Audio Engineering Society

Convention e-Brief 33

Presented at the 142nd Convention
2017 May 20–23, Berlin, Germany

Binaural spatialization methods for indoor navigation

Sylvain Ferrand[1], François Alouges[1], and Matthieu Aussal[1]

[1]*CMAP, Ecole polytechnique, CNRS, Université Paris-Saclay, 91128, Palaiseau, France.*

Correspondence should be addressed to Sylvain Ferrand (`sylvain.ferrand@polytechnique.edu`)

ABSTRACT

The visually impaired people are able to follow sound sources with a remarkable accuracy. They often use this ability to follow a guide in everyday activities or for practicing sports, like running or cycling. On the same principle, it is possible to guide people with spatialized sound. We have thus developed a navigation device to guide with sounds using binaural synthesis techniques. In this device, we are using both localization information provided by a precise and low latency positioning system and heading data computed from an Inertial Measurement Unit. These positioning data are feeding an HRTF based binaural engine, producing spatialized sound in real-time and guiding the user along a way. The user follows the sound, quite naturally and without initial training. Experiments show that it is possible to guide a walker with enough precision.

1 Introduction

Spatial audio guidance methods for navigation are using the natural ability of humans to localize directions of a sound. This capability results in the difference of signal received by each ear. These differences can be classified into three categories: the difference in arrival times between the ears (Interaural time differences, ITD), differences in sound level entering the ears (interaural level differences, ILD) and the effect of reflexions and diffractions from various parts of our head and shoulder, including pinnae, altering the sound asymmetrically. Head-related transfer functions (HRTF) are transfer functions that characterize how ears receive a sound from a point in space, they can be used to synthesize a binaural sound that seems to come from a specific location. It has been shown [1] that HRTF filtering allow a good accuracy for estimating the azimutal angle to a sound source. This precision can still be improved by taking into account the movement of the head [2] in real time.

If we are able to localize a sound, we are also able to follow a moving sound source. Visually impaired people are using this ability in everyday life, for example: to follow someone by listening to the voice or the foot paces. This ability is not specific to blind poeple, everybody is able to localize and follow a moving sound with a good accuracy and with a very minimal training.

2 Methods for audio guidance

Most studies on audio guidance focus on pedestrian navigation in the urban area, using beacon [5] or GPS-based route guidance [3, 5, 4]. With beacons, the user should head in the direction of beacons, choosing themselves the way to follow, while GPS based system performs turn-by-turn guidance changing sound source at every road intersection. A third approach consists of guiding people with a virtual sound source leading the user at a constant distance, like a *carrot on a stick*. The ARGUS project [6] used this method to guide individuals outdoor with an improved GPS system (GPS + IMU fusion). They successfully guide people on roads but the GPS does not permit to give an immediate feedback to the user because of it's low response time and limited precision.

In a different way, in this study, we are focusing on precision and performances for indoor or sports applications. Our goal is to guide people through 1 - 2 meter wide virtual corridors, typically to aid the visually impaired displacement indoor, in a warehouse or for practicing sports (for example, to permit a blind people to run on an athletic track).

In our experiments, the sound source can be static or on move in the experimental space. If the sound source is static, we are in the case similar to beacon guidance but suitable for indoor applications, it is then possible to lead someone to a precise location in a room. If the sound source is on move, we choose a *carrot on a stick* approach, and we can, therefore, guide someone along a predefined path with a good precision.

3 Device design

We designed an handled device which associates an accurate 2D positioning system (including a head tracker), a path tracking, and a binaural engine so as to simulate a virtual sound source regardless of user's movement, either a head rotation or displacement. Because head orientation and positioning have to be taken into account in real time in order to give an immediate feedback to the user, we pay particular attention to the development of a low latency and accurate positioning system.

The device is separated into two physical modules. The first module is placed on the user's head, it consists of the localization sensor and the head tracker attached to the audio headphone. The second module, the computational part, consists of a computer running the algorithm which implements the path tracking software and the binaural engine (tested on a laptop or ARM based embedded platform).

3.1 Software architecture

The software has been designed following two guidelines: high modularity and minimal latency operations. This is accomplished through a multi-tier architecture and each component is exchanging data via a REST [7] API. Figure 1 show the general software architecture of the system.

The main program is acting as a REST server and manages the communications with clients. It also includes the communication interface with sensors (head-tracker and positioning system) and the filtering system. All these tasks have been gathered in one multi-threaded process to minimize latency.

The binaural engine is a client which continuously gets localization data from the server through HTTP REST requests. The REST server is acting as an abstraction layer to sensors, in this way the binaural engine gets localization information regardless of localization technology (GPS, RTLS systems...).

We are using a second client for path tracking, visualization, and data recording. It may be noted that HTTP-based REST architecture could permit to locate this client on a remote computer. Typically, a handled embedded system could be used for hosting server part and binaural engine while a laptop with a wifi connexion could ensure the real-time visualization of data.

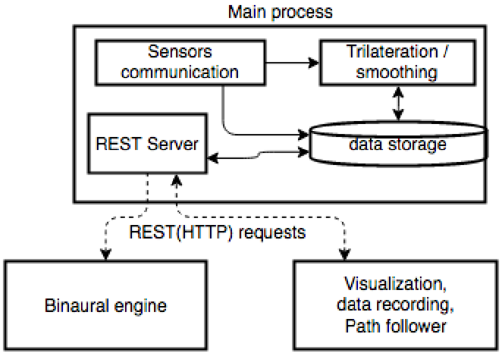

Fig. 1: Software architechture

Audio engine

We have to provide spatialized sound from a virtual source, which will give the user the direction to follow. For this purpose, we implemented an HRTF-based binaural engine [10, 11]. This software can spatialize a sound and allows real-time filter selection depending on the position of the user. For a better user experience, the software has been designed for low latency and imperceptible fast filter switching. We are using an HRTF data set derived from the subject 1040 of the *IRCAM listen database* [9]. To give accurate directions indications and lowering filters switching artifacts, HRTF measurements have been interpolated every 2° on a regular grid. For a more realistic rendering, the audio engine also takes into account the distance to the audio source and adapts the volume level accordingly.

3.2 Real Time Locating System (RTLS)

Traditional positioning systems like GPS typically provides signals with an accuracy in the range of 5-10 meters horizontal with an update rate in the order of the Hertz (see Table 1).

Technology	Positioning accuracy	Refresh rate
GPS	5 - 10 m	1 - 2 Hz
GPS + IMU	3 - 5 m	1 - 5 Hz
UWB	20 cm	10 - 15 Hz

Table 1: Comparison of positioning systems

In a sound guidance application, giving an immediate feedback to the moving user is absolutely crucial. For this reason, we are using a fast and accurate Real Time Locating System (RTLS) using the Ultra Wide Band (UWB) technology. UWB allows a fast and precise indoor or outdoor localization by utilizing the Time of Flight (ToF) [8] of a radio signal. The Two Way Ranging protocol (TWR) permits to measure the distance between two UWB transceivers, a master and a slave. In this protocol, we are measuring the round-trip time of a data frame sent from the master and replied by the slave. The elapsed time, between the device sending a message until a response is received by the master, permits to deduce a distance. This technology offers several advantages, it is robust in multi-path environments (relatively insensitive to shading of sensors and RF interference), low cost, low power and it gives high precision ranging (typical standard deviation about 10 cm) without complex clock synchronization infrastructure. The UWB transceivers modules we are using are based on DW1000 UWB chip from DecaWave.

In an area with several fixed UWB slaves (called Anchors) with known locations, trilateration algorithm can be used to determine the absolute position of the mobile master. A minimum of 3 anchors is required for 2D operation, 4 anchors in 3D. The master runs a loop and measures successively the distances to each anchor. With four anchors, our electronic device can achieve up to 12 cycles per second.

As an efficient framework for filtering and trilateration, we implemented a Kalman Filter.

Kalman filter design

Kalman filters are designed to estimate the state vector $x \in \mathbb{R}^{n_x}$ of discrete-time dynamic systems, which can be characterized by the relation :

$$x_k = Ax_{k-1} + Bu_{k-1} + w_{k-1} \tag{1}$$

$A \in \mathbb{R}^{n_x} \times \mathbb{R}^{n_x}$ is the state transition mathrice, $B \in \mathbb{R}^{n_x} \times \mathbb{R}^{n_u}$ is the input gain, $u \in \mathbb{R}^{n_u}$ is the input vector and $w \in \mathbb{R}^{n_x}$ the process noise

In our positioning problem, the input term is typically unknown, and x state vector is composed of X and Y components of position and speed x $= [x y \dot{x} \dot{y}]^T$

At time k a measurement z_k is made according to :

$$z_k = h(x_k) + v_k \tag{2}$$

$z \in \mathbb{R}^{n_z}$ represents the measurement vector, i.e. the distance measured by sensor $z = [dist_{a_1}\ dist_{a_2}\ \dots\ dist_{a_n}]^T$ and v_k the measurement noise (assumed to be gaussian) with $a_i = [a_{x,i}, a_{y,i}]$ the position of the anchor i

The measurement function $h(x_k)$ for a system with n anchors is given by the euclidian distance between each anchor and the tag :

$$h(x_k) = \begin{pmatrix} \sqrt{(x-a_{x,1})^2 + (y-a_{y,1})^2} \\ \sqrt{(x-a_{x,2})^2 + (y-a_{y,2})^2} \\ \vdots \\ \sqrt{(x-a_{x,n})^2 + (y-a_{y,n})^2} \end{pmatrix}. \tag{3}$$

Kalman filters are limited to linear dynamic systems but this measurement function is nonlinear. To address this issue, we are using an Unscented Kalman Filter (UKF) [12] . In most case, UKF is known to perform better than traditional Extended Kalman Filter based on Taylor series linearization. The UKF is adapted to deal with gaussian noise produced by UWB sensors and this nonlinear measurement function.

3.3 Head tracking

In order to guide the user to the appropriate direction, our application needs to know the orientation of the head of the user. For this purpose we are using a head tracker (Razor IMU [13]), giving the heading (i.e. the Euler's angles Yaw, Pitch, and Roll) of the user's head.

Heading and raw positioning data are multiplexed with a microcontroller and sent to the computer module of the device through a serial link. Data are retrieved by the interface thread and then processed by the filtering/trilateration thread which compute the positions of the user.

4 Preleminary results

In the following experiments, audio stimuli have been selected to be non-disturbing, pleasant to hear, with a rather continuous sound level and with a wide range of audible frequencies.

4.1 Following an unknown path

In this test, we were working in a 8 x 15 m area. We were defining a virtual course to follow in blind conditions. The virtual sound source moves dynamically along the path and lead the user at a certain distance ahead like a *carrot on a stick*.

The courses were chosen to be smooth, without sharp turn, and were changing at every try (typically we are using the sum of two sinusoids: $Asin(\omega_1 x+\phi_1)+Bsin(\omega_2 x+\phi_2)$, where A, B, ω_1, ω_2, ϕ_1 and ϕ_2 are randomly chosen). Figure 2 show a typical test. The line represents the virtual course and dots the position of the user.

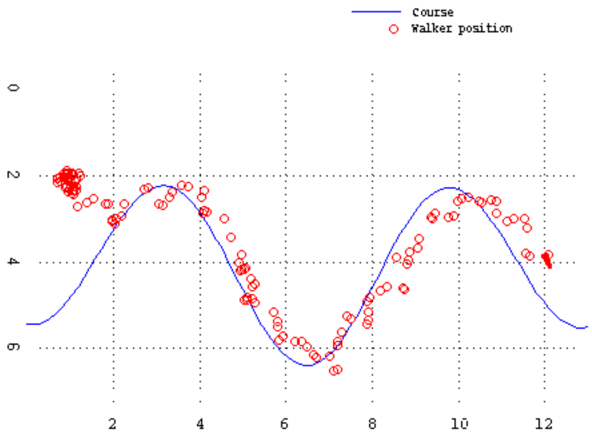

Fig. 2: Walking along an unknown path

This test has been done with five individuals including one person blind from birth. In blind conditions, with an unknown path, all of them followed the path without prior training. However, they were successful only at moderate walking speed (about 3-4 km/h).

4.2 A loop path

On the last test, users were limited by the length of the experimentation area. To allow then to pick up speed, we set up a wider loop circuit. In this experience, we were guiding walkers along an elliptic shaped course (approximatively 4 m and 6 m radiuses). Classical headphones or bones conductions headphone have been tested without a noticeable difference on guiding efficiency.

Figure 3 show the position of a blind person performing 3 laps after an initial 5 minutes training (with same path). This test has been performed with a bone conduction headset.

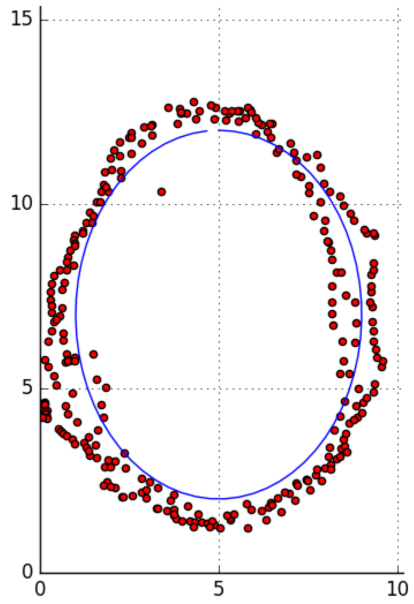

Fig. 3: Walking along a known path

In this example, the user went up to 7 km/h (typical speed of a brisk walking or a slow jogging) and followed the route with a sub-meter accuracy. At 7km/h, the mean distance to the ideal track was about 0.5m (max 1.3 m).

After a brief training, the user perceived the path configuration. As the result, the user experience became more comfortable and natural and the guiding system allowed a real time position adjustment without extra effort.

5 Conclusion

With these experimental setups, we were able to guide people at walking speed using only binaural sound cues. Giving the user an immediate feedback is a key factor in the success of these experiments, therefore designing an efficient positioning system associated with a low latency software had been decisive. The user reception of the device had been very positive, with a very short training, the user followed the sound naturally and quite comfortably.

This device opens lots of possibilities in the domain of the visually impaired guiding, notably for practicing sports or indoor movement. After these preliminary proof of concept with this device, we will investigate to improve the precision of guidance and the comfort of the user, a major objective is also to permit them to pick up more speed. To improve the user perception of sound directions, we will consider using individualized HRTFs. Another possibility to explore is to add some reverberations cues, indeed the perception of the lateral walls reverberation is an important facet for blind people displacement.

Furthermore, the limited size of the experimental area didn't allow them to guide the users at a high enough speed for running or cycling, however, we believe it has the potentials. For this purpose, we will try to improve the operating range of the RTLS system and consider other localization systems, for example, computer vision based systems.

References

[1] Hwang, S., Park, Y., & Park, Y. (2005, June). Sound source localization using HRTF database. In Proc. Int. Conf. on Control, Automation, and Systems (ICCAS2005) (pp. 751-755).

[2] Anderson, M. R. (2001). Direct Comparison of the Impact of Head Tracking, Reverberation, and Individualized Head-Related Transfer Functions on the Spatial Perception of a Virtual Speech Source4. J. Audio Eng. Soc, 49(10), 904-916.

[3] Loomis, J. M., Golledge, R. G., & Klatzky, R. L. (1998). Navigation system for the blind: Auditory display modes and guidance. Presence: Teleoperators and Virtual Environments, 7(2), 193-203.

[4] Strachan, S., Eslambolchilar, P., Murray-Smith, R., Hughes, S., & O'Modhrain, S. (2005, September). GpsTunes: controlling navigation via audio feedback. In Proceedings of the 7th international conference on Human computer interaction with mobile devices & services (pp. 275-278). ACM.

[5] Albrecht, R., Vaananen, R., and Lokki, T. (2016). Guided by music: pedestrian and cyclist navigation with route and beacon guidance. Personal and Ubiquitous Computing, 20(1), 121-145.

[6] Carrasco, E., Loyo, E., Otaegui, O., Fösleitner, C., Dubielzig, M., Olmedo, R., ... & Spiller, J. (2014, July). ARGUS autonomous navigation system for people with visual impairments. In International Conference on Computers for Handicapped Persons (pp. 100-107). Springer International Publishing.

[7] Fielding, R. T. (2000). Architectural styles and the design of network-based software architectures (Doctoral dissertation, University of California, Irvine).

[8] Sahinoglu, Z., Gezici, S., & Guvenc, I. (2008). Ultra-wideband positioning systems. Cambridge, New York.

[9] Warusfel, O. (2003). Listen HRTF database. online, IRCAM and AK, Available: http://recherche. ircam. fr/equipes/salles/listen/index. html (date last viewed 01/06/16).

[10] Aussal, M. (2014). Méthodes numériques pour la spatialisation sonore, de la simulation à la synthèse binaurale (Doctoral dissertation, Palaiseau, Ecole polytechnique).

[11] MyBino Binaural Engine, Available from: http://www.cmap.polytechnique.fr/xaudio/mybino (date last viewed 01/06/16).

[12] Wan, E. A., & Van Der Merwe, R. (2000). The unscented Kalman filter for nonlinear estimation. In Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000. AS-SPCC. The IEEE 2000 (pp. 153-158). Ieee.

[13] Razor 9DOF AHRS, Quality & Usability Lab, Deutsche Telekom Laboratories, TU Berlin & Peter Bartz, Available from : https://github.com/Razor-AHRS/ (date last viewed 01/06/16).