Audio Engineering Society

Convention Paper 7279

Presented at the 123rd Convention
2007 October 5–8 New York, USA

Real-time Spatial Representation of Moving Sound Sources

Christos Tsakostas[1] and Andreas Floros[2]

[1] HOLISTIKS Engineering Systems, Digeni Akrita 29, 122 43 Athens, Greece
tsakostas@holistiks.com

[2] Dept. of Audiovisual Arts, Ionian University, 49 100 Corfu, Greece
floros@ionio.gr

ABSTRACT

The simulation of moving sound sources represents a fundamental issue for efficiently representing virtual worlds and acoustic environments but it is limited by the Head Related Transfer Function resolution measurement, usually overcome by interpolation techniques. In this work, a novel time-varying binaural convolution / filtering algorithm is presented which, that takes into account both physical and psychoacoustic criteria, can efficiently simulate a moving sound source. It is shown that the proposed algorithm overcomes the excessive calculation load problems usually raised by legacy moving sound source spatial representation techniques, while high-quality 3D sound spatial quality is achieved in both terms of objective and subjective criteria.

1. INTRODUCTION

Spatial representation of a number of sound sources around a listener represents a significant feature for implementing realistic virtual world environments, game engines and special-purpose simulators. Although many alternative spatial sound reproduction technologies exist (such as analog and digital surround sound formats [1] – [2], wavefield synthesis [3] and ambisonics [4]), binaural reproduction represents a very promising and attractive format for immersive audio applications, especially when increased portability is required [5].

It is well-known that using binaural technology, the spatial placement of any virtual sound source is performed by filtering monaural recorded or synthesized sound with Head Related Transfer Function (HRTFs) [6]. In general, the latter functions describe the paths between a sound source and each ear of a human listener in terms of a) the interaural time difference (ITD) imposed by the different propagation times of the sound wave to the two (left and right) human ears and b) the interaural level difference (ILD) introduced by the different propagation path lengths, as well as the shadowing effect of the human head.

In real sound environments, it is very usual to listen to moving sound sources. Obviously, sound source (or receiver) movement introduces variable source – receiver spatial relative positions. Hence, the representation of a moving sound source using binaural technology requires the employment of a set of HRTFs that corresponds to all possible spatial locations. However, these functions are typically measured for a specific finite set of horizontal and vertical angles and positions. In order to overcome this restriction, HRTFs interpolation techniques can be employed [7]. As it will be explained later, this process induces additional, high computational load and renders real-time binaural processing implementations insufficient. Moreover, when moving sound sources (or receivers) are simulated, additional psychoacoustic parameters and cues should be considered for authentic sound reproduction, such as the Doppler Illusion [8] and the precedence effect [9], [6].

In this work, a novel frequency-based, time-varying filtering method is introduced that minimizes the need of additional computations for accurate moving sound-sources representation. This mechanism additionally takes into account psychoacoustic criteria and cues for perceptually optimizing the 3D audio representation performance.

The paper is organized as following: Section 2 presents a general overview of binaural technology and its application to moving sound sources simulation while in Section 3, a generic model for binaural sound motion simulation is introduced and analyzed. In Section 4, typical results of the performance of the proposed moving sound source simulation model are presented, using both objective and subjective criteria. Finally, Section 5 concludes this work.

2. BINAURAL SOUND MOTION SIMULATION BACKGROUND

Binaural technology is based on binaural synthesis [10], which employs the impulse responses (called HRTFs) from the audio source to each of the two human eardrums and allows for accurate sound source immersion. Binaural synthesis may also incorporate sound field models of closed enclosures for producing binaural room simulations, allowing for the accurate 3D sound representation in real or virtual spaces and rooms [11].

In the context of real-world applications (such as interactive virtual world representations) the realistic simulation of moving sound sources is nowadays more than required; However, as it was previously mentioned, it imposes a number of difficulties originating mainly from the discrete set of positions for which the HRTFs are usually measured and from the time - varying characteristics and parameters of such a system. Available digital signal processing methods for moving sound sources binaural synthesis are generally based on static structures, namely Linear - Time - Invariant (LTI) systems. Time - varying outputs can be obtained by either (a) *Output Cross – Fading* or (b) by *Parameter Cross – Fading*. In output cross – fading the output is a blend of input processed simultaneously with past and present parameters. On the other hand, parameter cross – fading relies on a varying set of rendering parameters [12].

Binaural sound motion simulation has already been intensively investigated by many researchers [13] – [17]. In general, such processing takes place in the time domain or in the frequency domain (overlap – add methods) or in time – frequency representations. Parameter cross – fading is usually used in the time domain, where the HRTFs are either approximated by IIR filters, or minimum phase filters where the interaural time differences ITDs are pre-calculated and simulated by time varying digital time delays. For the case of frequency domain processing output cross – fading can be applied (or alternatively an overlap on the input data), which smoothes out the transients occurring when filters are changed [18].

The main drawback of time – domain methods is that they are based on alteration of the features of the HRTFs (minimum-phase or IIR approximations). On the other hand, simple output cross-fading, apart from being computationally burdensome, as a mixture of two different systems might not resemble that of a single system intermediate between the two [12].

Authentic representation of moving sound sources in virtual auditory environments requires smooth rendering. Among others, this is accomplished with either high spatial resolution ($\leq 5^{\circ}$) of the HRTF measured catalogues or with appropriate interpolation algorithms [7].

A good summary of HRTF interpolation techniques has been reported by Hartung et al. [7]. In general, HRTF interpolation can be performed in the time domain [19], the frequency domain (i.e. in terms of magnitude and phase) or on other representations like principle - component - analysis - coefficients (PCA) [20], Karhunen - Loeve transformations [21], and pole - zero models [22]-[24].

3. EFFICIENT BINAURAL SOUND MOTION SIMULATION

During this work, the following requirements were considered as fundamental parameters for efficiently designing moving sound sources simulation techniques in real-time:

a) The computational complexity should be as low as possible in order to be able to execute in real-time. Towards this aim, FFT-based Overlap-Add block convolution was selected as the basis of the real – time processing.
b) All HRTFs employed [25] are pre-measured FIR filters. The HRTF length is 512 samples, while the sampling frequency equals to 44.1 kHz. Moreover, the HRTFs features are not altered in any way (like minimum phase, IIR approximations etc.). Only HRTF equalization is applied based on diffuse – field – equalization [11].
c) In order to assess the perceptual effect of the moving sound source simulation, no room-simulation was considered in this work.
d) For authentic sound source movement, psychoacoustic parameters and criteria must be considered, typically such as the *Minimum Audible Angles* (MAAs), the *Minimum Audible Movement Angles* (MAMAs), as well as the Doppler Illusion.

The above requirements and constraints are concentrated in a generic model for binaural sound motion simulation, depicted in Figure 1. In the above generic binaural sound motion simulation model it is assumed that a sound source can move from a position A to a position B, on an arbitrary trajectory in a 3-Dimensional space, as depicted in Figure 2. During this movement, the positions of both the sound source and the listener must be sampled and any changes should be tracked for updating all the necessary motion - related parameters.

The *Motion Parameters Estimation & Decision Module* is responsible for the estimation of the motion parameters (velocity, speed, acceleration). As it is shown in Figure 1, depending on the values of these parameters it is decided whether Doppler Illusion Simulation [8] should be applied and whether the Binaural Impulse Responses (BIRs) should be updated.

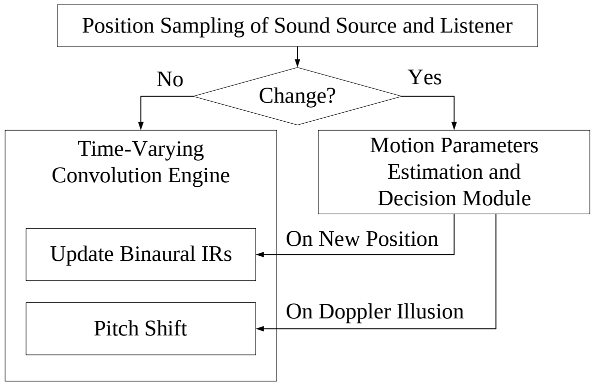

Figure 1 Architecture of the proposed generic sound motion algorithm

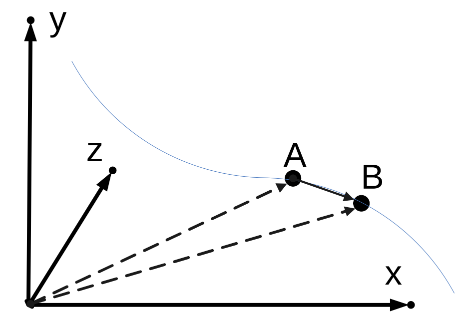

Figure 2 Sound source movement from position A to position B in 3D space

The *Time-Varying Convolution Engine* is primarily responsible for producing audio data with no artifacts for changing positions and thus filters coefficients. Additionally, it should simulate the Doppler Illusion. For the time - varying convolution engine, a *Time Frequency Representation* (TFR) based on *Short Time Fourier Transform* (STFT) signal processor was selected, followed by all the necessary spectral modifications [26].

3.1. Motion Parameters Estimation & Decision Module

As it was mentioned in the previous Section, this module is responsible for the estimation of the motion parameters (velocity, speed, acceleration). A critical factor for minimizing the computational complexity required is to process changes in the position of a moving sound source only when these are perceptually important. Psychoacoustical data concerning MAAs and MAMAs provide valuable information to this approach. A good summary of the available data is given in [27]. Most of these data have been encapsulated in the motion parameters & decision module as simple look - up tables, where the minimum values of MAAs and MAMAs have been considered. If, for example, a sound source moves with 128°/ second, according to Perrot & Tucker [28], the smallest MAMA is 6° at 950 Hz. Consequently, the binaural filters are sufficient to change every 128 / 6 ≈ 21 times per second. If, on the other hand, a block length of 1024 samples is used at a sampling frequency of 44100 Hz, it makes 44100 / 1024 ≈ 43 times per second, which is the maximum processing rate. Obviously, it is not necessary to change the filters 43 times per second when a change rate equal to 21Hz is perceptually enough. Thus, in this case, significant computational effort can be saved.

When FFT based convolution is applied, for each change of the position of a moving sound source, a complex FFT of the new HRTF must be calculated. Apart from that, as mentioned earlier, HRTF interpolation introduces additional computational load, which is far from negligible. In the current study HRTF catalogues of 5° degrees spatial resolution in azimuth were used [25], with no interpolation. This spatial resolution was considered to be adequate for the purposes of the current study. Nevertheless, in order to reduce the computational effort as much as possible, a simple *HRTF Encoding Scheme* was applied. According to this scheme the FFT, of predefined block length and sampling frequency, was taken for each pair of HRTFs and the complex coefficients were stored in a file, instead of the time domain coefficients.

Concerning the *Doppler Effect*, it is well-known that it refers to the change in frequency that occurs when there is relative motion between a wave - emitting source and an observer. In the auditory domain a sound source traveling at a constant velocity past a stationary observer will drop in observed frequency both as it approaches and departs. Here, the term *observed frequency* refers to the physical frequency of the sound, measured at the point of observation. Because no rise in frequency actually occurs, this perception is called the *Doppler Illusion* [8]. The motion parameters & decision module is also responsible for enabling / disabling the Doppler Illusion Simulation, that takes place in the time – varying convolution engine described in the following paragraph.

3.2. Time-Varying Convolution Engine

For real - time filtering of FIR filters, block convolution methods operating in frequency are a necessity. FFT based *Overlap - Add Method* for block convolution [29] is the basis of the processing for the analysis that follows. Figure 3 depicts the structure of a typical Overlap - Add FFT based block convolution.

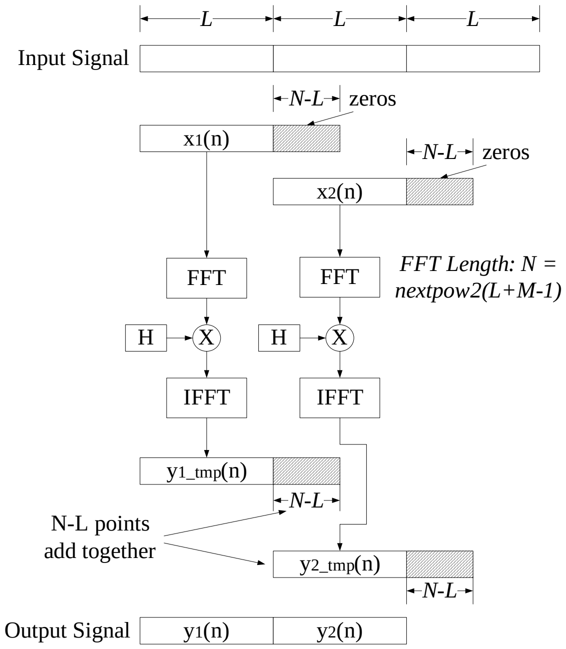

Figure 3 Typical Overlap - Add FFT block convolution

This kind of processing becomes inefficient when the filter coefficients change in successive blocks of input data. Figure 4 illustrates this problem. In this case, 2048 samples of a 120Hz sinewave are filtered in two successive blocks of 1024 samples. The first 1024 samples were convolved with the (left) HRTF at position A (0°) and the next 1024 are convolved with the (left) HRTF at position B (5°). The length of the HRTF filters is 512 samples at a sampling frequency of 44.1 kHz. Clearly, within the time interval that correspond to the samples between 1024 and 1535 (=1024+512-1), a waveform discontinuity occurs, which results into strong audible distortion.

This discontinuity can be resolved by applying output cross – fading processing, as illustrated in Figure 5. In this case, the input signal is typically processed twice; the results are multiplied with a fade-in or fade-out function and added together. However, as mentioned earlier, simple output cross-fading, is computationally burdensome, as two different systems must process in parallel.

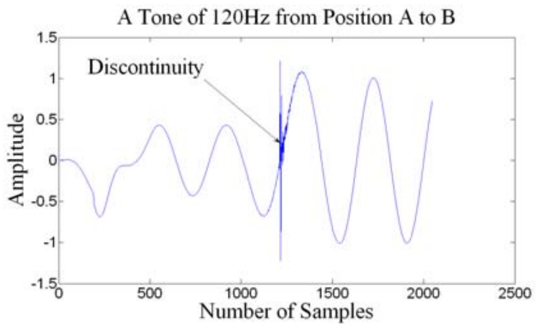

Figure 4 Discontinuity generated when a 120Hz tone is convolved with the left HRTF of 0° and 5°.

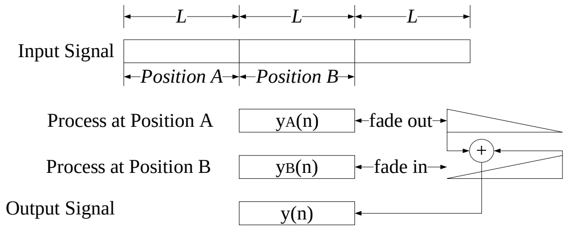

Figure 5 Simple Cross-Fading processing

In an initial effort towards the minimization of the computational complexity, the *Compact Cross – Fading* technique based on FFT overlap – add convolution was designed by the authors during this work. The basic functionality of this technique is described in Figure 6. The functional structure of the Compact Cross – Fading method is based on a combination of the overlap-add FFT convolution and the output cross – fading systems that were described previously. As far as computational effort is concerned, there is practically no gain. However, the benefit is that a single module can be considered as a time – varying convolution engine.

With these considerations in mind combined with the fact that Doppler Illusion simulation should somehow be encapsulated in the time – varying convolution engine, the attention of the authors was redirected to well – known spectral audio signal processing algorithms based on Short - Time Fourier Transform (STFT). It is well - known that the STFT is a powerful general - purpose tool for audio signal processing. It defines a particularly useful class of time - frequency distributions which specify complex amplitude versus time and frequency for any signal [26].

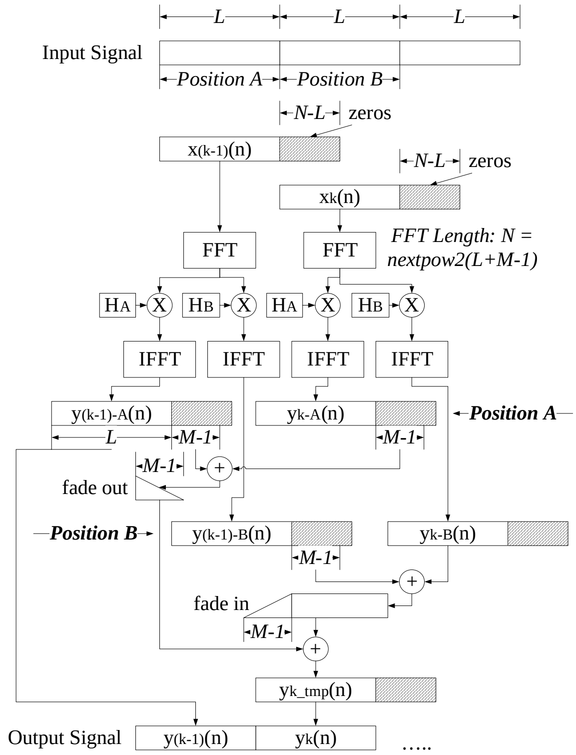

Figure 6 Architecture of the proposed Compact Cross-Fading

A very interesting (and intuitive) way of modifying a sound is to make a two-dimensional representation of it, modify this representation in some or other way and reconstruct a new signal from this representation. Consequently a digital audio effect based on TFR requires three steps: (a) an analysis, (b) a transformation and (c) a re-synthesis. The analysis / synthesis scheme is termed as *phase vocoder* [30].

The usual mathematical definition of STFT for discrete time signals is:

$$X_m(\omega) = \sum_{n=-\infty}^{\infty} x_m(n) = \sum_{n=-\infty}^{\infty} x(n)w(n-mR)e^{-j\omega n} \quad (1)$$

where $x(n)$ is the input signal time representation, and $w(n)$ is the window function of length M.

Additionally, $X_m(\omega)$ denotes the Discrete Time Fourier Transform (DTFT) of windowed data centered about time mR and R corresponds to the hop size, in samples, between successive windows.

If the window $w(n)$ has the *Constant Overlap - Add* (COLA) property at hop size R, that is:

$$\sum_{m=-\infty}^{\infty} w(n-mR) = 1, \forall n \in Z, (w \in COLA(R)) \qquad (2)$$

then according to [26], the sum of the successive DTFTs over time equals the DTFT of the initial signal, that is

$$\sum_{m=-\infty}^{\infty} X_m(\omega) = X(\omega) = DTFT_{\omega}(x) \qquad (3)$$

For windows that satisfy the COLA property, it is possible to obtain an *Overlap – Add STFT Processor* similar to the typical Overlap – Add FFT processor (illustrated in Figure 3). This is accomplished by multiplying each spectral frame $X_m(\omega)$ by some filter frequency response $H_m(\omega)$. One should note that H_m can be different for each frame, providing a certain class of time - varying filters. Additional spectral modifications may be applied for the simulation of the Doppler Illusion, by legacy pitch shifting [30]. The above discussion and mathematical relationships are synopsized in Figure 7, where the proposed architecture for binaural simulation of moving sound sources is depicted.

Practical implementations of the above relationships use the Fast – Fourier Transform (FFT), which may be regarded as a sampled DTFT. A very good overview of the "tips & tricks" for a correct practical implementation of the above structure are given in [31]. Briefly, the key – points, among others, have to do with the choice of the framing window and hop size, the correct window periodicity, the FFT centering and the zero – padding of frames.

Moreover, it is important to note, that with this structure, the computational effort is the same for either static sound sources or moving sound sources. For moving sound sources a small overhead is only introduced because of the calculation of the FFT of the filters that correspond to the new position. However, this calculation overhead can be eliminated using the HRTF Encoding Scheme that was described previously in Section 3.1.

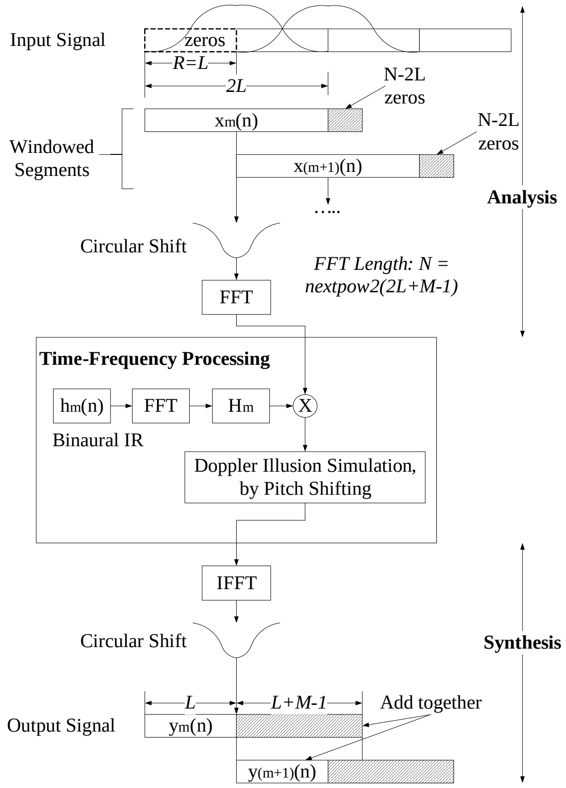

Figure 7 Architecture of STFT signal processor for the binaural simulation of moving sound sources

Additionally, for Doppler Illusion representation, a simulation method provided in [17] was employed during this work. Using this method, the scaling of the frequency measured at the sound source to that measured at the receiver position can be approximated by the function:

$$g(t) = 1 - \frac{\bar{x}x(t) + \bar{y}y(t) + \bar{z}z(t)}{cr(t)} \qquad (4)$$

where

$$r(t) = \sqrt{x^2(t) + y^2(t) + z^2(t)} \qquad (5)$$

and $\bar{x}$, $\bar{y}$, $\bar{z}$ are the vector elements of the sound source velocity with respect to the listener, $x(t)$, $y(t)$ and $z(t)$ are the positions of the sound source with respect to the listener, and c is speed of sound in air. The Doppler Effect measured at the receiver position can be viewed as an instantaneous frequency modulation imposed on each frequency component of the sound source, such that $f_i^R(t) = f_i^S g(t)$, for the *i-th* frequency component of the signal. Thus, a general time dependent signal can be defined:

$$s_i(t) = A_i \cos[\theta_i(t)] \tag{6}$$

where $\theta_i(t)$ is the instantaneous phase. The instantaneous frequency at the receiver can be defined in terms of the instantaneous phase as

$$f_i^R(t) = \frac{1}{2\pi}\frac{d\theta_i(t)}{dt} \tag{7}$$

and $\theta_i(t)$ can be obtained by simple integration, that is

$$\theta_i(t) = 2\pi f_i^S \int_0^t g(\lambda)d\lambda \tag{8}$$

Appropriate substitution and evaluation of the above integral yields to a remarkably simple instantaneous phase generation function for Doppler (after removal of a constant phase term):

$$\theta_i(t) = 2\pi f_i^S (t - \frac{r(t)}{c}) \tag{9}$$

where $r(t)$ is the time – varying range.

4. RESULTS

The proposed sound source motion simulation algorithms was implemented as an extension of the Amphiotik 3D Audio Engine [11] developed by the authors and it was evaluated using the Amphiotik Synthesis application [32]. The latter software application provides a robust and user friendly user interface for defining source and listener positions as well as binaural simulation parameters in real-time.

In order to evaluate the performance of the proposed moving sound source simulation technique using objective criteria, a white noise test signal of 5.94 seconds duration was generated and spatialized with the following parameters:

- Block length 2048 samples.
- Sampling frequency 44.1 kHz.
- Initial source position: 2 meters away from the listener at 0 degrees azimuth and 0 degrees elevation.

Finally, the sound source motion was defined as a circle trajectory surrounding the listener, in the horizontal plane, with a speed of 60 degrees / second. With this speed and test signal duration a full circle is covered.

The binaural output from the sound motion simulation was stored in a file and it was further analyzed in order to acquire the Head - Related Spectrograms (HRS) of the left and right channels. As defined in [33], the HRS encompasses the modification imposed on a sound signal between its source transcribing a particular trajectory and its detection at the ears. Hence, they represent the HRTF variation smoothness during the sound source movement.

Figure 8 depicts the HRS obtained using the STFT sound source movement simulation method. For comparison reasons, the corresponding HRS obtained using the Compact Cross – Fading algorithm, are also shown in Figure 9.

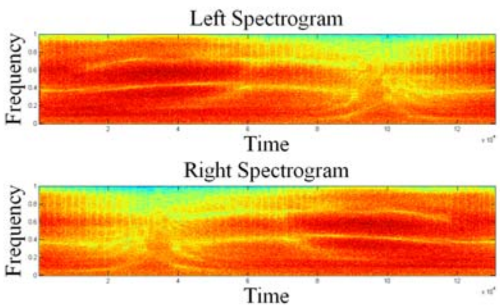

Figure 8 Measured HRS using STFT

It should be noted that the darker areas correspond to more signal power than the lighter areas. The spectrograms of both methods show a smooth transition between the various positions. The only noticeable difference is that for the STFT method, and when the power is low for high frequencies (above approximately 10 kHz), small fluctuations occur for these frequencies. The reasons of this observation are currently investigated by the authors. However, a number of perceptual tests performed using trained audience has shown that these fluctuations are not audible.

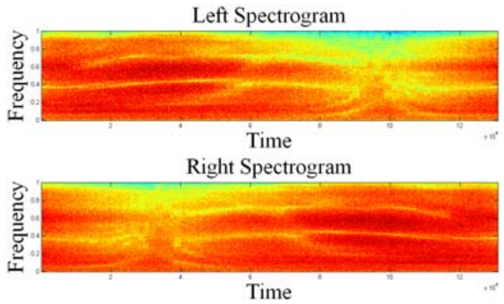

Figure 9 Measured HRS using Compact Cross Fading

5. CONCLUSIONS

The problem of simulating moving sound sources in real - time is mainly originating from the high computational load that legacy simulation techniques impose. In this work a novel frequency - based, time - varying filtering method is introduced that minimizes the need of additional computations for accurate moving sound - sources representation.

The performance of the proposed time - varying HRTF frequency - based method was objectively evaluated by Head - Related Spectrograms (HRS) that showed that the sound source movement is in terms of HRS very smooth, while the spatial performance achieved is very high. Moreover, compared to other moving sound sources simulation methods, the proposed technique does not introduce any additional computational load, especially when the available HRTF data are available as complex representations in the frequency domain.

Additionally, a number of subjective listening tests has verified the above results. More specifically, from these tests it was found that the sound source movement is perceived very smoothed, rendering the proposed method suitable for high - quality virtual reality and multimedia applications.

6. REFERENCES

[1] M. Davis, "The AC-3 Multichannel Coder", presented at the AES 95th Convention, AES, New York, October 1993, (preprint 3774).

[2] K. Brandenburg and M. Bosi, "ISO/IEC MPEG-2 Advanced Audio Coding: Overview and Applications"," presented at the AES 103rd Convention, New York, September 1997, (preprint 4641).

[3] A. J. Berkhout, P. Vogel, and D. de Vries, “Use of wave field synthesis for natural reinforced sound”, presented at the AES 92nd Convention, 1992, (preprint 3299).

[4] M. A. Gerzon, „Ambisonics in multichannel broadcasting and video”, J. Audio Eng. Soc., Vol. 33, No. 11, Nov. 1985, pp. 859–871.

[5] AES Technical Staff Writer, “Binaural Technology for Mobile Applications”, J. Audio Eng. Soc., Vol. 54, No. 10, Oct. 2006, pp. 990 – 995.

[6] J. Blauert, “Spatial Hearing” (revised edition), The MIT Press, Cambridge, Massachusetts, 1997.

[7] K. Hartung, J. Braasch and S. Sterbing, "Comparison of Different Methods for the Interpolation of Head-Related Transfer Functions", AES 16th Int. Conf. on Spatial Sound Reproduction, 1999, pp. 319-329.

[8] J. G. Neuhoff and M. K. McBeath, "The Doppler Illusion: The Influence of Dynamic Intensity Change on Perceived Pitch", Journal of Experimental Psychology, Human Perception and Performance, Vol. 2, No. 4, 1996, pp. 970-985.

[9] R. Litovsky, S. Colburn, W. Yost and S. Guzman, “The precedence effect”, J. Acoust. Soc. Am., Vol. 106, 1999, pp. 1633-1654.

[10] J. M. Jot, S. Wardle and V. Larcher, “Approaches to Binaural Synthesis” presented at the AES 105th Convention, California, Sep. 1998 (preprint 4861).

[11] Ch. Tsakostas and A. Floros, “Optimized Binaural Modeling for Immersive Audio Applications”, presented at the AES 122nd Convention, Vienna, May 2007 (preprint 7100).

[12] E. M. Wenzel, J. D. Miller, J.S. and Abel, "Sound Lab: A real-time, software-based system for the study of spatial hearing", presented at the AES 108th Convention, Paris, 2000 (preprint 5140).

[13] K. Tsujino, W. Kobayashi, T. Onoye and Y. Nakamura, "Efficient 3D Sound Movement with Time Varying IIR Filters", Special Section on

Multimedia and Mobile Signal Processing, IEICE Trans. Fundamentals, Vol.E90–A, No. 3, March 2007.

[14] J. C. Sarris, S. A. Dalianis, and Ge. E. Cambourakis, “Simulation of Sound Source Motion by time-frequency filtering”, in Proc. of the 6th Int. Conference on Digital Audio Effects (DAFX-03), London, UK, September 8-11, 2003.

[15] C. I. Cheng and G. H. Wakefield, "Moving Sound Source Synthesis for Binaural Electroacoustic Music Using Interpolated Head-Related Transfer Functions (HRTFs)" Computer Music Journal, Volume 25, issue 4 (Winter 2001), pp. 57-80.

[16] M. Matsumoto, S. Yamanaka, M. Toyama and H. Nomura, "Moving sound image representation method", in Proc. of the EUSIPCO'2002 Conference, 2002.

[17] R. L. Jenison, M. F. Neelon, R. A. Reale, and J. F. Brugge, "Synthesis of Virtual Motion in 3D Auditory Space", In Proc. of the 20th Annual International Conference of the IEEE, Vol. 3, Nov. 1998, pp. 1096 - 1100.

[18] A. W. Bronkhorst, "Localization of real and virtual sources", Journal of the Acoustical Society of America, Vol. 98, 1995, pp. 2542-2553.

[19] D.R. Begault, "3D Sound for Virtual Reality and Multimedia", Academic Press, Cambridge, 1994.

[20] R. Jenison, K. Fissel, "A Spherical Basis Function Neural Network for Modeling Auditory Space", Neural Computation, Vol. 8, 1996, pp. 115-128.

[21] J. Chen, B. D. van Veen and K. E. Hecox, "A Spatial Feature Extraction and Regularization Model for the Head-Related Transfer Function", J. Acoust. Soc. Am., Vol. 97, No. 1, 1995, pp 439-452.

[22] R. Jenison, "A Spherical Basis Function Neural Network for Pole-Zero Modeling of Head-Related Transfer Functions", in Proc. of IEEE Workshop in Applications of Signal Processing to Audio and Acoustics, New York, 1995.

[23] P. R. Runkle, M. A. Blommer and G. H. Wakefield, "Comparison of Head-Related Transfer Interpolation Methods", in Proc. of IEEE Workshop in Applications of Signal Processing to Audio and Acoustics, New York, 1995.

[24] V. Larcher and J. Jot, "Techniques d’Interpolation de Filters Audio – Numeriques", Actes du 4eme Congres Francais d’Acoustique, Vol. 1, April 14-18, 1997, pp 97-100.

[25] CIAIR, Nagoya University, HRTF database: http://www.itakura.nuee.nagoya-u.ac.jp/HRTF

[26] J. O. Smith, “Spectral Audio Signal Processing”, on-line book, http://ccrma.stanford.edu/~jos/sasp/

[27] A. Scharine and T. R. Letowski, "Factors Affecting Auditory Localization and Situational Awareness in the Urban Battlefield", http://handle.dtic.mil/100.2/ADA431963

[28] D. R. Perrott and J. Tucker, "Minimum audible movement angle as a function of signal frequency and the velocity of the source", Journal of the Acoustical Society of America, Vol. 83, 1988, pp. 1522-1527.

[29] J. Proakis and D. Manolakis, "Digital Signal Processing: Principals, Algorithms and Applications", Macmillan Publishing Company, New York, 1992.

[30] U. Zoelzer, "DAFX – Digital Audio Effects", John Wiley & Sons LTD, ISBN 0 471 49078 4

[31] A. Götzen, N. Bernardini and D. Arfib, "Traditional (?) Implementations of a Phase – Vocoder: The Tricks of the Trade", In Proc. of the COST G-6 Conference on Digital Audio Effects (DAFX-00), Verona, Italy, December 7-9, 2000.

[32] Ch. Tsakostas, A. Floros, and Y. Deligiannis, “Real-time Spatial Mixing Using Binaural Processing”, In Proc. of the 4th Sound and Music Computing Conference (SMC ‘07), 11 – 13 July 2007, Lefkada, Greece, pp. 291 - 294.

[33] M. J. Evans, “Synthesizing Moving Sounds”, In Proc. of the ICAD Conference, 1998.