Implementation of 3D Audio Effects using Head Related Transfer Function (HRTF) for Real Time Application using Blackfin Processor

Ganesh V N
1 MIT E/Department of E&C, Moodabidri, Karnataka, India
Email: ganesh@mite.ac.in

Abstract— Usually, in order to create 3D audio sound to make 3D effects using binaural systems, we use Head-related transfer functions (HRTFs) to describe the spectral filtering that occurs between a source sound and the listener’s eardrum. Because of the characteristics of HRTF some three-dimensional effects in the area of a cone of confusion between front and back directions can be declined.

In order to make the practical implementation of 3D audio from a mono-aural sound a large set of HRTFs for different azimuth and elevations must take into consideration, since HRTFs vary as a function of relative source location and subject. This paper describes 3D audio to make 3D effects using mono-aural sound. This can be achieved by using bank of filters called head related impulse responses in time domain at different 3D positions.

In this paper we proposed a real-time DSP implementation of a 3D sound localization algorithm to be run on high speed processor (Blackfin-537), on which sound quality is evaluated by a subjective listening test.

Index Terms— HRTF, binaural sound systems, FIR, IIR filters, acoustics, spatialization.

I. INTRODUCTION

3D sound is becoming most important for scientific, commercial, and entertainment systems [3]. It increases the auditory interfaces to Computers to improve the sense of presence for virtual reality simulations, and add excitement to computer games. 3D sound places the virtual sound source with a mono-aural sound in a given 3D space by adding the pitch, tone and sense of direction and distance. In general the most common implementation of surround systems uses 5.1 channels is DVD standards to do audio systems. Actually we need six speakers for a system with 5.1 channel speakers, which take large space and huge money. [4] Hence, binaural systems that uses only 2 channels come into the spotlight.

The recent sound systems have used HRTF to compose virtual 5.1 channels. HRIRs in time domain or HRTFs in frequency domain uses a series of algorithms utilized to synthesize simulated binaural signals from a monaural source. It includes the cues from the scattering process of sound from the user’s ears, body, and head [5]. However, because of the listener’s individualistic qualities and physical characteristics, the use of non-individual HRTF can create confusion in front/back and up/down directional perception. The listener especially cannot separate the difference of each direction in the cone-of-confusion with just a difference of time or difference of level. To eliminate this problem we need a specific HRTF for each individual which cannot be accomplished in the real world.

DOI: 03.AETS.2014.5.362
© Association of Computer Electronics and Electrical Engineers, 2014
In HRTF 3D effect is accomplished by using a series of FIR filters, commonly called head-related impulse responses (HRIRs) in time domain or head-related transfer functions (HRTFs) in frequency domain. A mono sound source can be perceived as coming from 30° right with respect to the front of the listener by filtering (or convolving) the original signal with the impulse responses of two HRIR filters (left and right) that correspond to the 3D position of 30° right and 0° elevation. This filtering reconstructs the acoustic pressures at the listener’s ears that would occur with a free-field sound source at the desired location.

A. HRTF Measurement and Filter Generation

In recent years, in the field of acoustic signal processing, several sophisticated approaches have been attempted for realizing 3D sound effects [3], which are based mainly on the so-called Head-Related-Transfer Function (HRTF), which is influenced in the human ears [6]. In general, given a sound source, the 3D localization of the sound can be realized on the basis of the HRTFs from the source to the right and left ears. HRTF measurement is required to produce 3D sound that is designed for a specific person. Usually, these measurements are done using a dummy head and miniature microphones placed in the ear. This allows to microphones to pick up sound after it has been altered by both head shadowing and the pinna. One such measurement is summarized by Klaus Hartung et al [11]. A dummy head and in-ear microphones were used to record sound generated by 17 loudspeakers in an anechoic chamber. The data recorded was then processed to create the coefficients for an FIR filter that replicates the transfer function.

B. HRTF analysis and synthesis

The most common method of measuring the HRTF of an individual is to place tiny probe microphones inside a listener’s left and right ear canals, place a speaker at a known location relative to the listener, play a known signal through that speaker, and record the microphone signals. By comparing the resulting impulse response with the original signal, a single filter in the HRTF set has been found. After moving the speaker to a new location, the process is repeated until an entire, spherical map of filter sets has been devised. Every individual has a unique set of HRTFs, also called an ear print. However, HRTFs are interchangeable, and the HRTF of a person that can localize well in the real world will let most people localize well in a simulated world. While generic, interchangeable HRTFs are suitable for general applications such as video conferencing or games, individualized HRTFs are useful for performance critical applications of binaural audio, such as jet fighter cockpit threat warning systems, or air traffic control systems. Once an HRTF has been devised, real-time DSP (digital signal processing) software and algorithms are designed. This software has to be able to pick out the critical (psycho-acoustically relevant) features of a filter and apply them in real-time to an incoming audio signal to “spatialize” it. The system works correctly if a listener cannot tell the difference between listening to a sound over the speaker setup from the analysis process above (the speaker is in a specific position), and the same sound played back by a computer and filtered by the HRTF impulse response corresponding to the original speaker location.

C. Literature Survey

General information about the 3D localization of the sound can be realized on the basis of the HRTFs from the source to the right and left ears are discussed by Chong-Jin and Woon-Seng Gan [1], E. M.Wenzel [2]. The robustness analysis, localization characteristics, structural model for binaural sound synthesis and head related transfer functions are discussed in detail by C. Phillip Brown and Richard O. Duda [3], K. Koo and H. Cha [4]. The concept of virtual audio signal and localized spatial audio signal concepts are considered by D. N. Zotkin and R. Duraiswami [5]. The implementation concepts about 3D audio signals are explained by Gardner W [7]. The basic concepts 3D sound systems are explained in detail by E. M. Wenzel [2]. An efficient HRTF model for 3D sound is proposed by Brown, C.P., Duda, R.O., [8]. Comparison of different methods used for HRTF is discussed by K. Hartung, J. Braasch, and S. J. Sterbing, [9]. Wakefield and Durant propose a genetic algorithm to generate a pole-zero HRTF file from measured impulse responses. Using the genetic algorithm, it is possible to use a directional transfer function and common transfer function (CTF) to generate the HRTF. Vikas C. Raykar, Ramani Duraiswami, Larry Davis, B. Yegnanarayana[12] considered the significance of group delay function in signal reconstruction from spectral magnitude and phase of the signal in realising 3D audio effects.

As the work is implemented using ADSP- 537 EZKIT reference is made to the manual of Analog Devices. The complete work proposed in this paper is implemented using ADSP-537 Blackfin processor.
II. DESIGN AND IMPLEMENTATION

The Head Related Transfer Functions (HRTFs) are acoustic filters that vary both with frequency and with azimuth, elevation and range to the source. If a monaural sound signal representing the source is passed through these filters and heard through Headphones, the listener will hear a sound that seems to come from a particular location in space. Appropriate variation of the filter characteristics will cause the sound to appear to come from any desired spatial location.

A. HRTF Implementation

The software program uses the function “filter” to create a binaural (left & right) pair of sounds from a monaural input “wave” file. The “filter” function performs the convolution of the digitized input sound with the impulse response of the appropriate HRTF. The impulse response of a filter is the output sequence obtained from the filter when the input is a discrete impulse. In the case of an HRTF we call the impulse response a “Head Related Impulse Response” or “HRIR”. The operations performed by software when the “filter” routine is executed are as shown in Figure 1.

\[ y(n) = b(1) \cdot x(n) + b(2) \cdot x(n-1) + \ldots + b(n) \cdot x(n-n) - a(2) \cdot y(n-1) - \ldots - a(n) \cdot y(n-n) \quad \ldots (1) \]

Where \( n-1 \) is the filter order.

Alternatively, the operation of filter at sample \( m \) is given by the time domain difference equations,

\[ y(m) = b(1) \cdot x(m) + z_1(m-1) \quad \ldots (2) \]

\[ z_1(m) = b(2) \cdot x(m) + z_2(m-1) - a(2) \cdot y(m) \quad \ldots (3) \]

\[ z_{n-2}(m) = b(n-1) \cdot x(m) + z_{n-1}(m-1) - a(n-1) \cdot y(m) \quad \ldots (4) \]

\[ z_{n-1}(m) = b(n) \cdot x(m) - a(n) \cdot y(m) \quad \ldots (5) \]

The input-output description of this filtering operation in the z-transform domain is a rational transfer function,

\[ Y(z) = \frac{b(1) + b(2)z^{-1} + \ldots + b(n)z^{-n}}{1 + a(2)z^{-1} + \ldots + a(n+1)z^{-n}} \cdot X(z) \quad \ldots (6) \]

Since the HRTFs measured for the purpose of 3D sound spatialization are modelled as non-recursive filters (i.e., no internal feedback is assumed in the HRTF models), all the “a” coefficients in the above equations will be assumed to have a value of zero, which simplifies the calculations significantly. In fact, under these conditions, the “b” coefficients in the above equations are directly the numerical values that constitute the HRIRs.

B. Results And Discussions

Unlike stereo sound, which only allows lateral listening, 3D sound enables sound to be positioned around the listener’s head at different directions with only a pair of loudspeakers or a headphone. This effect is achieved by using a set of FIR filters, commonly called head-related impulse responses (HRIRs) in time domain or head-related transfer functions (HRTFs) in frequency domain. For example, Figure 2 shows that a mono sound source can be perceived as coming from 30° right with respect to the front of the listener by convolving (or filtering) the original signal with the impulse responses of two HRIR filters (left and right) that correspond to the 3D position of 30° right and 0° elevation. This filtering reconstructs the acoustic pressures at the listener’s ears that would occur with a free-field sound source at the desired location. The HRIRs are usually measured with microphones inside the ears of a dummy head and loudspeakers that are placed at different positions that emit white noise [10]. A popular HRIR data set which consists of 710 different positions at elevations from −40° (i.e., 40° below the horizontal plane) to +90° (directly overhead).
At each elevation, 360° of azimuth was sampled in equal increments, and the definition of the angle of horizontal azimuth is shown in Figure 2.

The convolution of a mono sound source with a set of HRIR filters at an azimuth of 30° and 0° elevation. The Plot of the HRIR and note the differences between the two HRIRs at an azimuth of 30° as shown in Figure 3.

Note that the right HRIR has higher amplitude and responds earlier compared with the left HRIR. The HRIR at the azimuth of 30° is transformed to frequency domain to obtain the HRTF plots as shown in Figure 4.
In Figure 4 we note that the right HRTF (solid line) has a higher magnitude (>10 dB) than the left HRTF (dashed line), and the peaks and notches of the HRTFs occur at different frequencies. These different spectral envelopes contain important cues to convey the spatial information of the sound source. We also can hear the original and filtered signals with a headphone at different azimuth angles of 0°, 30°, 60°, 90°, 120°, 150°, 180°. The simulation results of HRIR for different azimuth angles are shown in Figure 5, Figure 6, Figure 7 and Figure 8 respectively.

The Blackfin EZ-KIT allows users to select and play back signals with different 3D effects in real time. The audio signal can be continuously sampled at the audio input channels of the Blackfin processor, and 3D audio
processing is carried out. The 3D audio effects can be changed on the fly by pressing different switches on the EZKIT. This real-time processing platform supports a useful functionality test for interactive applications. The real-time implementation of binaural spatializer with the BF537 EZKIT, a stereo or mono sound source is connected to the audio input channels of the EZ-KIT, and the output of the EZ-KIT is connected to a headphone. The default starting mode of the spatializer is the “Static” mode, where sound can be moved from one position to another with every press of a button. The signal source is considered as a mono sound source.

![Diagram of two-source binaural spatializer](image)

Figure 9: Two-source binaural spatializer

The second mode is the “Mixed” mode. To switch to this mode, press the switch SW10 for the BF537. The “Mixed” mode implements the two source binaural spatializer as shown in Figure 9. This binaural spatializer places two sound sources in symmetric locations, and these sound sources can be perceived as moving clockwise or counter clockwise in tandem.

![Graph of IN and OUT signals](image)

Figure 10: Snapshot of Two HRTFs at Static mode

The third mode is the “Pass-through” mode, which passes the left and right input channels to the output channels without processing. This mode is used to compare the differences between the 3D audio effects and the “Pass-through” without processing. The final “Dynamic” mode in this project moves a single sound source around the head. The sound position is automatically changed by 30° clockwise in every second. A cross-fading method is used in this mode to avoid any sudden discontinuation that creates clicking distortion when moving from one position to another. If the switch SW10 for the BF537 is pressed again, it will rotate back to the “Static” mode. The results of 3D sound effects for different modes are done and the static mode is shown in snapshot of Figure 10.

III. CONCLUSIONS

The real time system performed very well, using inexpensive components the complete 3D sound system is implemented in real time. Additionally, without making any significant hardware changes the system provided enough flexibility to test different computationally intensive algorithms. The VisualDSP++ software provided a very easy to use interface to the real-time system. The user was able to fully operate the 3D sound system, including changing the position of virtual sound sources in the real time system. The simulation results obtained by VisualDSP++ are successfully compared and verified with MATLAB.
The multiprocessor system described in this paper offers a low cost and powerful real time system for the real time implementation of audio algorithms. The system is an excellent choice for the development and implementation of computationally intensive audio algorithms. This paper is an initial step in using head related transfer functions to create acoustic events. Further the system can be analyzed in terms of solving the “front-back confusion” problem using the following work. To extend existing HRTF databases new measurements are required with a better resolution of azimuth and elevation or by using interpolation to obtain a desired HRTF. In this thesis a FIR filter approach was used, but also an IIR approach filter could be considered to see if the system could be enhanced. The paper provides a basis for further developments like head-tracking or the simulation of room acoustics.

REFERENCES