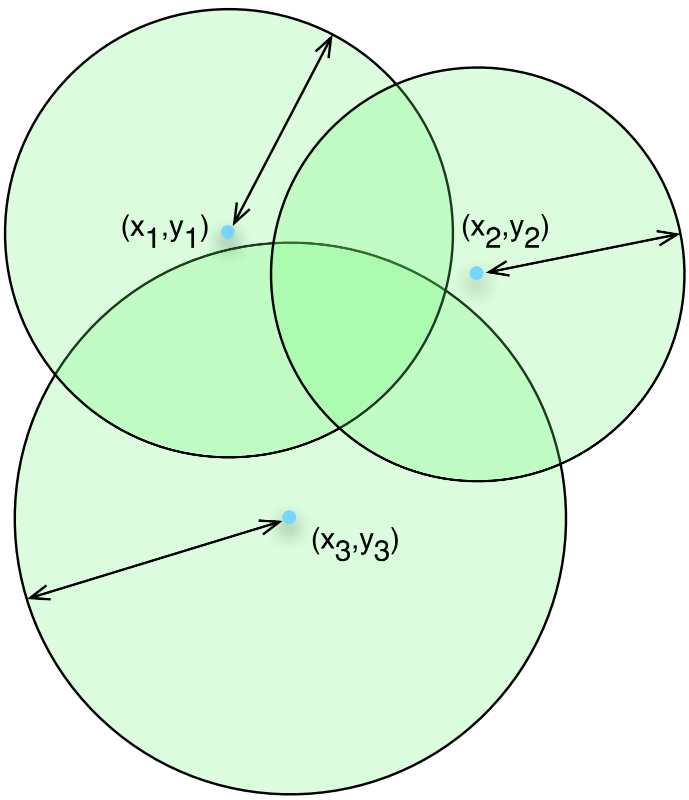

Figure 3.6: Localization with Approximate Distances: the actual position is somewhere in the area covered by all three circles.

always a 2×2 matrix which can easily be inverted.

$$\vec{x} = (A^T A)^{-1} A^T \vec{b} \quad (3.6)$$

Erroneous distance measurements have a high impact on the calculated location

We implemented this solution in our test application and it became clear that this algorithm is unsuitable for the nanoLOC system. The least squares method is susceptible to false distance measurements which have a high impact on the calculated position. There are different approaches to encounter this problem, but they all rely on a much higher number of measurements to exclude the outliers. Since we only have four values, there is some probability that correct values get rejected during this process because of a very false measurement.

3.2.2 Empirical Solution

The empirical solution is more robust against false distance measurements

Because we knew from some demo applications that the system can perform much better, we implemented an empirical approach which is more robust [Horn, 2007].
Since the measured distance d_{ri} to an anchor point i is mostly larger than the actual distance d_i, the correct position has to lie somewhere inside the circle described by the anchor point and d_{ri}. With four distance readings, the area where the actual position has to be in can be reduced to a small patch. The measured position is approximated at the barycenter of this patch as shown in figure 3.7. If there is a single wrong measurement, the other three still form a triangular patch, meaning that the localization is affected, but only to a certain point. The drawback of this procedure is that it requires distance measurements to all sides of the tag and thus, only works within the circumference of given anchor points. However, since we already decided to implement a roaming algorithm for the reasons explained in section 3.1.3, these restrictions do not apply to our use case.

The algorithm consists of these steps:

1. Rasterization of the space inside the circumference of the anchor points.
2. Calculation if a given rasterized point is within the measured distance to a given anchor point.

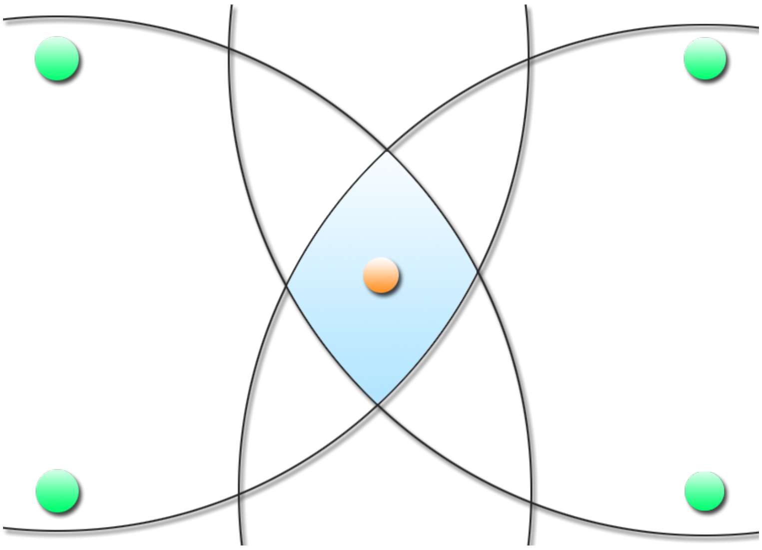

Figure 3.7: Circle Intersection: the measured position is in the center of the small patch resulting from the intersection of the circles described by the anchor points and the measured distances

3. Combination of these calculations by a logical AND operation.

4. Calculation of the center of the patch.

The space inside the circumference of the anchor points is rasterized with a step size of 1 m (cf. figure 3.8 a)). Then for each cell (x, y) and its center c_{xy} with distance $d_{ic_{xy}}$ to anchor point i and each measured distance r_i, we calculate the boolean

$$p_{xy} = \bigwedge_i \left(d_{ic_{xy}} \leq r_i + 1\right)$$

(cf. figure 3.8 b) and c)) to describe if the center of a cell lies within all four circles. We add 1 m to r_i to take into account the error introduced during the rasterization, otherwise it would be possible to get no marked patch at all and thus no result. Finally the barycenter of the resulting patch is calculated and returned as the calculated position (cf. figure 3.8 d)). The presented algorithm has an intrinsic error but it is acceptable since it is of the same size as the measurement error.

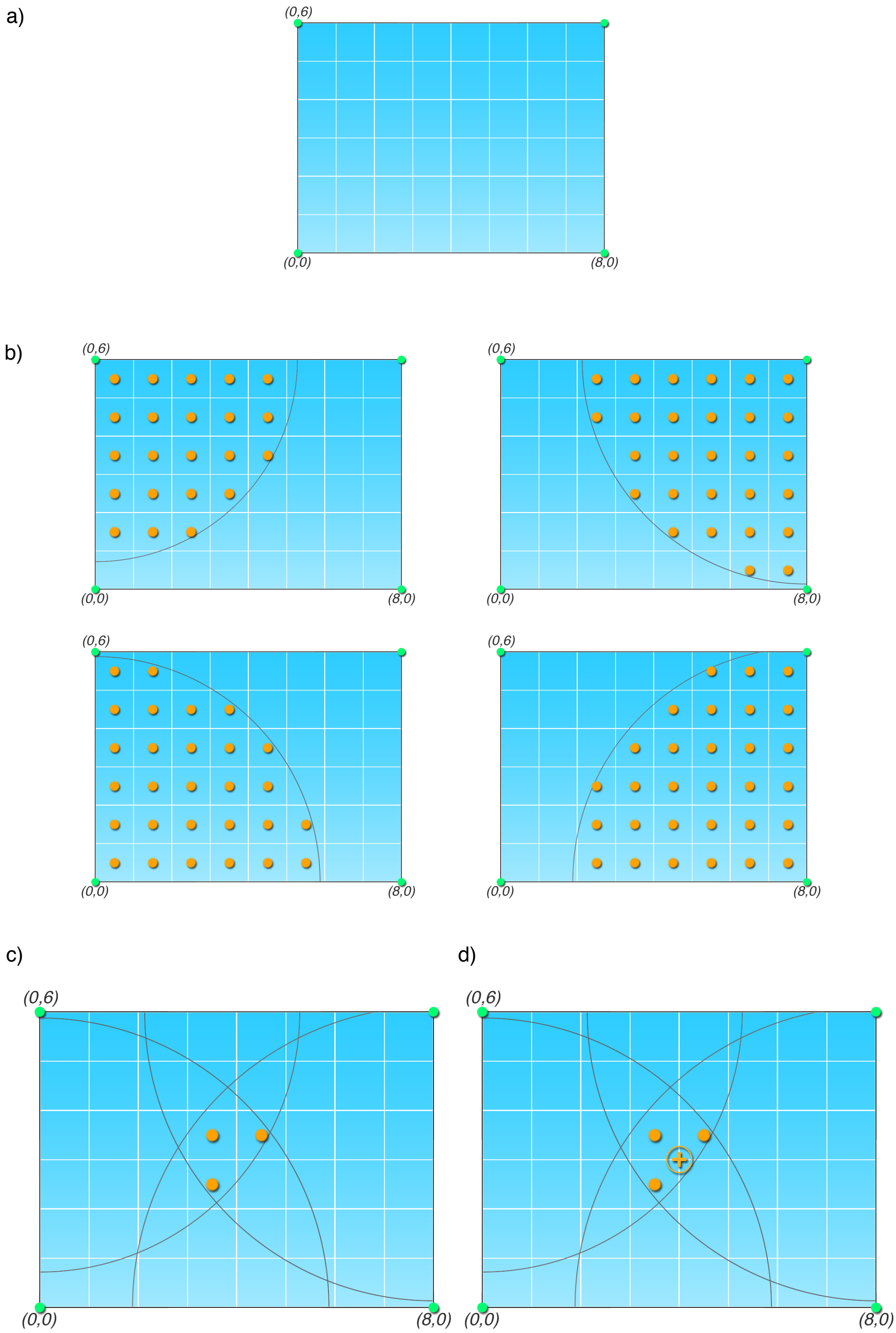

Figure 3.8: The Localization Algorithm: a) the circumference is rasterized **b)** fields where the distance from the center to the anchor point is smaller than d_{ri} get marked **c)** the markers are multiplied field-wise **d)** the calculated position is the center of gravity of all marked fields.

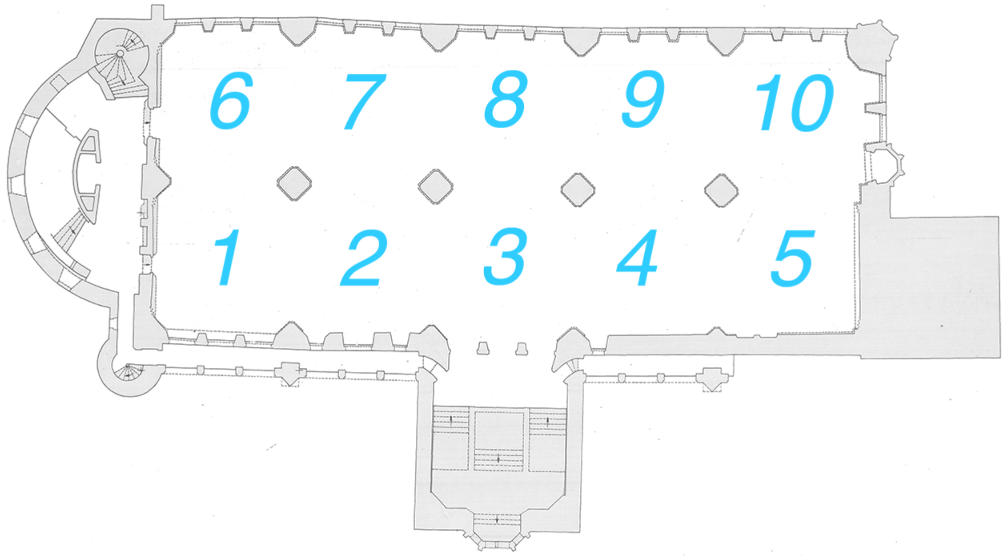

Figure 3.9: The different Zones in the Coronation Hall: The division in 10 zones allows to track more tags simultaneously, but also requires a roaming mechanism.

3.3 Roaming Procedure

For the reasons described in section 3.1.3 we need to divide the coronation hall in 10 zones as figure 3.9 shows. We designed a roaming mechanism to realize the handover between the different zones. Knowing the location of the tag, we can adapt the set of anchor points used for the localization accordingly. To accomplish this, we first store the neighborhood information for each zone as north, south, west, and east value if applicable. If there is no neighboring zone the corresponding property remains unassigned. Around the boundary of two zones, a four meter wide transition zone is defined (cf. figure 3.10). Whenever two consecutive localizations lie in this transition area, two of the four anchor points (e.g., 2 and 4) get changed (cf. figure 3.10 b)) such that the tag is always within the circumference of four anchor points. As the tag gets localized twice in the new zone, the two remaining anchor points (1 and 3) are changed as well (to 5 and 7) and the roaming procedure is complete.
This simple algorithm solves both the coverage problem as well as the scalability problem. However, keep in mind that

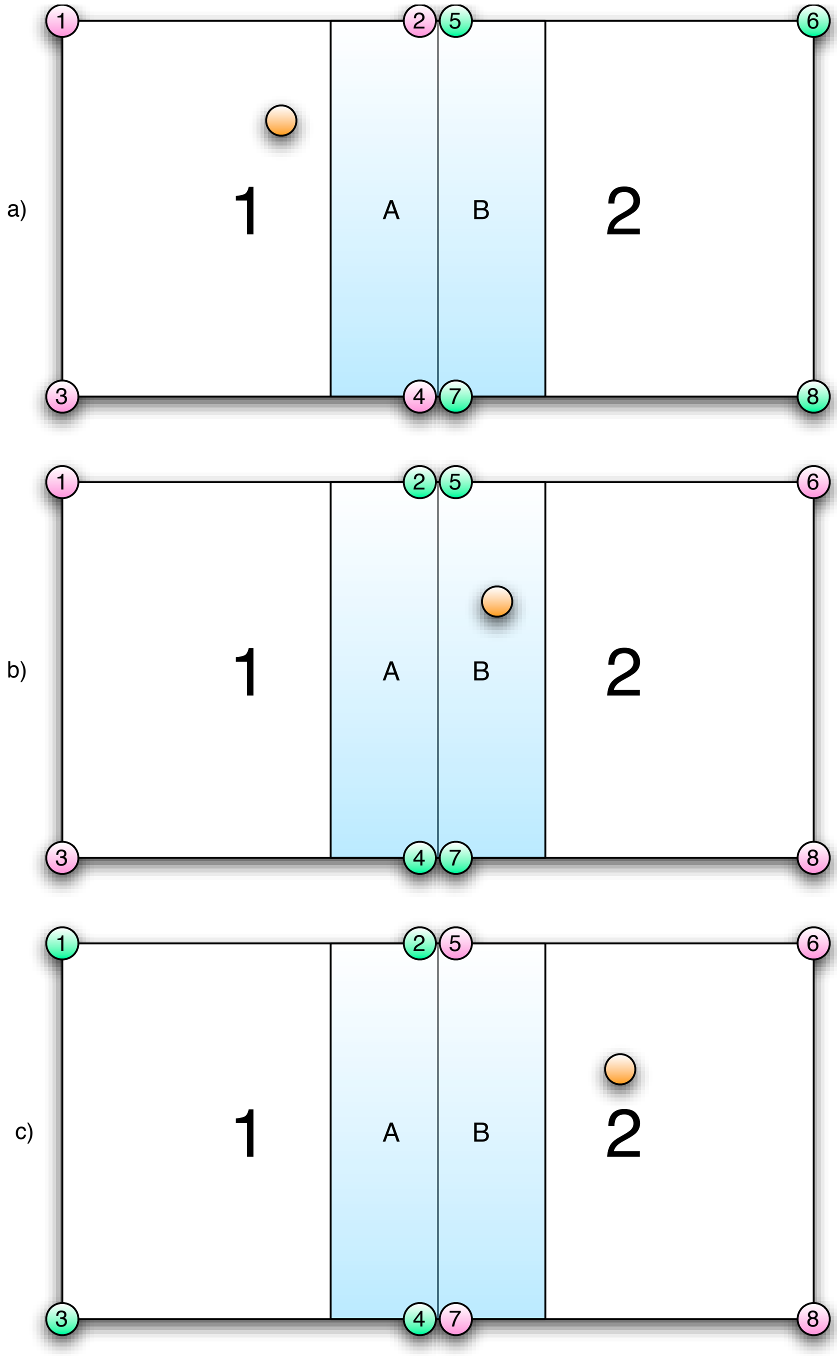

Figure 3.10: The Roaming Procedure: a) Starting in zone 1 with the four active anchor points 1-4. **b)** Whenever the tag is localized two consecutive times in the transition area (A+B), two of the four active anchor points get changed. **c)** When the tag gets localized twice in the new zone (2), but outside of the transition area B, the remaining anchor points also get changed to complete the procedure.

the maximum number of tags trackable per zone is slightly reduced because tags from neighboring zones might use some anchor points during the roaming.

3.4 The Headphone Module

To track the user's position and head orientation we developed a small hardware module that contains the nanoLOC AVR module, a three-axis compass with tilt compensation, and some additional hardware, i.e., power regulation and level converters. Our current prototype is depicted in figure 3.11. The compass obviously needs to be mounted on the head since we need to measure the head orientation and not the heading of the body. This is extremely important as tuning the head is the most important way to resolve the front/back ambiguity in sound source localization. The tracking parts however could have been placed on the mobile device since the position difference between device and head is negligible. The integration on the headphone has several reasons; First, a fixation on the mobile device is problematic when Wi-Fi is used because both systems work on the same frequency the very sensitive nanoLOC receiver would get saturated by the Wi-Fi signal.

The nanoLOC module comes with an Atmel ATmega644V microcontroller to drive the NA5TR1 nanoLOC chip. The ATmega platform is a very popular microcontroller family with a lot of integrated features, e.g., I^2C-bus, universal asynchronous receiver transmitter (UART) for serial communication, serial peripheral interface (SPI) and many others. The SPI-interface on the microcontroller is reserved for the nanoLOC-Chip, but all other connections are available through external pins. This makes it very easy for us to extend the board according to our needs [Nanotron Technologies GmbH, 2008]. The microcontroller serves as central processing and communication unit; it controls the nanoLOC chip, reads the heading from the compass and sends the motion tracking data to the mobile device. The logical connections are shown in figure 3.12 The three-dimensional compass is a HMC6343 from Honeywell [Honeywell, 2008] and connected to the ATmega

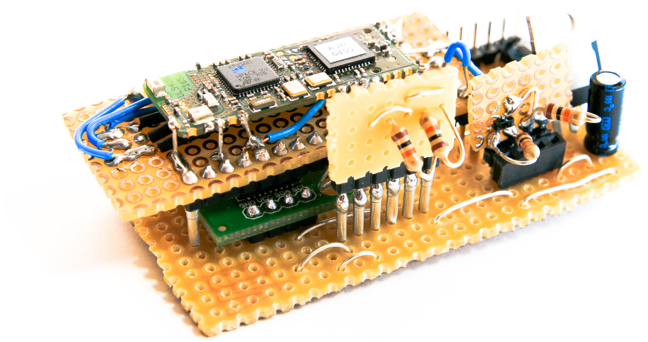

Figure 3.11: The CORONA Headphone Module Prototype: the nanoLOC AVR module is visible on top. The compass is located on the bottom left. The other components are level converters and the power supply.

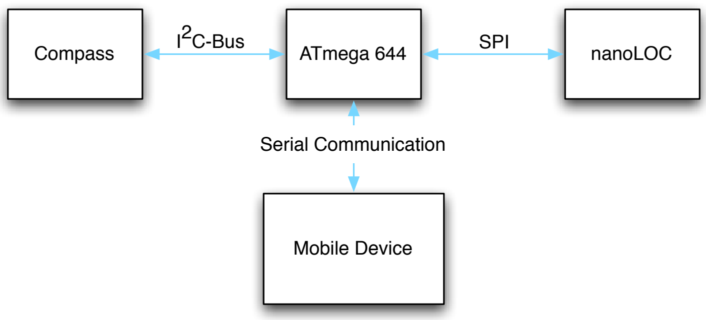

Figure 3.12: The Logical Connections on the Headphone Module.

Heading accuracy is more important than fast updates

via I^2C. The compass is set to continuous operation mode and the values are read out at the highest specified rate of 10 Hz. Our first prototype was equipped with a faster compass module (CMPS03), but without tilt compensation. This means that whenever the head is inclined, the heading information becomes inaccurate. A small preliminary test at our chair showed that a correct heading information independent of the head inclination is much more important than a fast refresh rate. The HMC6343 seems to be the most adequate as it is a fully integrated solution in a small form factor package. Even if its maximal refresh rate is much slower than our initial specifications, it suffices to create a plausible audio environment.

The headphone module is connected to the iPhone by serial line

The connection to the mobile device is done via serial line (universal asynchronous receiver/transmitter, UART) which is a very simple and widely available communication technology. The hardware connection requires two wires, receive (RX) and transmit (TX). On the software side, both ends have to be configured to the same parameters, i.e., speed (19200 baud), message length (eight bit), the number of stop bits (one), and the flow control (none) or as a short notation 19200 8N1. Flow control is used to manage the communication between two devices with different processing power. If the receiver cannot handle any more incoming data, the transmitter gets a signal that tells him to wait until the already sent data is processed. Flow control is not important in our case, as the amount of data transmitted is rather small and the receiver – the iPhone – has a substantially higher processing power than the ATmega microcontroller.

The headphone module is powered from the mobile device which offers a 3.3 V output for external hardware. The compass can directly be powered from this source whereas the nanoLOC board requires a power regulator that outputs 2.5 V. Because of the different core voltages, we need two level converters: one for the I^2C bus and one for the serial line such that the microcontroller can communicate with the higher voltage components. The complete wiring scheme can be found in appendix A.

To transmit the data collected from the compass and the tracking system and to react to commands from the host applications, we defined a simple communication protocol. Most messages are transmitted from the headphone module to the mobile device, the other direction is only used for short control messages. However, we need the bidirectional communication as we have to tell the nanoLOC which set of anchor points it has to use for localization during the roaming procedure.

Data Transmission

Hxxxy this message contains the heading information. The four characters xxxy have to be interpreted as a floating point value $xxx.y$ and represents the heading as a value between 0.0 and 359.9

RXYZ is a ranging information message. X and Y are the low and hight address byte respectively of the anchor point address. The remaining six characters (Z) represent the distance to the anchor point as a floating point value with three decimals. If a ranging operation fails, the value is -1.000

RC ranging complete message. Helps the mobile device to evaluate the incoming data.

Control Commands

Cx turns the compass on (x=1) or off (x=0)

Tx turns the ranging on (x=1) or off (x=0)

TAxyyyzzz changes the address of anchor point at position x to the address defined by the low byte yyy and the high byte zzz

TRxx sets the tracking refresh rate to a value between 1 Hz and 99 Hz.

3.5 Further Improvements

The refresh rate of one or two times per second is rather coarse, even if the variation in distance is less important than the orientation in the stereo panorama. To generate location estimations between the measurements, we could use a two dimensional Kalman filter. A Kalman filter is a probabilistic model to estimate the state of a dynamic system based on series of noisy measurements. Nevertheless, the filter affects the directness of the system as it adapts to every new incoming measurement. However, the previous information is integrated in the position calculation, which means that the position of a running visitor that suddenly stops will be estimated moving for some time after the stop.

Another option is to use the data available from the accelerometers placed in the iPhone and the compass module. If there is no measurable substantial movement, the position can be assumed stable and the localization can be stabilized by averaging over several measurements. The values of both sensors should be considered because the values from the accelerometer in the compass are affected by head movement even if the visitor is standing still.

3.6 OpenAL

This section is rather short because the audio rendering process is totally delegated to the OpenAL framework that provides a good programming interface. The reasons why we chose OpenAL are already described in section 2.4. A small OpenAL demo application like the one shown in figure 3.13 is a good starting point to grasp the effect of spatial audio rendering.

There are three basic primitives required for a simple example: sources, buffers and a listener. A source is a point in space that emits a sound to a certain direction, depending on a set of parameters. These can be omnidirectional or have a very small emitting angle. Every source has to be connected to a buffer which contains the audio data to be played. OpenAL itself does not provide any decoding mechanisms for complex audio file formats, but only accepts raw PCM data. To play a compressed audio file, e.g., MP3 or AAC, a decoding function has to be implemented that fills the buffer with a raw PCM stream.

OpenAL renders an acoustic scenery from given sound sources and a specific listener orientation and position

Finally, the listener is an object that gets assigned a position and orientation. The rendering engine then generates a stereo audio stream, according to the given parameters, that provides a plausible acoustic impression of the virtual scenery.

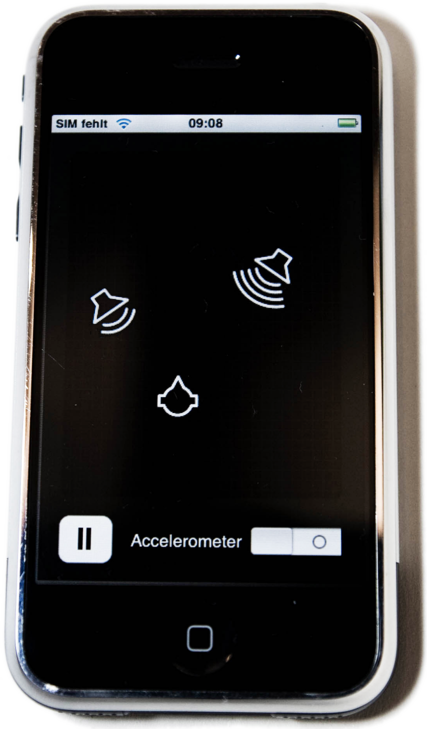

Figure 3.13: Simple OpenAL Demo Application Running on the iPhone: the listener and the sound sources can be moved around and the rendering engine adapts the audio stream accordingly.

In our implementation, we do not use the possibility of defining the elevation of a sound source because this complicates the rendering process and does not improve the experience as we do not plan to use sound sources with different elevations.

Using IMA4 encoding reduces the file size to a fourth

The iPhone only offers the possibility to decode one MP3/AAC stream at a time because this is done in a dedicated hardware unit. Decoding the several compressed audio stream in parallel on the main CPU would affect the rendering performance because of limited processing power. In contrast to the very disk space greedy PCM encoding used in AIFF and WAV files, there are other encoding techniques like IMA4 that reduce the size of a file to a fourth of the PCM variant. The use of such AIFC audio files is a bit more complex because the encoding has to be converted to PCM like for compressed audio, but this process does not require much processing power.

Chapter 4

Evaluation

In this chapter, we analyze how far we met the requirements described in chapter 1. On the on hand, we have the tracking system which should have an accuracy of less than two meters error, on the other hand we have the interaction with the audio space that should be plausible.

The analysis of the interaction with the audio space itself is currently work in progress. Especially the questions, how to deal with a location that is close to a sound source is very important. If the location jumps around a the source, the signal also jumps in the stereo panning which is a very unnatural and disturbing behavior.

4.1 Distance Measurements

To get a feeling for the accuracy of the nanoLOC AVR module, we first did a couple of simple ranging tests. We fixed the tag and one anchor point at a measured distance from each other. We then evaluated the distance measured by the nanoLOC system and got the results shown in table 4.1. As we can see, the error increases more or less with increasing distance. This can be explained by the increasing influence of multipath fading at large distances. During the experiments we also observed that the SMD-antenna on the nanoLOC AVR module does not have a totally equal

distance (in m)	average error (in m)	std. deviation (in m)	no. of measurements
1	0.78	0.20	63
2	0.96	0.23	54
4	0.66	0.32	39
6	1.44	0.85	31
8	1.40	1.04	40
10	2.08	1.35	41

Table 4.1: Distance Measurement Results.

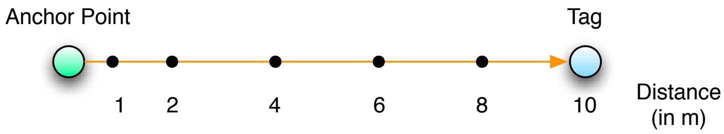

Figure 4.1: The Ranging Evaluation Experiment: the anchor point and the tag were placed at measured real world distances. We then evaluated the distance values returned by the nanoLOC system.

dispersion, i.e., depending on the orientation the measured values differ, like the five meter error in the six meter measurement. This can be attenuated by using external antennas on the anchor points with a more regular dispersion. The results presented in section 4.2 show, that the impact of these errors on the localization accuracy is small and acceptable for our needs.

4.2 Localization Accuracy

More important than the ranging accuracy is the localization accuracy as this is the value that is finally relevant for the user experience. To evaluate our success, we set up a square field of 10 m $\times$ 8 m covered by four anchor points. We also marked several measured positions and compared the real world coordinates with those that both, the least squares and the empirical algorithm returned (cf.

figure 4.2). The anchor points and the tag were mounted on tripods at one and a half meters height. As in preliminary test no difference in accuracy was determinable between the nanoLOC developer boards and the nanoLOC AVR module we used the developer boards because they can be powered with batteries, which makes it easier to arrange a test field. We used our prototype with the compass module as tag.

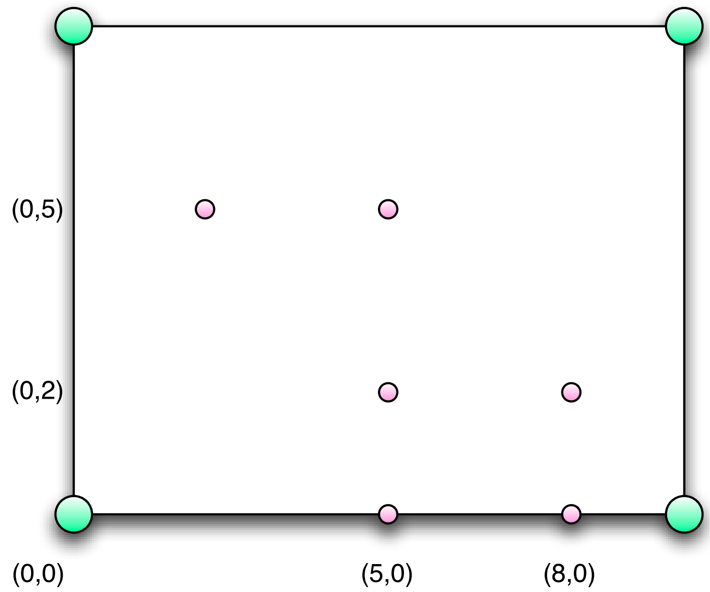

Figure 4.2: The Test Setup for the Localization Algorithms: the 10 m× 8 m wide area is covered by four anchor points. The small circles describe the positions of the location measurements

As we can see from the results (table 4.2, 4.3, and summarized in table 4.4), the empirical approach, except for one case, works better than the least squares approach. All results are statistically significant as a two-tailed unpaired t-test based on the original test data revealed (cf. table 4.5).

The performance at the borders can be improved

The bad performance near the borders can be explained by the design of the algorithm. The estimated position is defined as the barycenter of the marked patch. Because the rasterization takes only place within the circumference of

	x-axis		y-axis		total	
real position	mean error	std. deviation	mean error	std. deviation	mean error	std. deviation
2 × 5	0.64	0.28	0.58	0.21	0.90	0.15
5 × 0	1.01	0.47	2.15	1.06	2.51	0.59
5 × 2	0.16	0.11	1.91	0.16	1.91	0.17
5 × 5	2.16	0.44	1.04	0.27	2.40	0.51
8 × 0	0.27	0.16	4.42	0.20	4.43	0.20
8 × 2	1.20	0.20	0.42	0.39	1.33	0.16

Table 4.2: Localization Measurement Results for the Least Squares Algorithm. All values in m.

	x-axis		y-axis		total	
real position	mean error	std. deviation	mean error	std. deviation	mean error	std. deviation
2 × 5	0.15	0.12	0.64	0.28	0.29	0.14
5 × 0	1.12	0.24	2.63	0.25	2.79	0.17
5 × 2	0.26	0.19	1.74	0.26	1.77	0.26
5 × 5	0.37	0.11	0.45	0.12	0.60	0.14
8 × 0	0.32	0.11	2.25	0.16	2.57	0.16
8 × 2	0.90	0.09	0.59	0.10	1.20	0.20

Table 4.3: Localization Measurement Results for the Empirical Algorithm. All values in m.

the anchor points, the barycenter always tends to the inside as there is no counterweight that pulls it back to the border. To improve the algorithm the rasterized area should exceed the circumference by two meters, such that the balance on the borders is restored.

The achieved accuracy is within our expectations

The mean error of 1.54 m for the empirical approach is within our expectations, however, the standard deviation of 1.02 m poses some challenges for the creation of our user experience. A small test showed that the position jumps cause a perpetual movement in the stereo panorama, even if the position is far from the source. Further development is needed to smoothen those jumps, e.g., Kalman filtering or averaging over several measurements.

An evaluation of the impact of smaller step sizes in the rasterization process should also be done. This should in-

	least Squares		empirical	
real position	mean error	std. deviation	mean error	std. deviation
2×5	0.90	0.15	0.29	0.14
5×0	2.51	0.59	2.79	0.17
5×2	1.91	0.17	1.77	0.26
5×5	2.40	0.51	0.60	0.14
8×0	4.43	0.20	2.57	0.16
8×2	1.33	0.16	1.20	0.20
Overall	2.25	1.24	1.54	1.02

Table 4.4: Comparison of the Localization Algorithms. All values in m.

real position	$\hat{t}$	degrees of freedom	t-value	statistically significant
2×5	16.83	62	2.00	Yes
5×0	2.54	58	2.00	Yes
5×2	2.18	40	2.02	Yes
5×5	18.82	60	2.00	Yes
8×0	4.22	70	1.99	Yes
8×2	9.70	40	2.02	Yes

Table 4.5: The t-Test Results.

crease the sensitivity. The question to be answered is if this also increases the accuracy or reduces the variance.

Chapter 5

Summary and Future Work

The overall goal of this thesis was to provide the technical backbone for an interactive audio experience in the coronation hall. To implement such an experience we need a real time location tracking system and a spatial audio rendering engine. Both aspects have been handled in this thesis, with a focus on the location tracking. The audio rendering question was solved quite easily with the availability of the OpenAL framework for the iPhone, our designated mobile hardware platform. In this chapter we summarize our contributions and developments and we describe some aspects that could not be included in this thesis anymore.

5.1 Summary and contributions

We first evaluated several tracking technologies under specific assumptions for CORONA. Some of the analyzed real time location systems have an accuracy in the order of centimeters which is much better than ours, but they need at least a connection from the anchor points to a central server, which is unfeasible for a usage in the coronation hall. Other systems have a lower accuracy and base on popular technology like Wi-Fi or Bluetooth, but failed in our own tests. Finally, we decided to use the nanoLOC technology – even if the specified accuracy of two meters is not state-of-the-art – because of its completely autonomous and flexible architecture. The anchor points only require a power connection and the tags can easily be powered from a mobile device. The localization process completely takes place on the mobile device, so there is no need to worry about privacy issues. By combining the nanoLOC AVR module with a three dimensional compass we developed a small motion tracking device that provides good enough information to render our interactive audio space.

Motion Tracking Hardware

Localization Algorithms

We implemented and analyzed two localization algorithms for approximate distances. The more mathematical least squares approach is not feasible for our application as it is too sensitive to erroneous measurements. The empirical approach has an intrinsic error, but this is negligible regarding the accuracy of the ranging data. This approach is more robust and reaches the required accuracy of one to two meters.

Roaming Procedure

We designed a roaming procedure that is necessary to cope with the bad scalability of the nanoLOC system and the unfavorable architectural conditions of the coronation hall. The division of the hall in 10 zones covered by four anchor points each avoids coverage problems caused by the stone pillars and allows us to track approximately 100 tags simultaneously.

We developed a software that combines the ranging information returned from the nanoLOC system to a location using the empirical algorithm and renders the audio scene accordingly through OpenAL.

5.2 Future work

The empirical localization algorithm in it's current state performs bad on the borders of the circumference of the anchor points. We proposed some modifications to the original algorithm in section 4.2 that should significantly improve the accuracy at the borders.

Kalman Filter

Although we have a working prototype it still needs some improvements. The implementation of the Kalman filter and its evaluation are the next steps to go. The Kalman filter needs some time to react to state changes, which can possibly affect the responsiveness of the system. The virtual position should not continue moving when the visitor stops. We should also consider the possibility of omitting the position interpolation if the volume differences that come along with movement are negligible. Fine-tuning the attenuation factors in OpenAL is possibly more effective than the implementation of complex location interpolation mechanisms.

Visitors close to sound sources are problematic

The question of how to handle a visitor that is close to a sound source also needs to be analyzed. A visitor standing at the exact position of the sound source will experience jumps in the stereo panorama due to noisy and approximate localization. This is highly irritating and should be avoided by designing a clever strategy, e.g., that switches the source position from static to dynamic and back. Another possibility is to use the metaphor of story telling people that move aside to let you go past. This realistic behavior is easy to understand for the visitor and opens some possibilities for explorative features, e.g., following a moving sound source to hidden information.

The interaction with a sound source is also determined by the type of playback used. An endless loop is not useful for two reasons: first, the end and the beginning of the sample have to be connected somehow, otherwise, there would be an unnatural break. The second point is that it requires the visitor to stay at a location until the sample restarts to get the full context. However, reverting back to a trigger based variant does not allow navigation by ear which requires gentle sound information for navigation. A solution could be a gentle murmur and whispering is used for navigation and when the visitor comes close enough, the corresponding explanatory sample is played.

Appendix A

Technical Description of the Hardware

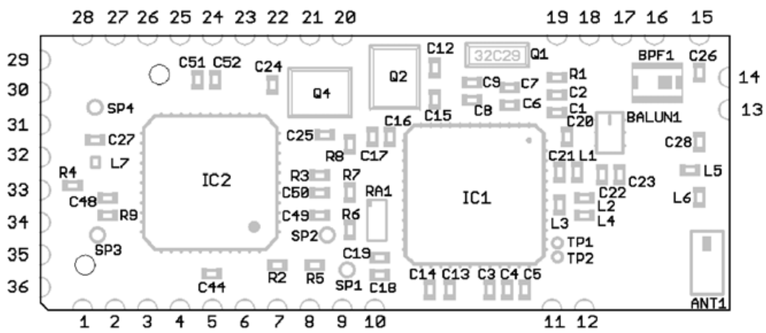

Figure A.1: The PCB Layout of the nanoLOC Module

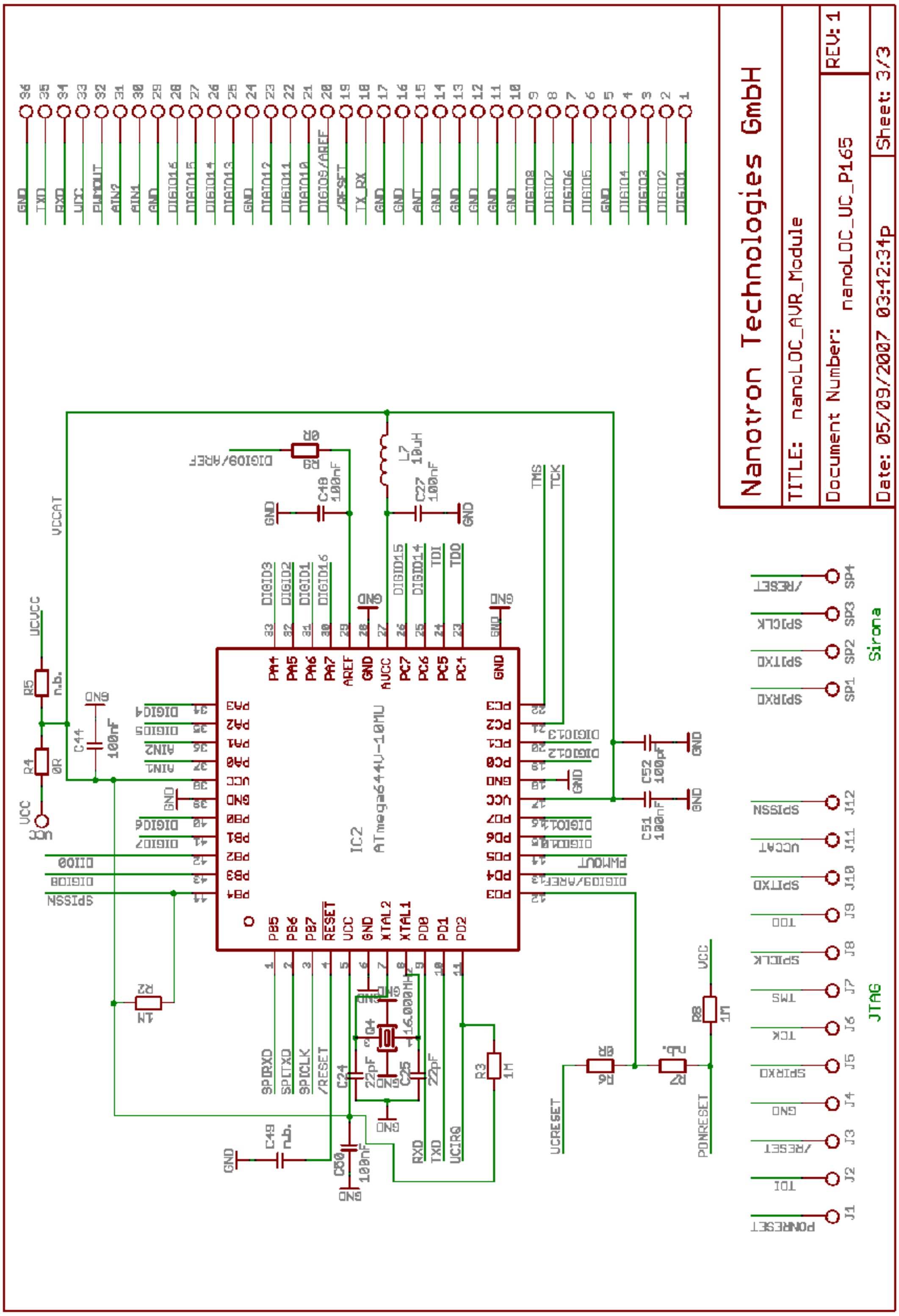

Figure A.2: Wiring Diagram of the nanoLOC Module

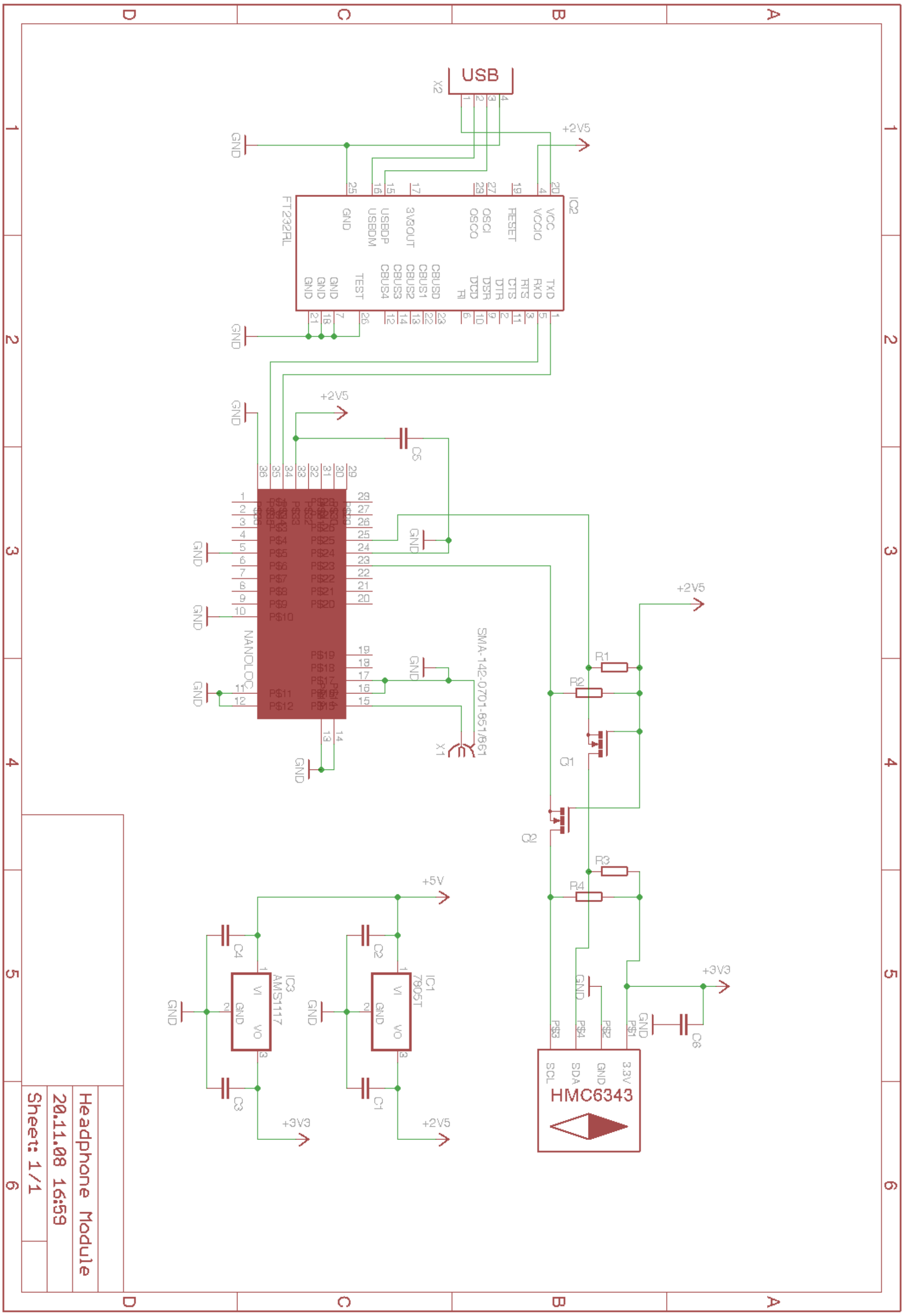

Figure A.3: Wiring Diagram of the Headphone Module

Bibliography

V.R. Algazi and R.O. Duda. Immersive Spatial Sound for Mobile Multimedia. In *Proceedings of the Seventh IEEE International Symposium on Multimedia*, pages 739–746. IEEE Computer Society Washington, DC, USA, 2005.

Paramvir Bahl and Venkata N. Padmanabhan. RADAR: An in-building RF-based user location and tracking system. In *INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies.*, volume 2, pages 775–784, 2000.

Benjamin B. Bederson. Audio augmented reality: a prototype automated tour guide. In *CHI '95: Conference companion on Human factors in computing systems*, pages 210–211, New York, NY, USA, 1995. ACM. ISBN 0-89791-755-3. doi: http://doi.acm.org/10.1145/223355.223526.

Björn Bischof. *[AudioCave] Mord im Bagdad-Express, eine auditiv-interaktive Hörspiel-Installation*, November 30th, 2006. URL `http://netzspannung.org/cat/servlet/CatServlet?cmd=netzkollektor&subCommand=showEntry&forward=&entryId=331966&version=print&print=all&lang=de`.

J. Breebaart, J. Herre, L. Villemoes, C. Jin, K. Kjöring, and J. Plogsties. Multi-Channel Goes Mobile: MPEG Surround Binaural Rendering. In *29th Audio Engineering Society Conference*, September 2006.

Gerhard Eckel. The vision of the LISTEN project. In *VSMM '01: Proceedings of the Seventh International Conference on Virtual Systems and Multimedia (VSMM'01)*, page 393, Washington, DC, USA, 2001. IEEE Computer Society. ISBN 0-7695-1402-2.

Michael Eisenacher. *Optimierung von Ultra-Wideband-Signalen (UWB)*, volume 16 of *Forschungsberichte aus dem Institut für Nachrichtentechnik der Universität Karlsruhe (TH)*. Inst. für Nachrichtentechnik, Karlsruhe, 2006. doi: http://digbib.ubka.uni-karlsruhe.de/volltexte/1000004809.

Sandra Gebbensleben, Jana Dittmann, and Claus Vielhauer. Multimodal audio guide for museums and exhibitions. Technical report, Otto-von-Guericke University of Magdeburg, Institute of Technical and Business Information Systems, 2006.

Alessandro Genco, Salvatore Sorce, and Giuseppe Scelfo. Bluetooth base station minimal deployment for high definition positioning. In *MOBIQUITOUS '05: Proceedings of the The Second Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services*, pages 454–460, Washington, DC, USA, 2005. IEEE Computer Society. ISBN 0-7695-2375-7. doi: http://dx.doi.org/10.1109/MOBIQUITOUS.2005.15.

S. Gezici, Z. Tian, GB Giannakis, H. Kobayashi, AF Molisch, HV Poor, and Z. Sahinoglu. Localization via ultra-wideband radios: a look at positioning aspects for future sensor networks. *Signal Processing Magazine, IEEE*, 22(4): 70–84, 2005.

J. Hallberg, M. Nilsson, and K. Synnes. Positioning with bluetooth. *Telecommunications, 2003. ICT 2003. 10th International Conference on*, 2:954–958 vol.2, Feb.-1 March 2003. doi: 10.1109/ICTEL.2003.1191568.

Georg K. Helg and Jürgen Linden. *Vom Kaiserglanz zur Bürgerfreiheit*. Shake Verlag, Aachen, 2006.

Honeywell. 3-Axis Compass with Algorithms. Datasheet, July 2008.

Dr. Johannes Horn. Location Engine 2006 - Functional Principle. Presentation, October 2007.

F. Izquierdo, M. Ciurana, F. Barcelo, J. Paradells, and E. Zola. Performance evaluation of a toa-based trilateration method to locate terminals in wlan. *Wireless Pervasive Computing, 2006 1st International Symposium on*, pages 1–6, Jan. 2006. doi: 10.1109/ISWPC.2006.1613598.

J. Kannisto, T. Vanhatupa, M. Hannikainen, and T.D. Hamalainen. Software and hardware prototypes of the IEEE 1588 precision time protocol on wireless lan. *Local and Metropolitan Area Networks, 2005. LANMAN 2005. The 14th IEEE Workshop on*, Sept. 2005. doi: 10.1109/LANMAN.2005.1541513.

Kent Lyons, Maribeth Gandy, and Thad Starner. Guided by voices: An audio augmented reality system. In *Proceedings of Intl. Conf. on Auditory Display (ICAD) 2000*, 2000.

MagicMap. Comparison to other systems. `http://www2.informatik.hu-berlin.de/rok/MagicMap/MagicMap_details.html`, February 16th, 2006.

John C. Middlebrooks and David M. Green. Sound Localization by Human Listeners. *Annual Reviews in Psychology*, 42(1):135–159, 1991.

William S. Murphy Jr. Determination of a position using approximate distances and trilateration. Master's thesis, Colorado School of Mines, July 2007.

Nanotron Technologies GmbH. nanoLOC AVR Module - Technical Description, February 2008.

Nanotron Technologies GmbH. nanoNET Chirp Based Wireless Networks. Whitepaper, February 2007a.

Nanotron Technologies GmbH. Real Time Location Systems (RTLS). Whitepaper, May 2007b.

Nissanka Bodhi Priyantha. *The Cricket Indoor Location System*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA,, June 2005.

Spread Spectrum Scene. An introduction to indoor radio propagation, June 1st, 2001. URL `http://www.sss-mag.com/indoor.html`.

J. W. Strutt. On our perception of sound direction. *Philos. Mag.*, 13:214–232, 1907.

Lucia Terrenghi and Andreas Zimmermann. Tailored audio augmented environments for museums. In *IUI '04: Proceedings of the 9th international conference on Intelligent user interfaces*, pages 334–336, New York, NY, USA, 2004.

ACM. ISBN 1-58113-815-6. doi: http://doi.acm.org/10.1145/964442.964523.

Héctor Velayos and Gunnar Karlsson. Limitations of range estimation in wireless LAN. In *Proceedings 1st Workshop on Positioning, Navigation and Communication*, March 2004.

Ron Wakkary and Marek Hatala. ec(h)o: situated play in a tangible and audio museum guide. In *DIS '06: Proceedings of the 6th conference on Designing Interactive systems*, pages 281–290, New York, NY, USA, 2006. ACM. ISBN 1-59593-367-0. doi: http://doi.acm.org/10.1145/1142405.1142448.

Olivier Warusfel and Gerhard Eckel. LISTEN - augmenting everyday environments through interactive soundscapes. Technical report, IRCAM, Paris, France and Fhg-IMK, Bonn, Germany, 2002.

DN Zotkin, R. Duraiswami, and LS Davis. Rendering localized spatial audio in a virtual auditory space. *Multimedia, IEEE Transactions on*, 6(4):553–564, 2004.

Index

Typeset March 2, 2009