Entertainment Computing 12 (2016) 9–27

Contents lists available at ScienceDirect

Entertainment Computing

journal homepage: ees.elsevier.com/entcom

A low-cost infrared-optical head tracking solution for virtual 3D audio environment using the Nintendo Wii-remote ☆

Yashar Deldjoo [a,b], Reza Ebrahimi Atani [b,*]

[a] Department of Signals and Systems, Chalmers University of Technology, Gothenburg, Sweden
[b] Department of Computer Engineering, University of Guilan, P.O. Box 3756, Rasht, Iran

ARTICLE INFO

Article history:
Received 29 September 2014
Revised 21 September 2015
Accepted 31 October 2015
Available online 10 November 2015

Keywords:
Head tracking
Rigid body
Degrees of freedom
Virtual 3D audio system
3D audio rendering
Wii-remote

ABSTRACT

A virtual audio system needs to track both the translation and rotation of an observer to simulate a realistic sound environment. Current existing virtual audio systems either do not fully account for rotation or require the user to carry a controller at all times. This paper presents a three-dimensional (3D) virtual audio system with a head tracking unit that fully accounts for both translation and rotation of a user without the need of a controller. The system consists of four infrared light-emitting diodes on the user's headset together with a Wii-remote to track their movement through a graphical user interface. The system was tested with a simulation that used a pinhole camera model to map the 3D-coordinates of each diode onto the two-dimensional (2D) camera plane. This simulation of 3D head movement yields 2D coordinate data that were put into the tracking algorithm and to reproduced the 3D motion. The results from a prototype system, assembled to track the 3D movements of a rigid object were also consistent with the simulation results. The tracking system has been integrated into an Ericsson 3D-audio system and its effectiveness has been verified in a headtracked virtual 3D-audio system with real-time animating graphical outputs.

1. Introduction

The human auditory system plays a major role in our three-dimensional (3D) spatial awareness and in extracting information from our environment. It is through hearing that we localize sound sources and identify and follow the movement of such sources around us. Human sound perception is based on binaural hearing, which results in the same sound reaching each ear at a slightly different time and with different intensity. The first phenomenon is called the interaural time difference (ITD) and the second the interaural intensity difference (IID). Further, sound waves interact with the torso, head, and especially the external ear or pinna, becoming further altered before reaching the eardrums. These interactions modify the frequency content of the signals by reinforcing some frequencies and attenuating others depending on the sound's direction of arrival. Therefore, the frequency spectrum reaching one ear will be slightly different from that reaching the other. The brain uses the IID, ITD, and spectral difference (spatial cues) between signals received by the ears to determine the location of the source [1,2].

Virtual 3D audio technologies use specialized filters known as head-related (HR) filters to render sound. Because HR filters contain all the information needed to locate a sound source, they can artificially spatialize sounds if the appropriate HR filters are known, a process known as binaural synthesis. HR filters are unique for every position and angle of incidence and are usually measured for sound sources in many locations relative to the head to obtain a database of hundreds of HR filters. Using HR filtering of the source signal, a virtual audio system can simulate 3D wave propagation that triggers the spatial hearing of the listener. As shown in Fig. 1, the listener experiences a 3D audio environment when listening to a generated sound signal, which is different from the 3D environment in which he or she actually exists in.

A major problem for virtual 3D audio systems is head movement [1,2]. This changes the direction of arrival (DOA), and hence the path of each sound wave to each eardrum, thereby changing the spatial information the brain needs to correctly locate the source. Therefore, to have a realistic virtual 3D audio system, head movements have to be tracked. When the listener's orientation is not tracked, the simulated 3D audio space moves with the head movements of the listener, which is not realistic. A dynamic virtual

☆ This paper has been recommended for acceptance by Ryohei Nakatsu.
* Corresponding author.
E-mail addresses: yashar.1984.deldjoo@ieee.org (Y. Deldjoo), rebrahimi@guilan.ac.ir (R.E. Atani).

http://dx.doi.org/10.1016/j.entcom.2015.10.005
1875-9521/

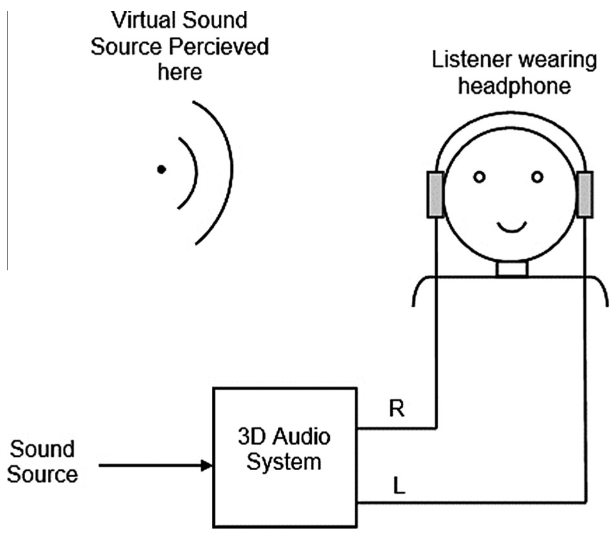

Fig. 1. 3D audio system using binaural synthesis.

3D audio system in which the user's movement alters the sound environment requires the following:

1. Head tracking (HT) with six DOFs to account for both position and orientation.
2. Fast processing of the tracking algorithm to make changes in virtual sound properly correspond with changes in the listener's position and orientation. Ideally the brain should not detect any delay between change in position/orientation and change in sound properties.
3. Accuracy in detecting events, specifically the type of event. For example, an upward–downward translation should not be reported as a rotation around the *y* or *x* axis (Fig. 2).

Visual HT systems can be categorized into two classes: marker-free systems and marker-based systems. Marker-free systems only exploit optical sensors to measure movements of the human head. They acquire the information directly from the recorded images without trying to build a 3D representation of the user's head. Building such systems usually involves tracking of facial features that are non-rigid and subject to various articulation and deformation due to muscle contraction and relaxation and usually results in movements in high degrees of freedoms (DOFs). These models are also known as appearance-based methods and are classified in 2D or 3D tracking methods [3–5]. On the other hand, marker-based systems apply identifiers which are placed upon the human head to capture movements. These systems build up a 3D model that represents human head and then try to estimate the 3 parameters that define the translation and 3 parameters that define the rotation with respect to reference coordinate system shown in Fig. 2. This results in tracking systems with 3-Degrees-Of-Freedoms (3DOF) or 6DOF in which the latter is considered as the most complete representation of a rigid object in three-dimensional space. The marker-based systems are referred to as model-based systems [4–6].

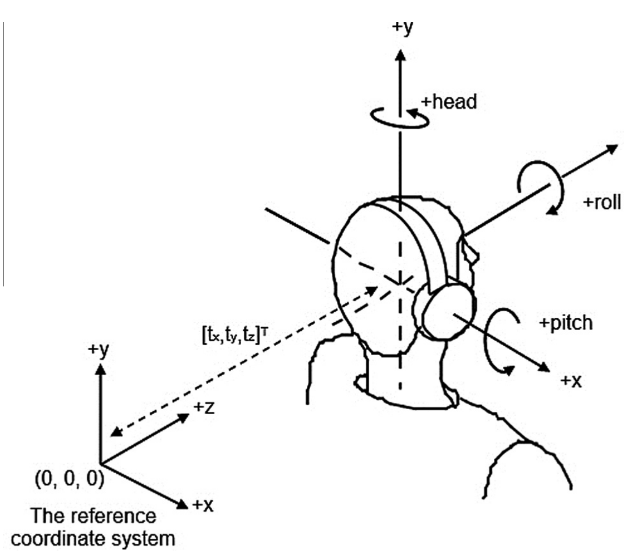

Fig. 2. Axis rotations (pitch, yaw and roll) and translation specified by a translation vector $[t_x, t_y, t_z]^T$ in a HT system.

Some 2D and 3D approaches have been developed for estimating head orientation. The 2D approaches [3] compare each new head image with a set of reference templates and then use the closest-matching template as the head pose. The advantages and disadvantages of such systems are discussed in [7]. Lee [8] has developed a 3D HT system that produces view-dependent images. This system uses a head-mounted sensor bar with a Nintendo Wii remote (Wiimote™) at the base of the display. It assumes that the observer is always looking into the middle of the screen, and does not account for head rotation. Kreylos' system [9] uses a direct rigid-body transformation to yield an approximate solution to the orientation problem, but that system requires the user to carry a controller and uses the controller's inertial data to achieve 6DOF tracking. Other devices that track all the DOFs similarly require the user to carry or wear a controller, which is inconvenient, cumbersome, and difficult to implement at home [10].

Wiimote is used in several other applications such as automated assembly simulation, head position tracking for gaze point estimation, home assessment of Parkinson's disease and as a tracking system for braille readers, etc. In [11] a low cost and simple location management system using the Wii remote controller and infrared LEDs is proposed in which a Wiimote controller is placed on a mobile robot pointing upward toward a number of IR LEDs placed on the ceiling. In [12] a Wiimote-based motion capture system is developed for automated assembly simulation. In [13] a pair of Nintendo Wiimote imaging sensors is used to create a stereo camera for 6DOF position tracking of the headset for eye gaze estimation. Application of the Nintendo sensor-rich data for building of a home-based assessment of Parkinson's disease (PD), known as WiiPD is presented in [14]. Another example is shown in [15] where a Wiimote and a refreshable braille display are used to build a cheap an easy-to-use finger tracking system for studying braille reading. A low-cost high accuracy 3D tracker is implemented in [16] using the Wiimote to detect the pose of a target. It applies the triangulation techniques to build the 3D location of the markers. In Fig. 3, the main criteria for design of a Wii-Based HT solution is presented. As can be seen, the problem of HT is an optimization problem on several parameters. In general, one of the main advantages of Wii-based HT systems is the low cost of building such systems. The next important factor is the precision of tracking which in the case of rigid objects can be performed with a precision of 3DOF to maximum 6DOF. On the other hand, there is a growing trend on using multiple Wii-motes to create a stereo-vision based system using triangulation technique which makes it easier to bring 6DOF tracking into reality. The camera mode is yet another determining factor in that the camera can be used in two stationary or non-stationary modes for the task of HT.

This research aims to solve the virtual 3D HT simulation system by developing a visual tracking system based on infrared positioning that updates the listener's orientation vectors without the cumbersome need of carrying a somewhat heavy Wii-remote