An augmented reality audio device helping blind people navigation

Sylvain Ferrand, Francois Alouges, and Matthieu Aussal

CMAP, Ecole polytechnique, CNRS, Université Paris-Saclay, 91128, Palaiseau, France,
sylvain.ferrand@polytechnique.edu,
http://www.cmap.polytechnique.edu/xaudio

Abstract. This paper describes a device conceived to guide blind people. In this device, we use both localization informations provided by a precise and low latency real-time positioning system and head orientation data measured with an Inertial Measurement Unit, coupled with a real-time binaural engine. In this paper we made a special focus on the localization techniques that we have studied. First experiments show that the approach is viable and the system can be used in practice ...

Keywords: Assistive technology, binaural sound, real-time localization, embedded systems

1 Introduction

The visually impaired are able to localize and follow sound sources with a remarkable accuracy. They often use this ability in everyday activities, for example, to follow someone by hearing his/her voice or footsteps.

In principle, we can use this ability to guide blind people with spatialized sound generated in real-time and listened with simple headphones. Indeed, if we can localize the person precisely in real-time, we can give him/her the illusion that the sound is leading him/her at a certain distance ahead, and guide him/her to follow a pre-defined path.

In this project, we are mainly interested in precision and performances for indoor or sports applications. Our goal, in the first instance, is to guide people in predictable areas with known geometries. Typically, we want to permit a subject to run and stay on his lane in a running track, or to guide him/her to follow a precise path in a gym.

Guiding methods have been experimented in various research project, for example GpsTunes and Ontrack [6, 5, 7] mostly based on GPS positioning systems. Unfortunately, GPS based systems are not able to give an immediate feedback to the user due to their high latency and imprecision. It is the reason why we developed a Real-Time Localization System (RTLS) especially adapted for our application. This precise localization allows us to position the virtual sound very close in front of the user to guide him/her with a very good accuracy.

This article describes the device that we designed, with a focus on the algorithms implemented for real time localization and control. We also briefly describe the audio renderer and show the results that we obtained during our experiments.

2 Device Overview

Our device couples an accurate 2D positioning system, a head tracker, and a sound renderier in order to simulate a virtual sound source regardless of user's motion. It takes into account both head rotation and user's displacement in real-time. We use sensors for the localization and head tracking connected to an embedded computer running the algorithm which implements the path tracking software and the audio engine. The sound is finally reproduced by a headphone.

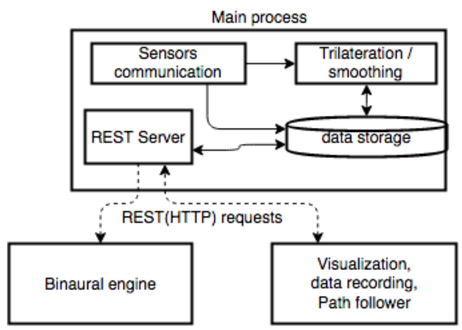

Fig. 1: Software architecture of the device.

The software to be highly modular with a minimal latency operations. This is accomplished through a multi-tier architecture and each component exchange data via a REST API (cf. figure 1). The main program, the REST server, manages the communications with multiple clients processes : communication interface with sensors (head-tracker and positioning system), filtering system, control system and audio rendering. This software, mostly written in python, has been successfully tested on architectures as various as : Mac and embedded computers (ARM 8 (Raspberry Pi 3), ARM *Heterogeneous Multi-Processing - Big-Little* (Odroid XU4 et Jetson TX1) and Intel Edison).

3 Accurate and fast environment sensing

In the literature [6, 5, 7], several applications use Global Navigation Satellite Systems (GNSS) to provide localization data. The accuracy of such localization systems is typically in the range of 2-5 meters horizontal with an update rate in the order of 1-10 Hertz (see table 1). Moreover, they are not usable indoor.

Technology	Positioning accuracy	Refresh rate	Usable place	Cost
GNSS (GPS)	2 - 5 m	1 - 10 Hz	outdoor	low
GPS + IMU	2 - 5 m	100 Hz	outdoor	low
UWB	10 cm	10 - 15 Hz	in / out	medium
Differential RTK GNSS	10 cm	1 - 10 Hz	outdoor	high

Table 1: Comparison of positioning systems

In our application, this is not suffisent, an giving an immediate feedback to the moving user is absolutely crucial. Therefore we propose two different systems which turn out to be faster, more accurate and better suited for our application.

3.1 A Real-Time Locating System (RTLS) with Ultra Wide Band

Ultra Wide Band (UWB) technology allows a fast and precise distance measurement by using the Time of Flight (ToF) of a radio signal. In association with the Two Way Ranging protocol (TWR), we can measure the distance between a master UWB transceiver (the Tag) and multiple slaves (the Anchors). With this protocol, we measure the round-trip time of a data frame sent from the Tag and replied by the Anchors. This elapsed time permits to deduce the distance between the Tag and the Anchors. Compared to systems based on the signal strength (e.g. Wifi, Bluetooth or radio beacons), this technology is very robust in multi-path environments and relatively insensitive to shading of the sensors and RF interference. It is also low-cost, low power and gives high precision measures. The UWB transceivers that we use are based on the DW1000 UWB chip from DecaWave.

In an area with several fixed UWB Anchors with fixed locations, a trilateration algorithm must be used to determine the absolute position of the mobile Tag. A minimum of 3 Anchors is required for 2D localization and 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 measure cycles per second.

In a system with n Anchors at coordinates $a_i = (a_{x,i}, a_{y,i})$, the distance between Anchor i and the Tag is given by :

$$d_i^2 = (x - a_{x,i})^2 + (y - a_{y,i})^2. \quad (1)$$

Basically, with known Anchor positions, the Tag coordinates can be estimated by minimizing Least Square Error (LSE) :

$$\sum_{i=1}^{n} \mid d_i^2 - ((x - a_{x,i})^2 + (y - a_{y,i}))^2 \mid. \quad (2)$$

It turns that, it is possible to localize automatically the Anchors. In an initialization step, we accumulate some distance measurements to each Anchor while moving around them with the device. Once enough data have been collected, we solve the system (1) for each position and each Anchor. The approximation of this overdetermined system of quadratic equations gives an estimation of the position of Anchors and the Tag along the realized path.

Next, for the real-time localization, we use the Anchor positions estimated as before and estimate the Tag position with a Kalman Filter with the measurement function

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

We use an Unscented Kalman Filter (UKF), adapted to deal with the Gaussian noise produced by UWB sensors and the nonlinear function given by (3) and a Position-Velocity (PV) model [8].

We find that, with 4 Anchors positioned at the corners of a 3 m side square, the precision of the localization of a fixed Tag with this system is about 1 cm RMSE and 3 cm max in line of sight conditions.

3.2 Data fusion for accurate long-range localization

If the previous setup is well suited for indoor localization, the limited range of UWB transceiver (about 50m) limits its use for larger areas, moreover it needs several Anchors.

To remedy these limitations we have considered an other strategy. We propose to use only one UWB Anchor and others sensors : a low-cost Inertial Movement Units (IMU) and a velocity sensor. The data measured by those three systems are fused together.

- **The IMU**. It consists of a low-cost Microelectromechanical sensor (MEMS) that measures simultaneously, in the three orthogonal axes, the acceleration, the rotational velocity and the earth's magnetic field. Additionally, a microcontroller can combine all of them to compute a drift-free 3D orientation vector (i.e. the Euler angles). A modern all-in-one chip like the Bosch BNO-055 can perform data fusion to provide orientation informations at 100Hz with 0-drift and very low noise. Notice that we attach the orientation sensor to the body of the user, assuming that he/she moves straight ahead in the direction measured by IMU without any side steps.
- **The velocity sensor**. Unfortunately, it's rather difficult to measure the velocity of a walking person, and most sensors typically used in robotics (e.g. encoding wheels) are unsuitable. We have therefore considered using optical flow sensors (see [1, 2] for instance).We use a very low cost optical flow sensor model DNS-3080 from Avago Technologies which is conventionally used in optical mouse. It can capture 30×30 pixels images at up to 6400 frames per second. In our setup, the optical flow sensor is pointing to the ground estimate the displacement, a 2D vector, by the comparison of successive images.

This sensor is positioned next to an IMU and we estimate the magnitude of the velocity, removing also the rotational component of the optical flow as in [2].

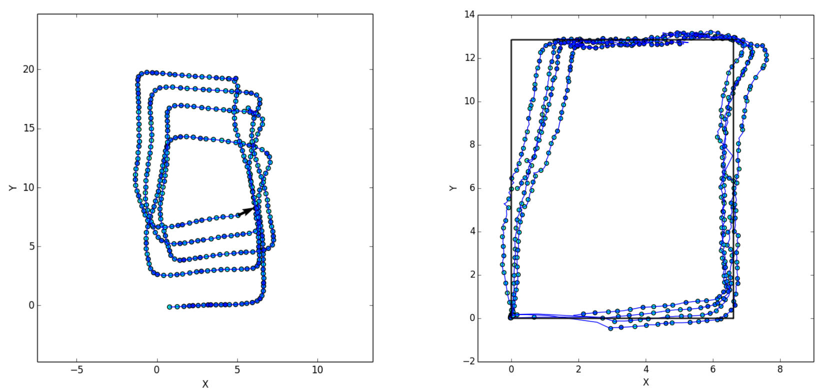

Fig. 2: Position of a user moving along a fixed rectangle estimated with IMU + Optical flow (left) and IMU + Optical Flow + UWB (right). Data fusion with three sensors give a more stable and accurate estimation.

Finally we fuse the measures issued by those sensors with the measure of the distance to a single UWB Anchor, using the measurement vector of the Kalman filter :

$$h(x_k) = \begin{pmatrix} \sqrt{\dot{x}^2/\dot{y}^2} \\ \arctan \frac{\dot{y}}{\dot{x}} \\ \sqrt{(x - a_{x,1})^2 + (y - a_{y,1})^2} \end{pmatrix}. \quad (4)$$

Experiments show that UWB permits to mitigate the accumulation of errors produced using the optical flow sensors only. In the experiments shown in Figure 2 a user is moving along a 6.5m × 12.5m rectangle. Without UWB measures, errors accumulate over time. The Root Mean Square Error of the full system stays below one meter along time. Precision could probably be increased by adding more UWB Anchors.

4 Audio renderer and control

Our ability to localize a sound source comes from the deformation and difference of signal received by each ear and interpreted by the brain. These differences are encoded in the so-called Head-Related Transfer Functions (HRTF)[9], that characterize how ears perceive a sound that comes from a specific direction in space. HRTF can be either measured or simulated, and then used to synthesize a binaural sound that, when listened with headphone, seems to come from this specific direction. In this project, we use the localization described before to create a virtual sound source placed in front of the user, a few meters ahead, in the direction he needs to follow. We render this source with the binaural engine that we implemented [3, 4] using a convolution with the HRTF. This software can spatialize a sound and allows real-time filter selection depending on the position of the user. Furthermore, we use a head tracker to keep the scene stable regardless of user's head movement. This is another IMU fixed on the user's head which measures its orientation. The software has been designed to be low latency and adapts the rendering in real-time to take into account all the positional and orientational parameters.

As mentioned before, the virtual sound source is moving along a predefined path. Nevertheless, we can consider to implement a closed loop Proportional Integral (PI) feedback to enhance the guiding precision. Namely, the sound source can be moved by a correction $F_t = K_p e_t + K_i \sum_j e_j$ proportional of the error e and its accumulation over a sliding window (see figure 3).

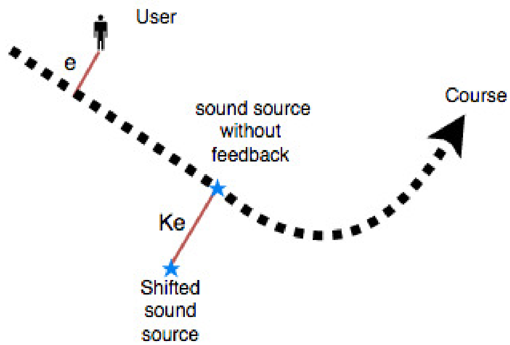

Fig. 3: Proportional feedback illustration

The experiments show that Proportional feedback improves the guiding precision. The Integral feedback is useful to compensate the inaccuracy of localization of the sound source by the user through the binaural rendering.

The whole software, including the audio renderer, has been implemented in Python and Cython for computing intensive tasks.

5 Results

We experimented the device in a 25m $\times$ 35m gym and we parametrized the system to guide the users along an elliptic shape course, as large as possible. The sound stimulus has been chosen to be appropriate for audio localisation, with rich spectral content and rather continuous level. We tested the system with various users, including two blind peoples. These experiments have been made with UWB localization for optimal precision.

Figures 4 and 5 show the localization of two blind users along the elliptic course performing several laps. With roller skates, the blind user nb. 2 has reached the speed of 12km/h. After a 500m long course, at an average speed of 8.7 km/h, he stayed closed to the ideal track with a mean error about 0.5m (max 1.99 m).

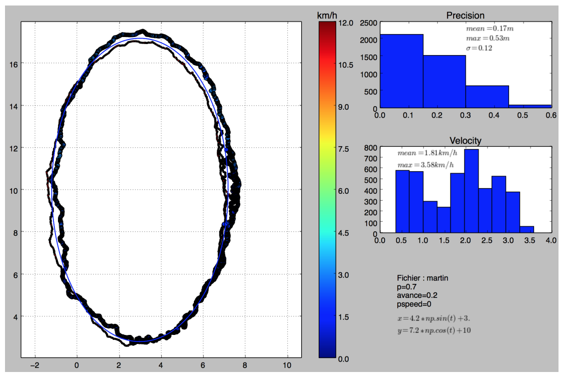

Fig. 4: Blind user nb. 1, walking

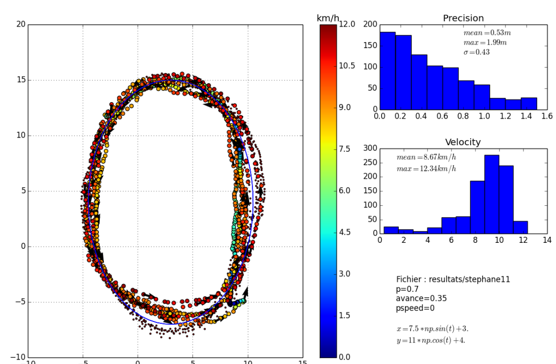

Fig. 5: Blind user nb. 2, roller skating

6 Conclusion

With this device we have been able to guide different peoples, including blind people, running or roller skating, using spatialized sound. 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 is decisive. The user reception of the device is very positive, with a very short training, the user follows the sound naturally and quite comfortably.

This device opens lots of possibilities in the domain of the visually impaired guiding, notably for practicing sports. Currently, we are exploring the effect of reverberations cues, because the perception of the lateral walls reverberation is an important facet of blind people displacement. To go faster, we will also try to improve the operating range of the RTLS system and consider other localization systems, for example, computer vision.

Since the beginning of this project, we involved blind people to test and help us to design this device. We are now committed with two associations to a true partnership for development and device improvement. This partnership have been decisive in the success of this project and continues today.

References

1. Gageik, N., Strohmeier, M., Montenegro, S. (2013). An autonomous UAV with an optical flow sensor for positioning and navigation. International Journal of Advanced Robotic Systems, 10(10), 341.
2. Lim, H., Lee, H., Kim, H. J. (2012, October). Onboard flight control of a micro quadrotor using single strapdown optical flow sensor. In Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on (pp. 495-500). IEEE.
3. Aussal, M. (2014). Méthodes numériques pour la spatialisation sonore, de la simulation la synthèse binaurale (Doctoral dissertation, Palaiseau, Ecole polytechnique).
4. MyBino Binaural Engine,
Available from: http://www.cmap.polytechnique.fr/xaudio/mybino (date last viewed 19/03/18).
5. 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.
6. Jones, M., Jones, S., Bradley, G., Warren, N., Bainbridge, D., & Holmes, G. (2008). ONTRACK: Dynamically adapting music playback to support navigation. Personal and Ubiquitous Computing, 12(7), 513-525.
7. Carrasco, E., Loyo, E., Otaegui, O., Fsleitner, 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.
8. Zhu, Y., & Shareef, A. (2006, August). Comparisons of three kalman filter tracking algorithms in sensor network. In Networking, Architecture, and Storages, 2006. IWNAS'06. International Workshop on (pp. 2-pp). IEEE.
9. Blauert, J. (1983). Spatial hearing.