Sociedade de Engenharia de Áudio

Artigo de Congresso

Apresentado no 11º Congresso de Engenharia de Áudio
17ª Convenção Nacional da AES Brasil
7 a 9 de Maio de 2013, São Paulo, SP

Real-Time Audio Augmented Reality System for Pervasive Applications

Sérgio I. Lopes,[1,2] André Oliveira,[2] José M. N. Vieira,[1,2] Guilherme Campos,[1,2] and Paulo Dias[1,2]

[1] Universidade de Aveiro, Departamento de Eletrónica, Telecomunicações e Informática
3810 Aveiro, Portugal

[2] Instituto de Engenharia Electrónica e Telemática de Aveiro, 3810 Aveiro, Portugal

sil@ua.pt, abo@ua.pt, jnvieira@ua.pt, guilherme.campos@ua.pt, paulo.dias@ua.pt

ABSTRACT

In this paper is presented a real-time audio augmented reality system that enables users to experiment binaural audio according to their position, proximity and head orientation to a particular point of interest. In the proposed system virtual audio sources are placed at points of interest in a specific room. The main goal is to create the illusion that an object acts as a sound source. As our system continuously measures the user head position in space and its orientation, we are able to create an artificial acoustic environment. Two main blocks form the proposed system: (i) a 2D indoor acoustic positioning system and (ii) an auralization system, that depending on the user position and head orientation, generates continuous binaural sound signals.

0 MOTIVATION

This work presents a real-time audio augmented reality system that enables users to experiment virtual acoustic environments based on information of its position and its head orientation to a particular point of interest. This way binaural audio can be generated in real-time to a particular virtual sound source. For example, this system can be used as an immersive audio guide for museums see figure 1. In this case users walking through a particular room could experience an audio guided tour to their position, proximity and head orientation to a particular point of interest. This way is possible to generate virtual binaural audio sources, creating the illusion that the point of interest, e.g. paintings, sculptures or any other work of art, act as a sound sources.

1 SYSTEM ARCHITECTURE

The system prototype is composed by two main modules: the positioning module and the auralization mo-

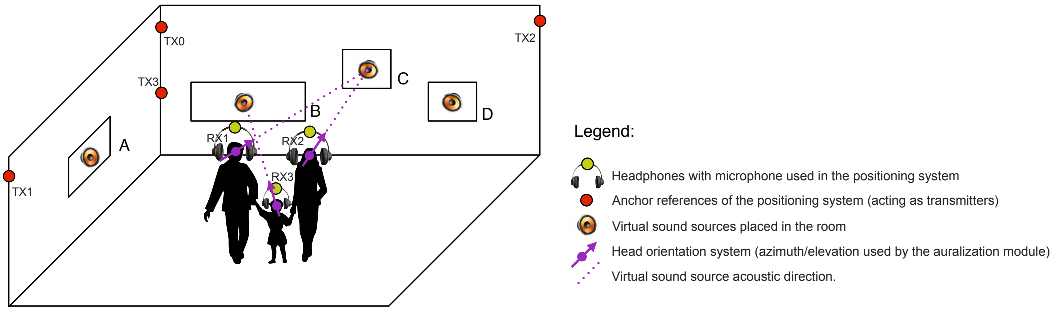

Figure 1: Example of the audio augmented reality concept used in an audio-guide for museums.

dule, see figure 2. The positioning module is used to obtain the user position in real-time. This module uses an infrastructure of acoustic beacons (in this case we used four beacons) and the number of mobile nodes has no limit (GPS-like). The user position is obtained using acoustic ranging. Distance measurements from all intervenient anchors to the mobile station are continuously obtained based on Time-of-Arrival (ToA) estimates. Using these distance measurements a position estimate is then obtained by multilateration using the minimum mean square error estimation method. Conventional low-cost transducers were combined in order to form an acoustic node. Results obtained showed to have sufficient accuracy for the application requirements, with a position estimation rate of $500ms$, an absolute 3D mean error of $20.2cm$ with a standard deviation of $2.3cm$.

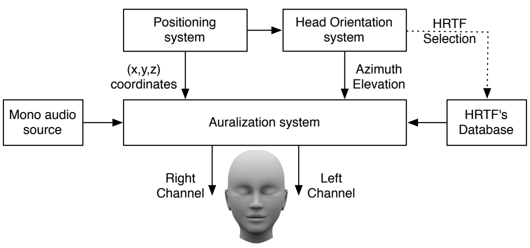

Figure 2: Block diagram of the proposed system.

The auralization system works as a binaural generator based on the position of the user and its head orientation. The core of the auralization is obtained using a set of Head-Related Transfer Functions (HRTF) [1] for spatialization of an anechoic/mono sound signal. Mono and anechoic audio signals are passed through a pair of HRTF filters to generate the binaural audio in order to create the virtual sound sources. The loudness of the virtual audio signals is both controlled by the orientation of the head, using the HRTFs, and by the use of the inverse law in which its amplitude is inversely proportional to the distance to the virtual sound source. The user head orientation is obtained using a commercial 3 DOF inertial orientation sensor, the InterSense InertiaCube3.

2 Acoustic Positioning System

The proposed indoor acoustic positioning system uses broadband ultrasonic signals and ToA measurements. Using low cost transducers we were able to use acoustic chirps of between 20 and 45KHz as pulse signals. This overcomes most of the problems faced by the narrow band signals usually used with common piezo-ultrasonic transducers, which include poor resolution, low environment noise immunity, short range and low robustness to the Doppler effect. Using a synchronized network of ultrasonic anchors in a centralized architecture and time division multiplexing to share the medium a GPS-like system was built [2].

A piezo-tweeter was used as transmitter and a Panasonic WM61A electret microphone as the receiver. We measured the frequency response of several unbranded and commercially available piezo-tweeters before choosing one. In figure 3 a) is presented the frequency response versus the sound pressure level per *Volt (SPL/V)* of the selected piezo-tweeter. In figure 3 b) is presen-

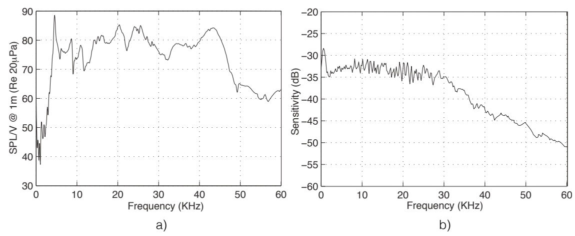

Figure 3: a) SPL Vs. Frequency in the piezo-tweeter @1m(re 20μPa) and b) Sensitivity Vs. Frequency in the Panasonic WM61-A electret microphone.

ted the sensitivity of the WM61A microphone versus its frequency response. Both graphs were obtained when the transmitter and receiver were at $1m$ of distance, respectively, considering the zero-degree case of its radiation pattern. We used a Brüel & Kjær 4954A reference microphone for measurement and calibration [2].

The system uses a standard soundcard with the sampling rate set to 96KHz to continuously measure the distance of each anchor to the mobile device using time division multiple access. The position of the mobile device is then estimated using the multilateration technique. We used four anchors in order to obtain 3D position estimates using the minimum mean square error (MMSE) estimation method, see [2] for more details.

2.1 Positioning System Validation

In figure 4 is presented the XY-grid plane used as the ground-truth in this experiment. The mobile device was placed at each test position (marked with a black cross), with a constant height of 1.75m. One hundred position estimations were computed for each test position. All the estimated measurements for all the positions are presented overlapped in the same XY-axis graph. Note that, no outlier measurements are present, which can be justified by the fact that measurements were taken in laboratory in a controlled acoustic environment, i.e. acoustic noise bellow 40dBSPL. In table 1 is presented a summary of the mean error for all independent coordinates and the correspondent absolute positioning error and standard deviation. In figure 5, a qualitative evaluation of the positioning system (2D case) can be observed when the mobile device was in a predefined moving trajectory. A moving person with the receiver on top of the head was used to evaluate the positioning system in the moving trajectory.

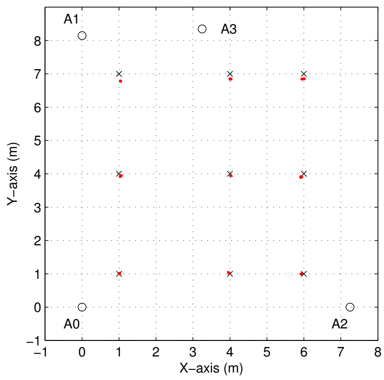

Figure 4: XY positioning results. Real positions - black cross; Estimated positions - red points; Anchor nodes - circles A0, A1, A2 and A3.

Table 1: Mean and absolute errors.

	x	y	z	**Mean**	**2D Abs**	**3D Abs**
Error (cm)	3.2	9.0	17.8	10.0	9.6	20.2
Std (cm)	0.7	0.4	2.2	1.1	0.8	2.3

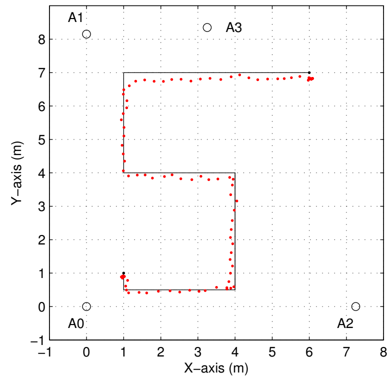

Figure 5: Qualitative results of the positioning system (2D case) for a moving trajectory. Real trajectory - solid black line; Estimated positions - red points; Anchor nodes - circles A0, A1, A2 and A3.

3 Auralization System

The implemented auralization system aims to produce a convincing binaural sound for a listener wandering in an interactive virtual room with virtual sound sources. Given the geometric model of the room, the position of each sound source, the position of the listener and its head orientation, the auralization system processes the monaural sound signal of each source to produce the corresponding auralized binaural sound (figure 6).

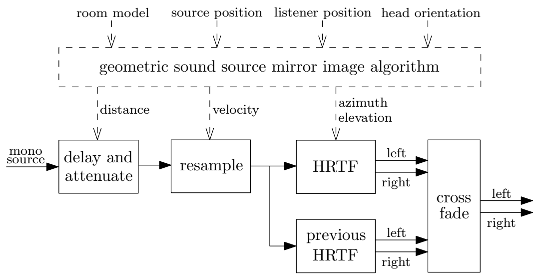

Figure 6: Auralization system block diagram.

The geometric part of the system (represented by dashed lines) continuously calculates the distance, velocity, azimuth and elevation of the listener relative to each sound source, and passes these parameters to the audio part of the system (represented by solid lines). The distance is used by the delay and attenuation block to simulate the propagation delay and attenuation of the sound signal in dry air. The velocity is used by the resampler block to continuously resample the sound signal frequency to simulate the Doppler effect as the listener moves towards, or away from, the source. The azimuth and elevation are used by the HRTF blocks to filter, using the overlap-add method [3], the monaural sound signal with the appropriate HRTF pair of the MIT KEMAR HRTF set [4] to produce the binaural sound. The crossfade block eliminates any discontinuities when switching from one HRTF pair to another, as the HRTF set contains measurements in discrete increments (10 degrees in elevation and mostly 5 degrees in azimuth).

The block diagram in figure 6 depicts the process for one source. The auralization system handles multiple sources by processing each source as described above, as if in parallel, and then summing the binaural outputs. When a geometric model of the room is supplied to the auralization system, it also generates the first reflection sounds for each source, in addition to the direct sounds, greatly enhancing the auralization experience. The image-source method [5] is used to calculate the mirror images for each virtual source. Each mirror image is then treated like if it was just another source, and processed the same, as described above.

The auralization system is implemented in C code, as a library, and is already fairly portable. The only external library used is FFTW [6], for the overlap-add method, but this dependency shall be removed in the future to make this auralization library completely portable and trivial to integrate into different applications. Tests showed that the current implementation is able to process up to 133 sources in real time on an ordinary 2-core 2.5GHz processor.

4 Experimental Validation

To evaluate the proposed system we used four isotropic virtual sources placed in distinct positions in the room. The trajectory presented in figure 7 was performed by two users. In the navigation process the

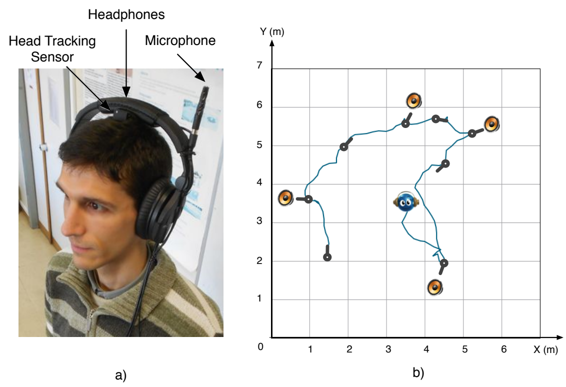

Figure 7: a) User equipped with the prototype. b) Example of trajectory used in experimental validation.

user was asked to rotate its head in the direction of the position of the virtual sound source. This enabled the qualitative evaluation of the proposed auralization system. Both users could effectively detect the azimuth of the virtual sound sources, when they were in the moving trajectory, but with considerable ambiguity in the distinction of front/rear situations. Both users also had problems in distinguish elevation in the virtual sound sources, when included. A video with demonstration can be consulted in the following URL: http://www.ieeta.pt/∼vieira/XXX/aesDemo.mp4.

5 Conclusions and Future Work

In this paper is proposed a reliable real-time audio augmented reality system that enables users to experiment binaural audio according to their position, proximity and head orientation to a particular point of interest. In the proposed system virtual audio sources are placed at points of interest in a specific room. The proposed system uses an acoustic localization system to obtain the user position in a 2D plane with a standard deviation error of less than $10cm$. Based on the HRTF data set from MIT [4] and using a commercial 3 DOF inertial orientation sensor we were able to implement a real-time auralization system that is controlled by the user head orientation and its position in a specific room populated with several virtual sound sources. In the current prototype, the overall system is running on a laptop computer in a wired version. At this moment we are working in the development of a handheld device version. Main changes will include the redesign of the positioning system to use signals in the audio band and TDoA estimation. To improve the performance of the auralization system a method to better distinguish the front/rear perception should be addressed. A possible direction to solve could be the inclusion of higher order room acoustics.

6 Acknowledgements

This work obtained national funding from the Fundação para a Ciência e a Tecnologia (Science and Technology Foundation) under the projects Pervasive Tourism XXXXXXXXXXXXXX and AcousticAVE PTDC/EEA-ELC/112137/2009.

REFERENCES

[1] Corey Cheng and Gregory Wakefield, "Introduction to Head-Related Transfer functions (HRTFs): Representation of HRTFs in Time, Frequency and Space," *J. Audio Eng. Soc. (AES)*, vol. 49, no. 4, 2001.

[2] Sérgio Lopes, José Vieira, and Daniel Albuquerque, "High Accuracy 3D Indoor Positioning using Broadband Ultrasonic Signals," in *Trust, Security and Privacy in Computing and Communications (TrustCom), 2012 IEEE 11th International Conference on*, 2012, pp. 2008–2014.

[3] Udo Zölzer, *Digital Audio Signal Processing*, Wiley-Blackwell, 2008.

[4] Bill Gardner and Keith Martin, "HRTF Measurements of a KEMAR Dummy-Head Microphone," 1994, http://sound.media.mit.edu/resources/KEMAR.html.

[5] Lauri Savioja, Jyri Huopaniemi, Tapio Lokki, and Riitta Väänänen, "Creating Interactive Virtual Acoustic Environments," *J. Audio Eng. Soc. (AES)*, vol. 47, no. 9, 1999.

[6] Matteo Frigo and Steven Johnson, "The Design and Implementation of FFTW3," *Proceedings of the IEEE*, vol. 93, no. 2, pp. 216–231, 2005.