Master Project
Interactive Media Design Master Program
Södertörn Högskola

Supervisor: Minna Räsänen
External: Annika Waern (MobileLife Centre)
Student: Syed Naseh

Title:
Mobile SoundAR: Your Phone on Your Head
Table of Contents

Abstract........................................................................................................................................ 3
Keywords......................................................................................................................................... 3
ACM Classification Keywords........................................................................................................ 3
General Terms.................................................................................................................................. 3
Introduction...................................................................................................................................... 4
Immersion issues without head-tracking..................................................................................... 5
  Front-back confusion..................................................................................................................... 5
  Interaural Time Difference ITD .................................................................................................... 5
  Interaural Intensity Difference IID .............................................................................................. 6
Design Problem............................................................................................................................... 7
Design Methodology........................................................................................................................ 7
Previous work.................................................................................................................................. 7
Capacity of modern phones.......................................................................................................... 7
Prototyping....................................................................................................................................... 8
Qualitative User-studies.................................................................................................................. 9
  Conclusion.................................................................................................................................. 9
Quantitative User-studies.............................................................................................................. 10
  Fitts Law.................................................................................................................................... 10
Method Description for Quantitative User-studies (Fitts Law)...................................................... 11
  Subjects:..................................................................................................................................... 11
  Equipment/ Technical Design......................................................................................................... 11
  Procedure.................................................................................................................................... 11
  Extra Description......................................................................................................................... 11
  Pilot.............................................................................................................................................. 12
  Setup.......................................................................................................................................... 12
  Participants................................................................................................................................. 12
  Problems..................................................................................................................................... 12
  Points.......................................................................................................................................... 12
  Results......................................................................................................................................... 12
  Conclusion.................................................................................................................................. 12
Conclusion and future work........................................................................................................... 13
Uses................................................................................................................................................ 13
  Museum....................................................................................................................................... 13
  TeleConferencing....................................................................................................................... 13
  Music Room............................................................................................................................... 13
Acknowledgements......................................................................................................................... 13
References....................................................................................................................................... 14
Abstract

Sound localization plays an important role in providing a believable sound based augmented reality. Human auditory system uses several cues for sound localization and thus it is important to render these cues in virtual environment as well. Though all cues complement each other, head motion is one cue that can work individually to help locate the direction of sound source. Affixing sensors on the head of the user have been previously tried to reintroduce head motion in virtual soundscape. Modern smart phones with motion detecting sensors are becoming highly pervasive in todays society. Such smart phones open up possibilities for early prototyping and testing of ideas, that previously required high fi gadgetry. Wearing the phone on head can track the head movement using gyroscope and accelerometer. This paper discusses development of prototypes to provide head tracking using iPhone4. A Fitts' Law study is formulated to learn about the properties of audio for target acquisition.

Keywords

Augmented Reality, Sound Based AR, Smart Phones, Directional hearing, Interaction Design, Experience Design

ACM Classification Keywords

H5.1 Multimedia Information System; Artificial, augmented and virtual realities. H.5.5 Sound and Music Computing; Methodologies and Techniques.

General Terms

Design, Prototyping, 3D Sound, Fitts Law
Introduction

Unlike virtual reality which is totally immersive, augmented reality users are aware of their surroundings while having access to additional sensory input and information. Humans hear with less spatial precision than they see [3] therefore the illusion of augmented reality could be quite strong with sound. However sound localization is critical for such immersion.

Sound localization refers to a listener's ability to identify the location or origin of a detected sound. This lays the basis for augmented sound illusion. The listener should locate the virtual sound as coming from real environment instead of the headphones. In comparison to vision, the localization of sound works in all directions. Human ear can locate sounds in three different planes namely elevation, distance & azimuth.

Illustration 1: Azimuth & Elevation

Elevation is the angle of the sound source from the ground. Distance is the distance of sound source from the listener and azimuth refers to the angle sound source is placed from the reference direction.

Human auditory system uses several cues for sound localization. These cues work together in complement with each other in real life. To create an augmented reality scenario these cues that exist in real world should be rendered through the headphones as well.

As sound travels from the source to the listener, it is
- reflected and refracted off the human body (Static cues),
- reflected and refracted off the environment (Environmental cues) and
- effected by the head motion (Dynamic Cues).

This is not an exhaustive list of cues that the auditory system uses. It is very common in synthesized or recorded sounds that static cues are ambiguous [4]. This is because every person has a different body size and therefore making sound perception very subjective. Normally when listening through headphones the sound source rotates with the listener if the listener rotates his head and is therefore perceived as being inside the head. If the sound source remains locked at a static position with respect to the movement of the head, it is instinctively thought of as belonging to the real environment and not coming through the headphones. This dynamic cue provided by head motion can help locate the azimuth of the sound source. Turning of face towards the sound source is a natural instinct with aim to bring the sound source in to the visual field [1]. In our prototype, head motion is reintroduced in the virtual sound scape by tracking the movement of the head using the sensors of an iPhone 4. We try to render sound so that they appear to come from some point in space in real world.
Immersion issues without head-tracking

Two channel sound is the standard format for home stereo receivers, television and radio. The simplest two-channel recordings, binaural recordings, are produced with two microphones set up to take place of human's two ears. When you listen to these two channels the setup recreates the experience of being present where the recording took place.

Front-back confusion

Head Related Transfer Functions (HRTF) are mathematical transformations that can be applied to mono sound signals and the resulting left and right signals are the same signals that the listener perceives when listening to a sound source positioned in real-life 3D space. HRTFs are used to recreate the static cues of sound being diffracted and reflected off the torso, head and pinna before reaching the ear drum. But because every individual has a different body and ear structure, individual HRTFs are required for a perfect scenario. Generalized HRTFs are therefore used.

Generalized HRTFs leads to two common confusions when simulating 3D sounds using earphones; front & back confusion and confusion pertaining to elevation.

A front-back confusion occurs when the listener locates the sound as being in the rear hemisphere when the actual source is in front[6]. So if the source is positioned at an azimuth of 30 degrees (which is in the front) the listener can mistake this as source located at an azimuth of 150 degrees (located behind the listener). Wenzel et al. in [5] has provided a confusion rate as high as 50%. The turning of head can eliminate front & back confusion. Only a small amount of head movement (5° or more) while a sound is occurring is sufficient to eliminate the front & back confusion. [Perrett & Noble, 1997]

![Illustration 2: Front & Back Confusion](image)

Interaural Time Difference ITD

The distance between the two ears of a listener is approximately 0.18 m. This is enough to perceive a time delay. The sound reaches the closest ear first and the listener can distinguish the direction. If the sound source is placed at a distance equally between the ears the time difference is about the same for the two ears [12]. Without head tracking the ITD remains the same for the sound through headphones and hence the sound moves with the user when the user moves his head.
**Interaural Intensity Difference IID**

If the sound source is placed at the left of the person, the head will hide the right ear. The sound intensity is thereby higher at the closest ear [12]. Its intensity will be less at the right ear than the left, both as a result of the screening effect of the head and, to a lesser extent, due to the extra distance travelled. This is referred to as the inter-aural intensity difference. The head interferes with the sound-wave, casting the auditory equivalent of a shadow on the far ear. And, this sound shadow is more effective for sounds of higher frequency.

*Illustration 3: Sound Shadow due to IID*
Design Problem

Knowing that movement of head can help with sound localization and sound localization is important for a believable augmentation, the design problem is how to use a modern day mobile phone to imitate the head movement in the virtual soundscape. The solution would minimize the front-back confusion and give the illusion that the sounds through headphones are actually part of real environment rather than the virtual.

Design Methodology

Design Oriented Research Methodology [2] is used where the knowledge gained by creating the prototypes is more important than the end product. The resulting artifact will be considered more as a means to gain new insights in sound based augmented reality systems. Rapid Application Development (RAD) is used to favor rapid prototyping. The prototype has been iteratively developed and has currently gone through two iterations. Both qualitative and quantitative analysis is planned to test the prototypes before feeding the results into the next iteration.

Previous work

Affixing sensors on the head of the user have been used previously to reintroduce head motion in virtual soundscape. Goudeseune et al. In [3] constructed a wearable outdoor computer system using off the shelf hardware and publicly available software. They presented sound and images placed at particular locations by means of headphones and a head-mounted display. The sounds and images were tied to GPS coordinates. The hardware used was a GPS receiver, flux-gate compass, tilt sensor and gyroscope. The gyroscope used was actually a gyroscope enabled computer mouse that works in mid air by pointing in different directions. The gyroscope, tilt sensor and compass data was combined together to track the head movement and provide spatial sound.

G. Healy et al. in [8] used generic set of wireless headphones with sensors connected on it to track the head movement and orientation. They used digital compass, ultrasound range detector and an accelerometer. A XBEE RF Module was used to wirelessly stream the sensor data back to the base station. They used UbiSense 2.0 3D positional tracking system to track the location of the subject within the experimental area. According to the authors the system provided accuracy of approx 15cm in x – y – z dimensions. This is quite good accuracy compared to current off the shelf GPS solutions that have their accuracy in meters. On the base station they used FMOD 3D audio production library in combination with a HRTF assistive sound card. The generated audio was streamed back to the user.

Capacity of modern phones

Rapid miniaturization of electronics has made it possible to fit an accelerometer and a gyroscope in an iPhone 4. The technology is now ripe and can be used in many ways which couldn't have been possible before. Using the device's motion data, now it is possible to find out the orientation of the phone.
Iphone 4 comes equipped with a microscopic electronic version of a vibrational MEMS gyroscope and a 3-axis MEMS accelerometer. MEMS is an abbreviation for microelectromechanical system. The Core Motion API for iOS4.2 provides the processed data from gyroscope and accelerometer using Apple's sensor fusion algorithms. This makes it easy to obtain the device's attitude at a point in time. “Attitude” refers to the orientation of the phone relative to a given frame of reference. The attitude is available as Euler angles (roll, pitch and yaw).

iOS4.2 supports OpenAL, which is used to playback positional audio. It is a free audio API designed for multichannel three dimensional positional audio. OpenAL converts mono sounds into stereo and pans them across depending on the position of the sound source and the listener in 3d plane. For current implementations only the azimuth of the sound source is taken into account ignoring the elevation of the sound source.

![Illustration 4: Euler Angles provided by Core Motion API](image)

**Prototyping**

Two interaction models were prototyped to recreate the head movement. One where the listener wears the phone on the head and second where the listener holds the phone in his hand and makes a sweeping motion, like a radar, to locate the sound source.

By wearing the phone on the head, it is possible to measure which way the person is looking, and keep track of head movement. When the iPhone is attached to the head of the user, the roll of the phone can be used to calculate the azimuth when user turns his head from left to right and vice versa. When held in the hand the yaw of the phone is used. In both the prototypes the given Euler angle (roll or yaw) is used to modify the audio and pan it appropriately with respect to the direction of the sound source.

An iPhone app is developed where sound source can be placed using the GUI around the user. Iphone's earphones are used. The data from the gyroscope is fed into 3D audio engine provided by Sonic Studios based on OpenAL. On receiving the data the engine generates a stereo signal of the panned sound to the listener. The listener can start and stop the sound using the phone's remote built in the headphones. The azimuth is updated at 60hertz.

A toggle switch is provided in the app to toggle between hand and head model. Another toggle switch is provided to reset and calibrate the Gyro.
The first iteration had a single sound of water dripping placed at an azimuth of 90 degrees, i.e. to the right of the listener. The sound was played continuously in a loop. The iPhone was attached to the front of a unisex winter cap which was later replaced by tying two iphone arm bands as shown in illustration 5.

The prototype was a positionally static application since at this early stage it was decided not to implement the global positioning of the sounds and the listener. Hence that user could only stand at one position and his movement within the experimental area has no effect on the sound. The height of the sound was kept static throughout as well and only the direction and distance of the sound could be manipulated in the application. Since the prototype was positionally static the movement of the listener had no effect on the distance of the sound.

**Qualitative User-studies**

The prototypes were informally presented at Mobile Future at Kista Science Tower on 11th November 2010. It was a noisy big hall which was considerably crowded. The visitors were asked to try to point out the direction the sound source using either the phone on their head or in their hand.

When using the phone like a radar, the visitors were aware of the presence of the phone in their hands at all times. The phone was acting as a means of a sound detector. But with the phone on the head they forgot the presence of the phone and were moving their head, and body, in a naturally instinct to locate the sound source.

Because the sound was placed to the right there was no front-back confusion involved and all those who wore the cap instantaneously pointed towards the sound. It was only after they moved their heads that they got the illusion as if the dripping sound was not coming through the headphones at all and was actually placed in the real environment. Few started walking towards the source to find it. The noisy crowded environment provided enough ambiance for the illusion to work.
Following the first iteration, multiple sounds were added in both the prototypes. A static scary ambient sound was added, the movement of the listener had no effect on the ambiance sound. A sound of a laughing witch and an owl was positioned at azimuth 30 and 270 respectively. The volume of owl was kept low enough so that it got merged with the ambience.

Second iteration was presented at the Open House at Mobile Life Centre on December 17th 2010. The demonstration was held in a darkened room measuring approx 20m². The visitors were asked to point out the direction of the witch. And were later asked to point out where the owl was.

**Conclusion**

Promising reactions were noted at Mobile Future especially where few participants started walking to find the sound. Where as at Open House a lot of participants were confused and couldn't localize the sounds. My hypothesizes is that augmentation is maximized at open noisy places and does not work that good in small rooms.

The participants were given an option to wear the phone either on their head or to use it in their hands. Those who wore it on the head were more comfortable to localize the sound as compared to those who used the device in their hands. A quantitative study is therefore planned to see which model suits best.

**Quantitative User-studies**

To learn more about the properties of audio for target acquisition a Fitts Law experiment is being formulated. The data from this quantitative study would be used to compare whether it is better to mount the sensor based phone on head or hand and to compare the differences in sound perception.

The study is carried out by the Mobile Life Centre at Stockholm University in collaboration with Helsinki Institute of Technology, as part of the EIT (European Institute of Innovation and Technology) initiative.

**Fitts Law**

In 1954 Paul Fitts published a paper about a mathematical model applied to human motor system[9]. Fitts' law predicts the time to reach a target if the width of the target and the distance to the target is known. The version of Fitts law coded in an ISO Standard [11] and commonly used for research purpose is:

\[ MT = a + b \log_2(A/W + 1) \]

Where MT denotes the movement time;
a, b are the regression coefficients;
A is the distance of movement from start to target center;
and W is the width of the target.

In [10] Marentakis et all. hypothesize that interaction in a spatial audio display is affected by the prominent variables of target width and distance to target in a manner similar to what is stated by Fitts’ law. Their empirical study shows that spatial audio target acquisition can be sufficiently described in terms of Fitts Law when proper width choices are made.

Below is the method description and the description of the pilot run.
Method Description for Quantitative User-studies (Fitts Law)

Subjects:
Training will be provided to subjects to use the device in both hand and head. They would be told how brain localizes the sound based on which ear is hearing what. They will be briefed about front and back confusion with sound localization and how moving the head is used to solve the error. They will be given 10 minutes to toy around with both head and hand technique and fill a small questionnaire.

Equipment/ Technical Design
- iPhone 4 would be used
  Vibrational MEMS gyroscope and a 3-axis MEMS accelerometer.
- Apple’s generic headphones for iPod/iPhone
- OpenAL to position sound. OpenAL converts mono sounds into stereo and pans them across depending on the position of the sound source and the listener in 3d plane
- Custom App would be used to randomly position sound and record following data
  - Time taken to acquire the sound
  - Distance moved (degrees)
  - Error in pointing (degrees)

Procedure
- The subject will either wear the device or hold it in hand.
- He will hear a sound of water dripping through his headphones. This would be a randomly placed sound.
- The subject would move as fast as possible to acquire the sound by pressing the mic button once. Acquiring the target means moving the head towards the direction of the sound. The direction of the sound is determined by the panning of the sound. Target is successfully acquired when sound in both the ears appear equal.
- Once the sound is acquired, either successfully or not, next sound would be played instantaneously. This sound would be randomly placed too.
- Subject can opt to stop any time by pressing the mic button twice.

Extra Description
- The direction of the sound source would be random. Distance, amplitude and other qualities would be kept the same for every sound played.
- Distance would be measured in degrees.
- Effective Width would be a point.

Illustration 6: iPhone headset with remote
Pilot
The pilot study was carried out in week 32. The basic aim of the study was to get an approximation of selection time.

Setup
12 points were randomly generated before hand. 3 points were distanced less then 30 degrees from each other and were ignored. Only 9 points were saved in the app. They were played in the saved order. Minimum distance between the points was 30 degrees where as the maximum distance was 120.

Participants
The study was performed with 3 students, male, with approx. age of 27 and no hearing problems (they were asked before hand if they are aware of any hearing impairments they might have). They were given 10-15 mins each to use the device before the study. Each participant did the test twice.

Problems
The app had a bug where the gyro was not calibrated and previous results were not flushed. This was realized when I analyzed the data of the first participant which was completely off the target because of the bug. The bug was fixed and the data collected via the buggy app was termed invalid.

Points
In total 36 Points for the head model and 27 Points for hand model were collected.

Results
Average selection time was 3.79 seconds where as the pointing was off target by an average of 8.35 degrees when the user wore the phone on his head.

Average selection time was 5.2 seconds where as the pointing was off target by an average of 8.67 degrees when the user used the phone in his hand making a radar like motion.

Conclusion
Although this is just a pilot run the head model has a better selection time with more or less the same error in pointing.
Conclusion and future work

The simple idea of putting the phone on the head works well for head tracking. The illusion is much stronger in large crowded noisy halls since it provides better augmentation compared to the darkened room. Head movement cues are important for human auditory system to determine the azimuth of the sound sources and by reintroducing this cue a believable sound augmentation is possible.

Head tracking with putting the phone on the forehead works for the augmented illusion but has limitations since the phone can not be used while it is being wore around. This project will further extend to track all the three axis of rotations (roll, yaw and pitch) of head movement to see how it effects the localization of the elevation of the sound source. Fitts Law study is planned in late September 2011.

Current smart phones are not just phones anymore. They provides means to quickly prototype and test ideas not only in the lab, but also outside the lab. I have noted few applications below where head tracking can be used. And without needing any sophisticated gadgetry these ideas can be tested using many of the current smart phones.

Uses

Museum
Head tracking enabled devices can provide museums with audio guides that can provided spatialized sound while breifing about certain artifact. These devices can also be used to help with directions within the museum permisies.

TeleConferencing
Head tracking enabled devices can be used in teleConferencing to position the participants just like they would have been if this conference was held in real life.

Music Room
Different music instruments can be placed around the user and the user can use his head orientation to focus on a certain instrument.

Acknowledgements
I would like to thank Minna Räsänen, Annika Waern and Martin Jonsson for their patience, support and supervision. I would like to acknowledge Kim Nevelsteen, Jon Back, Magnus Linell and Sergio for the valuable discussions at Mobile Life. Thanks to MobileLife Centre for their support.
References


SoundScape is an exploration on how mobile phones can be used in a sound based augmented reality scenario. The phone tracks the head movement in relation to the virtual sounds placed in the environment.

Contact:
Syed Naseh, syed.naseh@gmail.com
Annika Waern, annika@sics.se