Simplifying Orientation Measurement for Mobile Audio Augmented Reality Applications

Florian Heller **Aaron Krämer** **Jan Borchers**
RWTH Aachen University
52056 Aachen, Germany
{flo,aaron,borchers}@cs.rwth-aachen.de

ABSTRACT

Audio augmented reality systems overlay the physical world with a virtual audio space. Today's smartphones provide enough processing power to create the impression of virtual sound sources being located in the real world. To achieve this, information about the user's location and orientation is necessary which requires additional hardware. In a real-world installation, however, we observed that instead of turning their head to localize sounds, users tend to turn their entire body. Therefore, we suggest to simply measure orientation of the user's body — or even just the mobile device she is holding — to generate the spatial audio.

To verify this approach, we present two studies: Our first study in examines the user's head, body, and mobile device orientation when moving through an audio augmented reality system in a lab setting. Our second study analyzes the user experience in a real-world installation when using head, body, or device orientation to control the audio spatialization. We found that when navigating close to sound sources head tracking is necessary, but that it can potentially be replaced by device tracking in larger or more explorative usage scenarios. These findings help reduce the technical complexity of mobile audio augmented reality systems (MAARS), and enable their wider dissemination as mobile software-only apps.

Author Keywords

Audio Augmented Reality; Mobile Devices; Binaural Rendering; Spatial Audio; Orientation; Presence.

ACM Classification Keywords

H.5.1 Information Interfaces and Presentation (e.g. HCI): Multimedia Information Systems

INTRODUCTION

Spatial audio rendering applies special filters to recorded audio signals, giving the user the impression that the sound emanates from a source located in the physical space. Audio augmented reality applications use this technique to overlay the physical space with a virtual audio space that the user experiences via headphones. This virtual audio space can add atmosphere and context to the real world through an ambient soundscape [21], and it can contain virtual sound sources, possibly connected to physical objects [19].

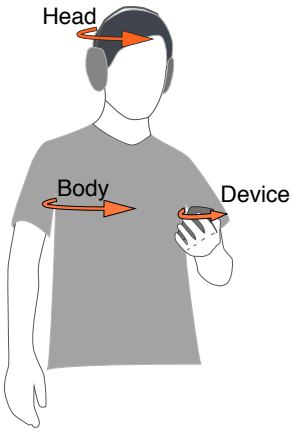

Figure 1. Three possible points of measurement that can be used to control orientation in virtual audio spaces. We found that head tracking is most realistic, but that device orientation is easier to measure and can be used as a proxy for body and head orientation in certain situations.

While traditionally the spatial audio was rendered on complex centralized hardware [7, 17], current mobile devices can render this audio without limitations in realism compared to desktop solutions [15], supporting large numbers of parallel users and allowing for easy dissemination of such systems. In addition, these mobile solutions suggest new applications, such as tools that do not require visual attention and solutions for the visually impaired. One example are auditory menus that are arranged spatially around the user's head [6]. Mobile audio augmented reality systems (MAARS) also allow for navigation systems that offer serendipitous discovery instead of guiding the user strictly towards a certain target [16, 19]. What all implementations of audio augmented reality have in common, however, is that they need to make the sound sources appear to be located in the physical space.

For this, a rendering algorithm needs to know position and head orientation of the listener to create this spatial impression. Under the assumption that a realistic rendering provides the best user experience, most audio augmented reality projects rely on complex dedicated hardware to determine the users position and head orientation. This is also true

CHI 2014, April 26–May 1, 2014, Toronto, Ontario, Canada.

ACM 978-1-4503-2473-1/14/04..$15.00.
http://dx.doi.org/10.1145/2556288.2557021

for CORONA, our own virtual audio space installation, which uses a headphone-mounted digital compass. However, in everyday use we observed that many CORONA users, instead of turning only their head to orient themselves in the audio space, turn their entire body. If this behavior prevails, then the experience could be implemented using the sensors available in a modern smartphone, thus avoiding the need for a separate compass. Such a system could then be easily deployed to a broader public as a software-only solution, e.g., via an app store.

We study the question which source of orientation (Figure 1) can be used for orientation measurement with two approaches: In our first experiment, we measure the difference between head, body, and device orientation for real spatial hearing and for headphone-based spatial audio rendering of the same scene. This deepens our understanding of how users orient towards sound sources and how much our rendering influences this behavior. To see if the effects measured in this experiment actually influence the perceived sense of presence and the ability to navigate in a virtual audio space, we conduct a second experiment with three different compass sensor placements.

Therefore, the key contributions of this work are:

- an analysis of orientation movements when navigating in virtual audio spaces and
- recommendations which point of measurement to use for specific use cases.

INSPIRATION: THE CORONA AUDIO SPACE

Our test case is an audio augmented reality experience deployed in the Coronation Hall (Figure 2) of the historic city hall in Aachen, Germany. This room was the location of coronation feasts for important emperors in medieval Europe, including Charlemagne. Of these festivities no apparent visual traces remain except for a series of coats of arms engraved in the pavement. To bring back the atmosphere of such festivities, we created an audio space that depicts the well-documented coronation feast of Charles V. from 1520. Virtual characters discuss different aspects of the ceremony, providing the visitor with insights in a more personal manner: Maids discuss the order of the dishes, characters at the window describe the festivities for the common people they are watching, and clerics and the king discuss the perils of the Black Death. Since the sound sources are not connected to concrete physical objects but to meaningful locations, we consider this an augmented environment. The CORONA audio space combines the atmosphere of a medieval coronation feast with educational content into an experience of serendipitous discovery. The installation is part of the exhibits in the city hall and open to the public every day.

RELATED WORK

Even without spatial rendering, audio augmented reality can be an immersive addition to an environment. *Riot! 1831* [14] covers an entire place in Bristol, UK with an audio space that users can discover through a scripted story. The space is divided into zones, each one connected with three distinct audio samples, one of which is played while the user is in the zone. On a smaller scale, the *ec(h)o* installation [21] augments a museum with ambient soundscapes related to the overarching topic of the exhibits. Using a portable, tangible artifact users can access more detailed information on the closest exhibit. Unlike Corona, both systems add an engaging atmosphere to a location, but do not focus on simulating a realistic auditory impression.

Figure 2. The historic Coronation Hall where the Corona virtual audio space is deployed.

The potential of such installations can be increased by making the audio display integrate both location and orientation data. Spatial hearing has been studied since the early 20th century [3, 13], and the process of auralization is also well-defined [20]. While early implementations required a complex hardware setup to simulate the basic cues of spatial hearing [7], modern implementations can create a realistic impression and run smoothly on a simple smartphone [15].

AudioGPS [5] is designed for pedestrian navigation. It communicates the designated walking direction through simple panning and uses a Geiger counter metaphor for distance cues. To avoid front/back confusion, the system uses two distinct sounds depending on whether the source is in the frontal hemisphere of the user or not. The heading is extrapolated only from GPS measurements, which means that if the user is stationary, no orientation change can be calculated and thus no directional hint can be given. As Holland et al. already mentioned, this problem could be solved by integrating a digital compass. The Roaring Navigator [16] is an auditory navigation system for a zoo that uses a head-mounted compass for orientation tracking. The different enclosures are represented using sounds of the animals living in it as beacons. On her way to the designated navigation target, the user possibly passes other enclosures and hears the according sounds, which might help to discover animals that she was not looking for. To avoid confusion by overloading the auditory channel, the number of sources is limited to four, and front-back confusion is avoided by muting sources that are behind the listener. Audio augmented reality outdoor navigation tools would clearly benefit from an easier implementation without specialized hardware, as they could use smartphone sensors and GPS and be deployed as software-only solution.

Indoors, different tracking technology is required to get precise location information because GPS is generally unavailable. Part of the *Listen* [17] project was the development of such a tracking system and combining it with a high-quality spatial audio display. The tracking system requires a complex installation and special headphones. With this high technical effort, the visitor has the impression of sound emerging directly from the paintings in the exhibition, but the system can only handle a small number of visitors simultaneously.

From the aforementioned projects, we can already extract one level of connection between the virtual and the physical space, similar to what was defined for visual augmented reality [9]. If the *environment* is augmented, the sound is either ambient or emerges from a location, but not from a specific physical object. Whenever the system generates the impression that the sound emerges from a specific physical object, we talk of augmented *objects*. The second level on which the virtual and physical space are connected is the semantic level. For example, in the Roaring Navigator the beacon sounds are related to enclosures, whereas in [19], statues in a park are augmented with animal sounds that do not relate to the monument. Although a semantic connection between physical landmark and audio representation is often used, the type of sound used can vary. If navigation efficiency is of primary interest, then a beacon-like sound should be used [18, 22] over a continuous sound. Ankolekar et al. [2] compared different types of auditory landmarks regarding the discovery of points of interest in the context of serendipitous exploration of unfamiliar places. They created *sound, speed, music,* and *mixed* audio samples connected to the POI and compared the identification effectiveness with an image. One finding is that musicons, pieces of music that closely match the nature of the landmark, are nearly as good as visual cues and leads to users reporting a more enjoyable experience.

The systems mentioned so far all use an exocentric audio space, i.e., one where the perceived positions of the virtual audio sources do not change when the user moves her head. In contrast to that, the audio space used in [6, 10] are egocentric, which means that the sound sources stay at their position relative to the head of the user. In a lab study, Marentakis et al. examine the impact of feedback and mobility on spatial audio target acquisition. The orientation was only measured at the participants waist and the pointing device. This kind of audio space can also be used in combination with auditory menus [6].

Loomis et al. [7] showed that people are able to navigate towards virtual audio sources even if the rendering only uses a simplified spatial model. Based on the same experimental design, Nicholas Mariette studied the interaction with virtual audio spaces regarding several technical parameters such as head-turn latency, rendering quality, and capture circle diameter for audio sources [11]. His experiments show that a rendering algorithm with a lower horizontal resolution, and thus higher positional blur, can successfully hide measurement errors of the positional sensors while still allowing for successful navigation. In fact, a complex rendering model suffers from sensor inaccuracy to a much higher extent. However, to get a better GPS reading, participants of the study were instructed to walk only in the direction their torso was facing and at a steady, medium pace. The subjects were also instructed explicitly to use head-turns to determine the source direction which may have confounded the findings deduced from the head-yaw measurements.

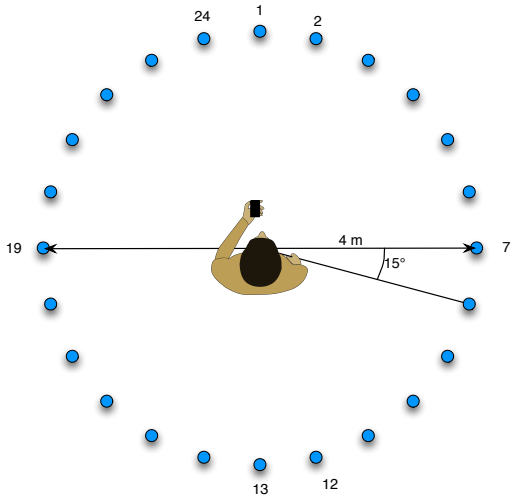

Figure 3. The test setup for the first experiment. We placed 24 loudspeakers in a circle and created a virtual representation of it. We compared head orientation when walking towards a sound source for real spatial hearing and spatial audio rendering.

Vazquez-Alvarez et al. [19] compared navigation performance with different auditory display types in a sound garden implemented in a municipal park. Interestingly, the task completion time in the fully spatialized condition was higher than with, e.g., earcons. However, the authors observed and participants confirmed that this is mostly because they stopped for some time and enjoyed the experience. Vazquez-Alvarez et al. also point out that measuring the difference between body and head orientation would help to further understand the behavior of people experiencing virtual audio spaces.

Overall, related work shows that simple hardware and simple rendering models do not necessarily result in a bad user experience. We wanted to answer the question if using the built-in compass of a smartphone is sufficient to measure listener orientation with a rendering algorithm that is readily available.

MOVEMENT AND ORIENTATION

The goal of our first experiment was to determine if there is a significant difference between the orientation of the head, body, and device when moving towards a sound source. This should give us an indicator where the compass should be placed when developing a virtual audio space. Since nearly every current smartphone has an integrated digital compass, using device orientation is very easy. If the display is not used, measuring body orientation can be achieved by attaching the smartphone such that it rotates with the torso (e.g., through a lanyard), whereas using head orientation requires additional hardware.

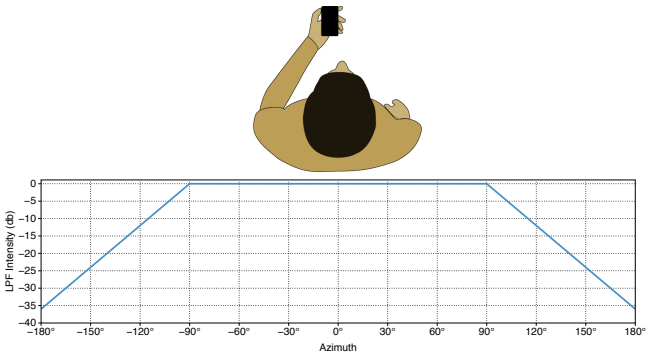

Figure 4. Our improvement of the available spatial audio rendering. To reduce front/back confusion, a low pass filter is applied to sources in the back of the listener. The intensity of the filter is interpolated linearly from 0 dB at 90° to -36 dB at 180°.

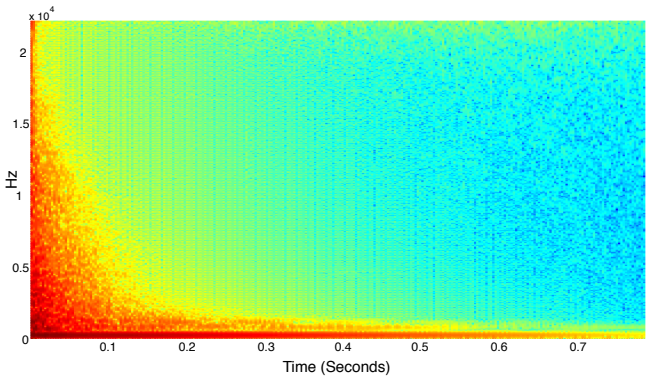

Figure 5. The frequency spectrum of the non-speech beacon sound. We used a drum sample with a broad frequency range as it fits the mental model of a sound that emanates from one specific location.

Technical Setup

To compare the behavior of real spatial hearing with the orientation in a virtual audio space, we created the following experimental setup. We placed 24 Wavemaster Mobi loudspeakers at 15° intervals in a circle with 4 m diameter. The loudspeakers are designed to stand upwards and emit sound in an omnidirectional pattern, which is how sound sources are modeled in our virtual space. We placed the loudspeakers at 140 cm height above ground to reduce the impact of elevation angle on the localization. The audio output of the smartphone was connected to the loudspeakers via a cable hanging from the ceiling. The cable was attached to the user's waist to avoid pulling forces on the device and to keep users from stumbling over it.

To quantify the influence of smartphone-based spatial audio rendering, we created a virtual audio space that represented this same setup. Positional tracking was performed with a Vicon optical tracking system with an update rate of 100 Hz. Spatial rendering was done on an Apple iPhone 4S running iOS 5.1.1 using the OpenAL library and presented using AKG K-512 headphones. The headphones fit firmly and have a supple cable so as to reduce the impact on the amount of head turning. Since we needed the optical tracking markers for the head in both conditions, participants had to wear a headband during the loudspeaker trials, which balances this influencing factor. While state of the art spatial audio rendering technology achieves astonishing results[1], the auralization results of this framework are less realistic. We decided to use this one, as it is a representative for a variety of spatial audio rendering frameworks available for mobile phones, comparable to, e.g., the AM3D Framework[2] used in [10] or the Java Advanced Multimedia Supplements used in [19]. We used the OpenAL extension `ALC_EXT_MAC_OSX` which provides a better spatialization based on a spherical head model and including the following filter factors: interaural level difference, iteraural time difference, head filtering, and frequency dependent distance filtering. To improve the perception of sources that are behind the head we used the `ALC_EXT_ASA` extension that enables additional effects, such as reverb, obstruction and occlusion. As the rendering suffers from front-back confusion, which is a common problem in spatial audio rendering [13], we added a low-pass filter that muffles the sounds that are behind the listener.

The low-pass filter intensity is interpolated linearly between 0 dB and 36 dB for sources with an azimuth angle between 90° and 180° (Figure 4). For the reverb, we used the medium room preset which best matched our impression of the physical room's characteristics. We also tuned the audio rendering parameters to make the scene sound as similar as possible in both conditions.

No delay of location or orientation measurement was perceived. With a specified latency of 2.5 ms of the Vicon Tracker and an average round trip time for the WiFi connection of 4.7 ms, we are below the limit of 376 ms total system latency defined in [11] and the 80 ms head tracker latency defined in [4].

Conditions & Methodology

Since related research indicates that there are performance differences in the localization of different source sound types [18, 22], we decided to use a non-speech sound and a speech sample. As a non-speech sound, we chose a drum sample that covers a large frequency range (Figure 5) and is repeated every second. The repetition rate was chosen based on the recommendations in [18]. We favored the drum sound over artificial sounds, e.g., noise or a square waveform, because it fits the mental model of a sound emerging from a precise single location. The speech sample is a continuous monologue of a male voice, which is close to our use case. Together with the two presentation modalities headphone and speaker, we have four conditions that were balanced across all participants. Our 24 participants, 3 female, age 19-53 (average 26), mostly had no prior experience with spatial audio and did not report any known hearing problems. The position of the sound source was randomized, such that each participant had to navigate to each of the 24 sources under every condition.

We recorded the position of the head as well as head, body, and device orientation. Due to a technical issue, headphone measurements for source no. 7 were discarded, leaving 23

[1] Virtual Barber Shop: **`http://youtu.be/IUDTlvagjJA`**
[2] **`http://www.am3d.com`**

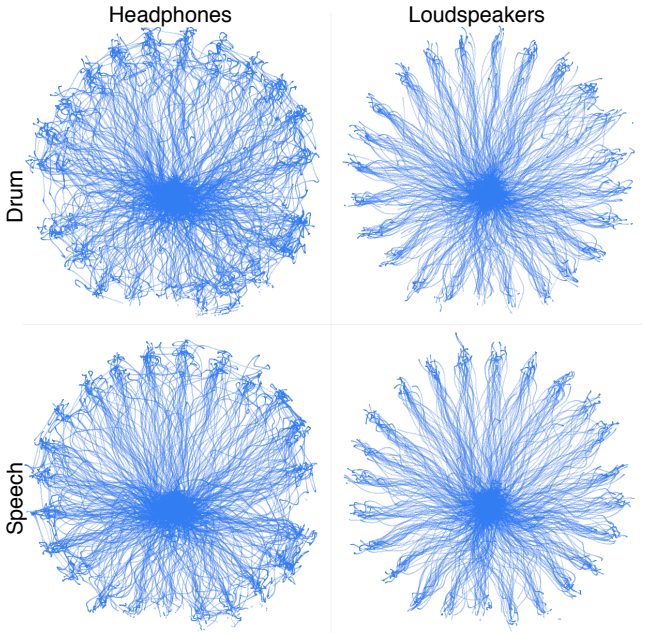

Figure 6. Paths on the way from the start in the center to the sources on the periphery of the circle. We see that the paths in the speaker condition are headed more directly to the source than with virtual audio rendering.

sources for the evaluation. We did not measure task completion time, since this is highly dependent on the source position (you have to turn around to reach a source in your back), and it is dependent on the type of beacon sound as a pulsed signal like the drum only allows localization in bursts in contrast to a continuous signal.

To be close to our designated use-case in the CORONA audio space, which might be similar to implementations that use the display to provide additional information, participants had to hold a smartphone in their hand. Participants were instructed to start the task using a button on the smartphone, go to the sound source currently playing until it was directly in front of them, and end the task using the stop button on the device. Participants practiced all conditions in a 12 trial training session before the actual experiment.

Results

By looking at the recorded paths of the participants, we can already see that the rendering has an effect, as the paths in the speaker condition are much smoother and lead towards the target more directly (Figure 6). From the three orientation measurements head, body, and device, we calculated their relative angles. Following the definition in [11], we define head-yaw (θ_h) as the relative angle of the head to the body, device-yaw (θ_d) as the relative angle between device and body, and head-device-yaw (θ_{hd}) as the relative angle between head and device. We transformed the values from their reported range of $0°$ to $360°$ to $[-180, ..., 180]°$, with $0°$ being the direction of the user's torso. We subtracted the initial difference between head, body, and device measurement at the beginning of each trial, since this difference is caused by the placement of the tracking markers.

Most of the time, body and head are aligned, as the means for θ_H are close to 0 for both conditions (Headphones: $M = -1.57°, SD = 15.83$, Speaker: $M = -2.24°, SD = 19.98$). The kurtosis[3] for the headphone condition is a bit higher (Headphones: Kurtosis $= 3.08$, Loudspeaker: Kurtosis $= 2.21$), which indicates that the participants turned their head less in the headphone condition. Although the headphone we used has a comparably long and flexible cable, we cannot totally exclude this as an influencing factor on the amount and range of head rotations. Overall, body and device are aligned most of the time, since we have $M = -0.18°, SD = 8.62°$, Kurtosis $= 85.7$ with headphones and $M = -0.35°, SD = 13.05°$, Kurtosis $= 5.92$ for loudspeakers.

Since positive and negative angles cancel each other out when calculating the arithmetic mean, we calculated the root mean square (RMS) head-yaw and device-yaw deviation ($\theta_{h(\text{RMS})}$ and $\theta_{d(\text{RMS})}$), which gives us the average amount of head and device turns. After performing a log-transform on the RMS data, a repeated measures ANOVA revealed a major effect of the used rendering (headphones or speaker) on $\theta_{h(\text{RMS})}$ ($F(1, 2083) = 132.76, p < .0001$) However, if we take a closer look at the data, we see that the RMS means only differ by about $4°$ (Headphones: $M_{\text{RMS}} = 13.86°, SD = 8.05$, Loudspeakers: $M_{\text{RMS}} = 17.75°, SD = 9.93$), which places it in the order of the just noticeable difference of the rendering. This slight difference is also noticeable in the head-yaw (θ_h) distribution. The RMS means for the angle between head and device ($\theta_{hd(\text{RMS})}$) are in the same range as for those between head and body (Headphones: $M_{\text{RMS}} = 15.06°, SD = 9.19$, Loudspeakers: $M_{\text{RMS}} = 19.73°, SD = 11.84$), which shows that body and device orientation can be considered equal in this setting.

Not surprisingly, the source position also has a major effect on $\theta_{h(\text{RMS})}$ ($F(23, 2083) = 22.63, p < .0001$). When orienting towards a source in your back, the amount of head turns will of course be larger. If we look at the values for the individual sources however, we cannot attribute this effect to sources in a specific location.

Discussion

The head-yaw tracks over time look very similar for both rendering conditions (Figure 7). The high fluctuations at the end of the tasks are caused by the fact that two participants took exceptionally long and turned their head extensively to discriminate between two possible candidates. The overall observation is that after a larger initial head-turn to get an orientation, the head-yaw stays within a $10°$ angle to both sides. If we just look at $\theta_{h(\text{RMS})}$, the mean value of $14°$ is not extensively large, taking into account that our rendering has a just noticeable difference of about $4°$. Similar to Fitts' law tasks, we have a large movement at the beginning which is then slowed down to achieve a precise homing. By using body or device orientation, we risk loosing the large head-turns we see at the beginning of each trial. This might lead

[3] *Kurtosis* is a statistical measure that describes the distribution of data around the mean. A positive or high kurtosis characterizes a sharp, peaked distribution.

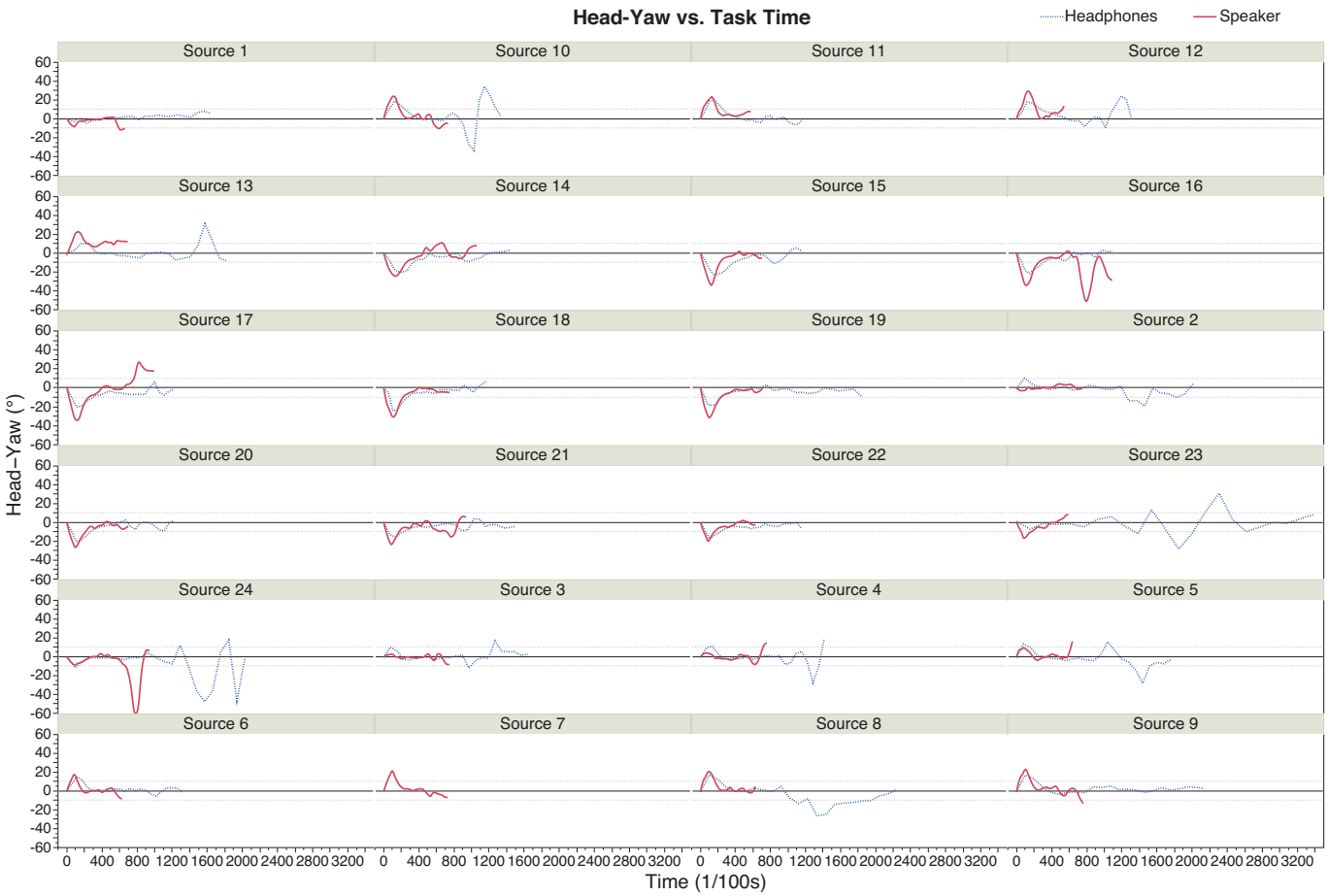

Figure 7. Mean head yaw per source. Sources are numbered from 1 to 24 clockwise, starting at 12 o'clock. The large deviations at the end come from two users that took exceptionally long and turned their head extensively at the end of the task.

to a seriously degraded sense of presence in the virtual space as these rotations are necessary to get the initial orientation. The mean duration of the peaks exceeding 15° occurring in the first 4 s of each trial (cf. Figure 8) is 590 ms in the headphone conditions (SD = 750). This could be considered as an additional head-tracker latency, as the body follows the head with some delay. These 600 ms are too large to stay unnoticed by the user, but completing navigational tasks is still possible [11]. Depending on the rendering resolution, technical setup, and designated use case this might be tolerable.

In our setting the device did not show any relevant information, but was used to start and end the trial. Nevertheless, all users held it in their hand in front of the body. If the smartphone is used to display some additional information, holding it in this position is further encouraged, which means that body and device will be aligned most of the time. Using the device orientation could also allow for different interactions with the audio space, such as using it as a virtual directional microphone.

ORIENTATION MEASUREMENT AND PRESENCE

Our first experiment showed that to locate and move towards sound sources in the near field, the initial head turn is a natural behavior, which is not severely influenced by our rendering algorithm. This initial head turn, however, makes it difficult to use other sources than head tracking for orientation measurement. The tracking speed and accuracy necessary to measure and analyze this behavior is only achievable in a lab setting. As many of the existing implementations are deployed in much larger areas [16, 19, 8, 12], the question remains if the use of a different device orientation in a larger setting has an influence on the perceived presence and navigation performance. As indicated in [19], users might move slower, enjoy the experience, and pay less attention to the realism of the installation. To draw the right conclusions for practical installations from the results of the first study, we conducted a second experiment in a real-world setting.

Optical tracking is unfeasible for such a scenario, as it requires some kind of marker to be placed on the headphones and a considerable amount of cameras to cover large areas. GPS and magnetometers are less precise and may introduce higher latency and larger error to the measurements fed into the rendering algorithm, which might have an influence on the perceived presence.

To account for these different types of installations, we conducted a second experiment using sensors appropriate for larger implementations.

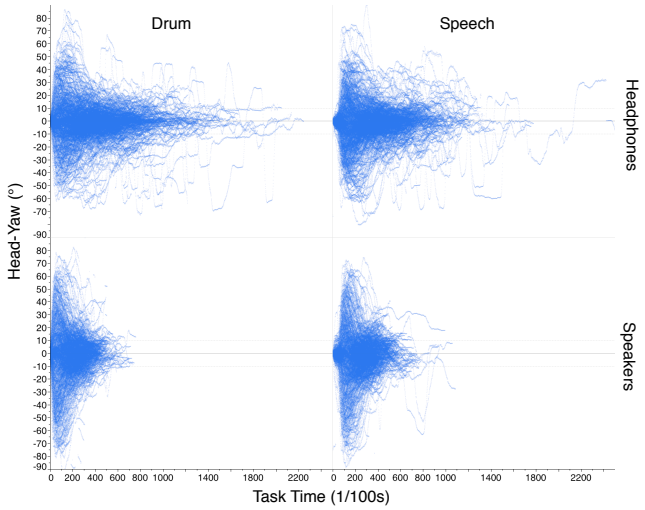

Figure 8. Head-yaw over task time of all users in the four different conditions. The initial head turn is present in both conditions.

Technical Setup

Since optical tracking is not feasible for large areas such as the Coronation Hall (45×20 m), we used a Ubisense RTLS[4] with an accuracy of 15 cm in the center of the covered area, 50 cm at the outer borders, and a refresh rate of approximately 10 Hz. The location measurement has a specified latency of 234 ms and the WiFi connection used to transmit the location data to the smartphone adds an additional average latency of 42 ms, which results in a total location update latency of around 276 ms. The orientation measurement was performed using a tilt-compensated compass HMC6343 with a refresh rate of 10 Hz. Audio rendering was performed with the same engine as in the previous experiment.

Conditions & Methodology

We varied the placement of the compass, which was attached either to the middle of the headstrap of the headphones, to the left shoulder, or to the smartphone. To create comparable measurements, we used the same chip for all three measurements even though the smartphone had a built-in compass. Participants were not told which sensor placement was actively used in the respective trial.

We created a series of six audio samples enumerating specific classes of objects, i.e., colors, first names, drinks, fruits, animals and cities, using a text-to-speech system. The participants were instructed to walk to the source shown on the smartphone display using text and an image. Upon successful arrival at a source, i.e., entering the capture radius [11, 22] of 2 m, a short sound sample notified the user. We created three distinct paths that were randomly assigned to the conditions. We randomized the sound sample played at a specific source position on that path and balanced the order of the compass placement across participants. The measurements recorded during the experiment include the path taken, the orientation of the compass, and the orientation of the smartphone compass. After each trial, we asked the participants to fill out the presence questionnaire proposed (Figure 9) in [23], omitting the questions only related to vision or touch. Participants had to walk around through the audio space to get acquainted to it before the first trial.

Results

We collected data from 9 users, 2 female, age 20-25 (average 24), who all successfully completed the tasks. All questions were answered on a 7 point Likert scale, with 1 being the lowest and 7 the highest score. An analysis of the questionnaires showed no substantial difference between the three different compass placements. Head tracking receives the best overall scores ($M = 5.3, SD = 0.8$), but the difference to device tracking is very small ($M = 4.9, SD = 0.5$) (cf. Figure 9). Since the perceived presence questionnaire [23] is quite long we will only report the most interesting results. For the question *How natural did your interaction with Corona seem?*, the head compass got the best results with an average score of 5.9 ($SD = 0.8$), followed by device tracking ($M = 5.44, SD = 1.1$), and body ($M = 5.0, SD = 1.3$). A pairwise Tukey-HSD test showed no significant difference with the smallest $p = 0.27$. The responsiveness of the environment was rated on a similarly high level: head: $M = 5.8, SD = 0.89$; device: $M = 5.3, SD = 1.4$; body: $M = 5.2, SD = 1.2$. The stability of the sources in space was perceived better in the head ($M = 6.1, SD = 1.0$) and device ($M = 5.3, SD = 1.2$) conditions than with body orientation ($M = 4.4, SD = 1.9$). For this question, the difference between head and body tracking is marginally significant with $p = 0.06$. The participants adjusted quickly to the virtual environment experience, again, with a slight but not significant advantage for head tracking. Some users mentioned that, although they were able to complete the task, they felt confused by the body tracking. The ratings indicate that the perception of the virtual audio space is not heavily affected by the different orientation measurements. This supports our hypothesis that head tracking is best, but device tracking sufficient for certain applications.

From the log files, we calculated the relative angle between head and device, which we know from the first experiment to be a good approximation for body orientation. Since the hardware changed, the results are not directly comparable to those gained in the first experiment. As the compass chip uses accelerometer data to compensate tilt, the stability of the reading is reduced while walking. The average θ_H ($M = -8.4°, SD = 33.0$) and $\theta_{H(\text{RMS})} = 34.1°$ are around double the results from the first experiment, which can be partly explained by the high fluctuation while walking. Future developments should take different filtering approaches into consideration and measure their influence on the overall latency. The task completion times for the three conditions showed a distribution similar to the ratings from the questionnaire. Head tracking was fastest with $M = 192\ s, SD = 63$, followed by device tracking with $M = 198\ s, SD = 62$, whereas body tracking was considerably slower with $M = 245\ s, SD = 106$.

[4] `http://www.ubisense.net`

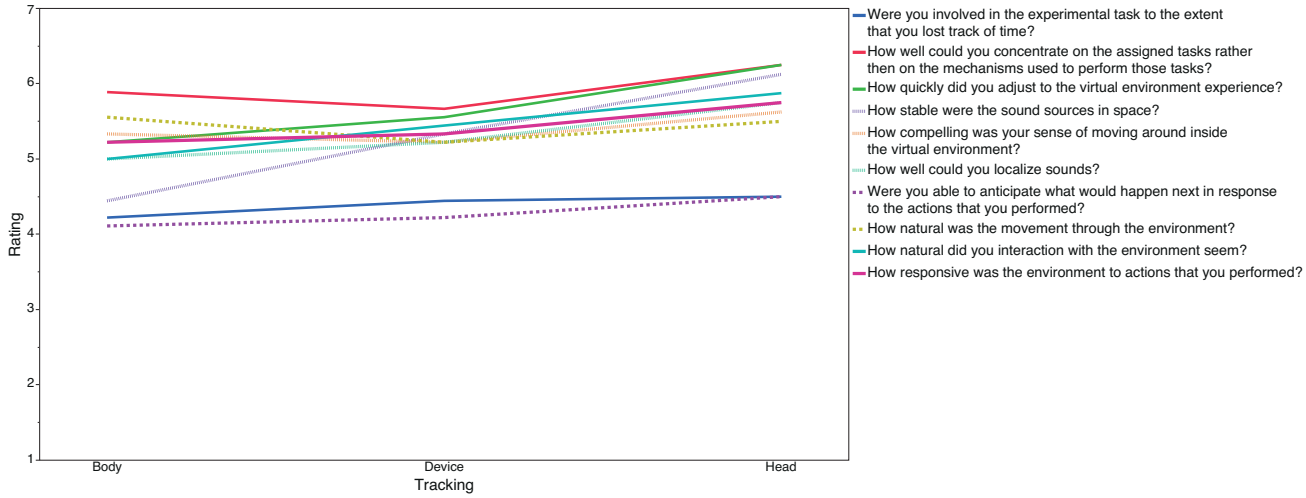

Figure 9. Mean ratings for the different compass placements on the presence questionnaire. Overall ratings are fairly high with a slight, although not significant advantage for the head tracking.

CONCLUSION

The human brain is really good at covering up errors in the audio simulation. When physical artifacts are augmented with virtual sound, we can observe the "ventriloquist effect" [1]: smaller errors in the combination of tracking and rendering are simply ignored, and the sound source "snaps" to the object. When no physical anchor is present, the source position only needs to be perceived as stable, as exact positioning is not required. Even a total failure in the tracking system can be interpreted into something meaningful. In one case we encountered problems with the transmission of location data from the server to the client, so only orientation was updated. The user of the affected device commented this with *"That was amazing! After some time, the voices started walking with me!"*

As a conclusion from this work, we recommend to use head tracking if realism and navigation close to the virtual sound sources is important, e.g., when exploring small artifacts in a museum. If the sources are close to each other, this becomes even more important, as the differentiation between two potential sources will result in additional head turns. However, in our experiment, the differences in the ratings between the three orientation measurements are very small and not statistically significant. Considering the small number of participants, this is not surprising, but we expect this not to change with a larger sample. If the focus of the implementation is rather on serendipitous discovery, the sources are further apart, or the use of additional hardware poses a problem, using the available sensors of a smartphone may be sufficient. Even in the case of nearby sources as in the first experiment, we think it is more a matter of communicating the functionality. Using your device as a virtual directional microphone might not give you the sense of total realism. Nevertheless, it could trigger your sense of exploration and motivate to point the device towards different artifacts to listen to their sonification.

Looking at the use of audio augmented reality as a navigational tool, the dimensions of the audio space increase dramatically, e.g., for city wide navigation. In such a scenario, the sources would probably become larger, blurring the error of the measurement. Since audio-based navigation is mostly designed to keep the hands and eyes free, placing the smartphone into a shirt pocket is probably good enough.

All these applications have to be tested, of course, but we are confident that reducing the technological effort required can bring audio augmented reality to a larger audience.

SUMMARY AND FUTURE WORK

In this paper, we looked at the natural behavior when moving towards a sound source with the goal to find out where the compass should be placed when implementing an audio augmented reality application. To analyze the orientation in a virtual audio space and to quantify the impact of our simple spatial audio rendering, we conducted a first experiment where participants had to walk to a sound source. The general observation is that after a large, initial head turn, the body follows and the head movement stays within a fairly small angle. We found that the rather simple audio rendering we used does not substantially change this behavior, although the amount of head turns is slightly smaller for the virtual condition. The delays and offsets introduced by the use of the device compass would probably be noticeable, although not necessarily prevent successful navigation in the audio space.

A series of audio augmented reality implementations [14, 16, 19] cover a much larger area than our first setup. Due to technical limitations, the tracking on this scale is less precise than the one we used in our first experiment. To account for these installations, we ran a second study, comparing three different compass placements (head, shoulder, device) and their impact on the perceived presence in the virtual environment. The head tracking received slightly, but not significantly better ratings, followed by device and body tracking.

From the data we collected during these experiments, we recommend placing the compass on the head in implementations that focus on natural behavior and quick movement to a closely located sound source. If the audio space is large or the focus is more on serendipitous discovery, using device tracking is potentially sufficient to create an immersive user experience. The reduced technical requirements allow for the distribution of software-only solutions that can spread audio augmented reality applications to a broad public.

Future work should take a closer look at the impact of using the device compass as orientation source on the behavior when interacting with nearby sources. It is possible that people easily adapt to the restriction, or even develop new interaction methods, for example, use the device as a virtual directional microphone. In our studies, we used a rather simple rendering algorithm, similar to those available in smartphone SDKs or as commercial frameworks. It uses the inter-aural level difference (ILD) as principal cue for localization, which is the simplest to implement and very robust as most people understand it easily. When we look at high-end audio rendering, possibly with HRTFs individual or adapted to the user [24], the ILD becomes a less important source of information [13]. With this kind of rendering, gauging the source position using head turns is possibly not necessary. The effect of low-resolution location measurement, e.g., GPS on high-end rendering algorithms is an interesting area as well.

ACKNOWLEDGMENTS

This work was financed by the German B-IT Foundation and the state of Northrine Westphalia through its EU-ERDF program “Ziel 2”.

REFERENCES

1. Alais, D., and Burr, D. The ventriloquist effect results from near-optimal bimodal integration. *Current biology 14*, 3 (2004), 257–262.
2. Ankolekar, A., Sandholm, T., and Yu, L. Play it by ear: a case for serendipitous discovery of places with musicons. In *Proc. CHI ’13*, ACM (2013), 2959–2968.
3. Blauert, J. *Spatial Hearing: Psychophysics of Human Sound Localization*, 2nd ed. MIT Press, 1996.
4. Brungart, D. S., Simpson, B. D., and Kordik, A. J. The detectability of headtracker latency in virtual audio displays. In *Proc. ICAD ’05* (2005), 37–42.
5. Holland, S., Morse, D. R., and Gedenryd, H. AudioGPS: Spatial Audio Navigation with a Minimal Attention Interface. *Personal and Ubiquitous computing 6*, 4 (Jan. 2002).
6. Kajastila, R., and Lokki, T. Eyes-free methods for accessing large auditory menus. In *Proc. ICAD ’10* (2010), 223–230.
7. Loomis, J. M., Hebert, C., and Cicinelli, J. G. Active localization of virtual sounds. *The Journal of the Acoustical Society of America 88* (1990), 1757.
8. Lyons, K., Gandy, M., and Starner, T. Guided by voices: An audio augmented reality system. In *Proc. ICAD ’00* (2000).
9. Mackay, W. E. Augmented reality: linking real and virtual worlds: a new paradigm for interacting with computers. In *Proc. AVI ’98*, ACM (1998), 13–21.
10. Marentakis, G. N., and Brewster, S. A. Effects of feedback, mobility and index of difficulty on deictic spatial audio target acquisition in the horizontal plane. In *Proc. CHI ’06*, ACM (2006), 359–368.
11. Mariette, N. Navigation performance effects of render method and head-turn latency in mobile audio augmented reality. In *Proc. ICAD ’09*. Springer, Copenhagen, 2010, 239–265.
12. McGookin, D., Brewster, S., and Priego, P. Audio Bubbles: Employing Non-speech Audio to Support Tourist Wayfinding. In *Proc. HAID ’09*. Springer, Dresden, Germany, 2009, 41–50.
13. Middlebrooks, J. C., and Green, D. M. Sound localization by human listeners. *Annual review of psychology 42*, 1 (1991), 135–159.
14. Reid, J., Hull, R., Cater, K., and Fleuriot, C. Magic moments in situated mediascapes. In *Proc. ACE ’05*, ACM (2005), 290–293.
15. Sander, C., Wefers, F., and Leckschat, D. Scalable Binaural Synthesis on Mobile Devices. In *AES Convention 133* (Oct. 2012).
16. Stahl, C. The roaring navigator: a group guide for the zoo with shared auditory landmark display. In *Proc. MobileHCI ’07*, ACM (2007).
17. Terrenghi, L., and Zimmermann, A. Tailored audio augmented environments for museums. In *Proc. IUI ’04*, ACM (2004).
18. Tran, T. V., Letowski, T., and Abouchacra, K. S. Evaluation of acoustic beacon characteristics for navigation tasks. *Ergonomics 43*, 6 (2000), 807–827.
19. Vazquez-Alvarez, Y., Oakley, I., and Brewster, S. Auditory display design for exploration in mobile audio-augmented reality. *Personal and Ubiquitous computing 16*, 8 (2012), 987–999.
20. Vorländer, M. *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*, 1st ed. Springer, 2007.
21. Wakkary, R., and Hatala, M. ec(h)o: situated play in a tangible and audio museum guide. In *Proc. DIS ’06*, ACM (2006).
22. Walker, B. N., and Lindsay, J. Navigation Performance With a Virtual Auditory Display: Effects of Beacon Sound, Capture Radius, and Practice. *Human Factors 48*, 2 (2006), 265–278.
23. Witmer, B. G., and Singer, M. J. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoper. Virtual Environ. 7*, 3 (1998), 225–240.
24. Zotkin, D. N., Duraiswami, R., and Davis, L. Rendering localized spatial audio in a virtual auditory space. *IEEE Transactions on Multimedia 6*, 4 (2004), 553–564.