SYNTHESIS OF VIRTUAL MOTION IN 3D AUDITORY SPACE

Rick L. Jenison*, Michael F. Neelon*, Richard A. Reale**, and John F. Brugge**
* Dept. of Psychology, University of Wisconsin, Madison, WI 53705
** Dept. of Physiology and Waisman Center, University of Wisconsin, Madison, WI 53705
*E-mail: jenison@wavelet.psych.wisc.edu

Abstract - This paper introduces a novel approach to the synthesis of acoustic waveforms that mimic those present near the eardrums of a listener when a sound source is moving through three-dimensional (3D) space. The approach combines well-known physical and biophysical sound-location cues into a discrete synthesis system. Physical cues include both the dependence of intensity on distance from the sound source and the *Doppler* effect. Biophysical features include the direction-dependent transfer functions imposed by a listener's head and ears and interaural time differences. This technique requires very little computer-disk storage since it relies on low-dimensional functional models that relate cue values to arbitrary locations in a virtual acoustic space.

I. INTRODUCTION

Virtual auditory displays that are employed in both experimental and commercial applications [1] require that for a source location in 3D virtual space the sound field generated at the eardrums of a listener be identical to the sound field generated from the same location in the free field. Prior work has employed static, impoverished or discontinuous virtual sound sources for neurophysiological and psychophysical studies of spatial hearing [2][4][15]. In this paper, a method is described that affords veridical simulation of the complicated acoustic waveforms that result from a sound source in continuous motion. Controlled movement of a sound source in the free field is difficult to achieve and is often severely limited by physical constraints. Virtual environments, by comparison, offer unlimited flexibility in the design and effects of movements.

The synthesis of the receiver characteristics for auditory motion may be divided into two parts. In the first part a sufficient set of theoretical physical laws governing the acoustics of a single sound source and a binaural receiver are employed to convey information for motion [12]. This part requires that at the receiver, the intensity of the sound varies as a function of distance, and that the frequency and interaural phase of the sound vary as a function of relative velocity with respect to the receiver [12]. The second part is based on a functional approximation to the direction-dependent spectral filtering introduced by the ear and head, which is analytically captured by the head related transfer function (HRTF). Figure 1 illustrates a trajectory of a sound source moving along a linear path in virtual auditory space. The analysis and synthesis of kinematic acoustics is most efficient using cartesian coordinates [12], where the y-axis is aligned with the listener's nose. On the other hand, functional approximation of HRTFs is optimal when performed on spherical coordinates. Presently, the sound source is treated as a omni-directional, monopole point source, however future work will entail multipole sound sources.

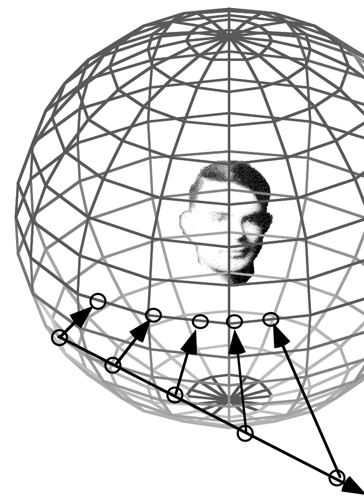

Fig. 1. Schematized sound source path relative to the spherical projection surface of a binaural receiver (listener).

II. SYNTHESIS METHODS

A. Modulation of Frequency (FM) and Amplitude (AM)

When a moving sound source has a component of its relative velocity along a line connecting it to a receiver, the frequency at which the receiver encounters radiated waves from the source depends upon the direction of the relative movements (i.e. toward or away from each other). This observation is referred to as the *Doppler* effect [14]. The scaling of the frequency measured at the source to that measured at the receiver(listener) can be approximated by the function

$$g(t) = 1 - \frac{\dot{x}\, x(t) + \dot{y}\, y(t) + \dot{z}\, z(t)}{c\, r(t)} \qquad (1)$$

where

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

$\dot{x}$, $\dot{y}$, and $\dot{z}$ are the vector elements of the sound source velocity with respect to 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 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 *ith* frequency component of the signal. We can define a general time-dependent signal

$$s_i(t) = A_i \cos\left[\theta_i(t)\right] \quad (3)$$

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

$$f_i^R(t) = \frac{1}{2\pi} \frac{d\theta_i(t)}{dt} \quad (4)$$

and $\theta_i(t)$ can be obtained by integration

$$\theta_i(t) = 2\pi f_i^S \int_0^t g(\lambda)\, d\lambda. \quad (5)$$

Substituting (1) into (5) and evaluating the integral yields a remarkably simple instantaneous phase generation function for Doppler (after removal of a constant phase term)

$$\theta_i(t) = 2\pi f_i^S \left(t - \frac{r(t)}{c}\right) \quad (6)$$

where r(t) is the time-varying range.

In an anechoic environment the measured sound intensity follows an inverse-square law, where the intensity varies inversely as the square of the range r(t) between the receiver and source. Pressure, therefore, varies as the inverse of r(t), such that

$$\alpha(t) = \frac{\eta}{r(t)} \quad (7)$$

where η is a constant related to intrinsic power.

B. Modulation of Interaural Phase

Interaural Time Difference (ITD) is the difference between the time of arrival of a sound at each ear caused by the reflection and diffraction of the head. In general, ITD depends on both frequency and the angle of incidence, although there are several approximations that simplify these dependencies [13]. In the present simulations, the ITD is defined by

$$\delta(t) = \frac{2a}{c} \frac{x(t)}{r(t)} \quad (8)$$

where *a* is the radius of the average head (8.75 cm), x(t) is the coordinate on the x-axis, and r(t) is again the time-varying range from the receiver to the source. Frequency dependent interaural phase can then be defined as

$$\phi_i(t) = 2\pi f_i^S g(t)\, \delta(t) \quad (9)$$

C. Static Sound Source Analysis and Motion Synthesis

There are several possible transforms that could be used to decompose a static sound source into sinusoidal bases. The obvious choice is the Fourier Transform. A window of a sampled static sound of size N is selected and the Fourier coefficients A_i and B_i are extracted. Fourier summation is then used to synthesis each of the binaural signals

$$s_L(t) = \alpha(t) \sum_{i=1}^{N} \left[A_i \cos\left(\theta_i(t) - \frac{\phi_i(t)}{2}\right) + B_i \sin\left(\theta_i(t) - \frac{\phi_i(t)}{2}\right)\right] \quad (10)$$

and

$$s_R(t) = \alpha(t) \sum_{i=1}^{N} \left[A_i \cos\left(\theta_i(t) + \frac{\phi_i(t)}{2}\right) + B_i \sin\left(\theta_i(t) + \frac{\phi_i(t)}{2}\right)\right] \quad (11)$$

D. Head-Related Transfer Function (HRTF)

The HRTF expresses, for a given sound-source direction and over a specified range of frequency, the transformation of sound pressure in the free field to the pressure measured near the eardrum. The present simulations are based on a regularly spaced set of HRTFs measured in the left and right ears [16][17][18]. In order to reduce the dimensionality of sampled frequencies (which are often on the order of several hundred), a principal component analysis is applied to the measured set of HRTFs [17]. This linear transformation typically reduces the dimensionality by at least an order of magnitude. In order to provide interpolation for directions not measured in the original set, a spherical approximation technique is employed. The von Mises basis function (VMBF) network depends on spherical, rather than Cartesian input coordinates, and is a natural approach to problems on the sphere [8][10][11]. The expression for the von Mises basis function is given by

$$\begin{aligned} &\mathrm{VMBF}\left(az(t), el(t); \gamma, \beta, \kappa\right) = \\ &\qquad \exp\left\{\kappa\left[\sin el(t) \sin\beta \cos\left(az(t)-\gamma\right) + \cos el(t) \cos\beta\right]\right\} \end{aligned} \quad (12)$$

where the input parameters correspond to a sample location in azimuth and elevation, a centroid in azimuth and elevation (γ, β) and the concentration parameter (κ). Estimated HRTFs are computed as a linear summation of an appropriate number of weighted VMBFs

$$\widehat{\mathrm{HRTF}}\left(az(t), el(t)\right) = \sum_{j=1}^{J} \mathbf{w}_j \, \mathrm{VMBF}\left(az(t), el(t); \gamma_j, \beta_j, \kappa_j\right) \quad (13)$$

where $\mathrm{VMBF}\left(az(t), el(t); \gamma_j, \beta_j, \kappa_j\right)$ is the output of the *j*th von Mises basis function and $\mathbf{w}_j$ is its *j*th weight vector. The principal components of the measured set of HRTFs serve as the training vector, and the VMBF parameters are learned through gradient descent methods. Mapping from cartesian to spherical coordinates, where azimuth is measured with respect to the positive y-axis, is accomplished by

$$az(t) = \tan^{-1}\left(\frac{x(t)}{y(t)}\right) \quad (14)$$

and

$$el(t) = \tan^{-1}\left(\frac{z(t)}{\sqrt{x(t)^2 + y(t)^2}}\right) \quad (15)$$

The synthesized binaural signals are currently convolved with time-varying HRTFs based on the spherical projections illustrated in Fig. 1 at discrete time samples.

E. Earphone Delivery

Successful simulation of the receiver signals corresponding to sound-source movement ultimately depends on the fidelity of their delivery through a headphone sound delivery system. In experimental work, we employ a specially designed inset earphone that incorporates a probe microphone for sound measurements near the eardrum [3]. Least-squares FIR filters are used to compensate for the non-ideal frequency response of these transducers [15].

III. EXAMPLES

A 100 ms digitized sample of a mosquito in flight was Fourier analyzed (Fig. 2), and the Fourier coefficients were then used to synthesize 3 seconds of virtual auditory motion. The spectrogram from the resulting Fourier summation is shown in Fig. 3 illustrating the Doppler shifted harmonics and change in intensity. The closest point of approach occurs at 1.5 s following the signal onset, which corresponds to the point where the sound intensity and the derivative of the Doppler function are at their maximum.

IV. DISCUSSION

The synthesis of sounds in both experimental and commercial virtual 3D displays will demand simulations of

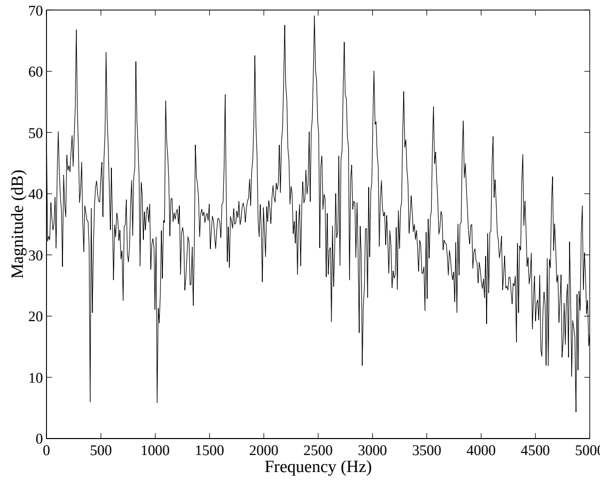

Fig. 2. Snapshop magnitude spectrum of a mosquito in flight.

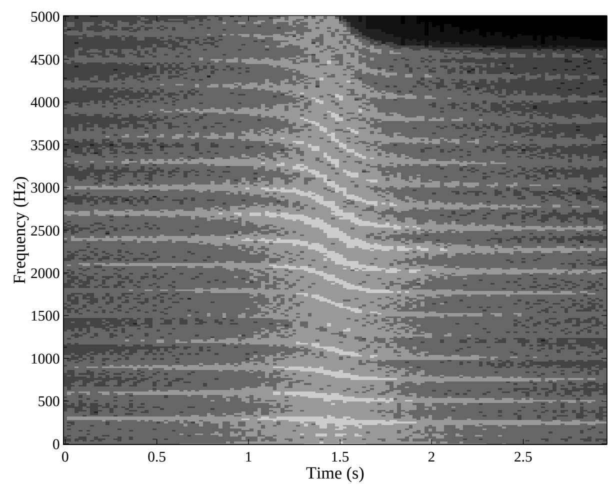

Fig. 3. Spectrogram of synthesized mosquito traveling along a linear path in virtual auditory space. The brightness of the image is proportional to the log intensity.

moving sound sources. A procedure that combines major physical and biophysical cues of sound-source location into a virtual space paradigm that can accommodate moving sound sources has been described. A schematic summarizing the steps required for this synthesis is shown in Fig. 4. The application of frequency modulation for synthesizing complex audio spectra is well-known in the computer music literature [6]. Analytical solutions for instantaneous frequency integrals have generally been restricted to simple sinusoidal FM, however we have extended this approach to include Doppler FM. We have employed these techniques in the past for psychophysical studies of auditory motion perception, specifically for the task of estimating object time-to-arrival [9]. The ability to analyze environmental sounds and synthesize controlled paths of motion have important implications for the study of ecological perception. For example, an interesting characteristic of winged insect sound radiation is that the source behaves as an acoustic dipole radiator [7]. Thus, the directivity pattern is no longer omnidirectional. As a dipole sound source moves relative to a receiver, it induces a characteristic amplitude modulation that affords information about the flying insect. The acoustic signatures of natural moving sound sources has largely been ignored, and represents another aspect of virtual auditory motion to be investigated. Other signal analysis methods are also under consideration, including the Discrete Cosine Transform and Wavelet Analysis. These alternative methods have certain advantages over the Fourier Transform in terms of data compression efficiency.

ACKNOWLEDGMENT

This work was supported by NIH Grants, DC00116, DC03352, and DC02804.

REFERENCES

[1] D. R. Begault, *3-D Sound For Virtual Reality and Multimedia* , Academic Press, London, pp. 39-76, 1994.

[2] J.F. Brugge, R.A. Reale, J.E. Hind, J.C. Chan, A.D. Musicant, and P. W. Poon, "Simulation of Free-Field Sound Sources and its Application to Studies of Cortical Mechanisms of Sound Localization in the Cat," *Hear. Res.*, vol. 73, pp. 67-84, 1994.

[3] J. C. Chan, A.D. Musicant, and J.E. Hind, "An Insert Earphone System for Delivery of Spectrally Shaped Signals for Physiological Studies," *J. Acoust. Soc. Am.*, vol. 93, pp. 1496-1501, 1993.

[4] J. C. Chen, B. VanVeen, and K. E. Hecox, "Synthesis of 3D virtual auditory space via a spatial feature extraction and Regularization Model," in *IEEE Virtual Reality Annu. Int. Symp.*, pp. 188-193, 1993.

[5] J. C. Chen, Z. Wu, and R. A. Reale, "Applications of Least-Squares FIR Filters to Virtual Acoustic Space," *Hear. Res.*, vol. 80, 153-166, 1994.

[6] J. M. Chowning, "The synthesis of complex audio spectra by means of frequency modulation," *J. Aud. Engr. Soc*., vol. 21, 1973.

[7] N. F. Fletcher, *Acoustic Systems in Biology*, Oxford Press, Oxford, 1992, pp. 119.

[8] R. L. Jenison, "Von Mises Basis Function Neural Networks for Spherical Approximation," in *Proceedings of the World Congress on Neural Networks* , Lawrence Erlbaum Associates, Washington, DC, 1995, pp. 610-614.

[9] R. L. Jenison, "Estimation of time-to-arrival from acoustic information. *Acoustica*, vol. 82, pp. 96, 1996.

[10] R. L. Jenison, "A Spherical Basis Function Neural Network for approximating acoustic scatter," *J. Acoust. Soc. Am.*, vol. 99, pp. 3242-324, 1996.

[11] R. L. Jenison and K. Fissell "A Spherical Basis Function Neural Network for Modeling Auditory Space," *Neural Comp*., vol. 8, pp.115-128.,1996.

[12] R. L. Jenison, "On Acoustic Information for Motion," *Ecol. Psych.*, vol. 9, pp. 131-151, 1997.

[13] G. F. Kuhn, "Model for the Interaural Time Differences in the Horizontal Plane," *J. Acoust. Soc. Am.* , vol. 29, 988-991, 1977.

[14] P. M. Morse and K. U. Ingard, *Theoretical Acoustics*, Princeton University Press, Princeton, 1986.

[15] R. A. Reale, J. Chen, J. E. Hind, J. F. Brugge, "An Implementation of Virtual Acoustic Space for Neurophysiological Studies of Directional Hearing," in *Virtual Auditory Space: Generation and Applications*, edited by S. Carlile, R. G. Landes, Austin, TX, 1996, pp. 153-170.

[16] F. L. Wightman and D. J. Ki,stler *Sound Localization*, in *Human Psychophysics* , edited by W. Yost, A. Popper and R. Fay, Springer-Verlag, Berlin, 1993, pp. 155-193.

[17] F. L. Wightman. and D. J. Kistler, "Headphone Simulation of Free-Field Listening. I: Stimulus Synthesis," *J. Acoust. Soc. Am.*, vol. 85, pp. 858-867, 1989.

[18] F. L. Wightman and R. L. Jenison, *Auditory spatial layout*, in *Perception of Space and Motion.*, edited by W. Epstien and S. Rogers, Academic Press, 1995, pp. 365-400.

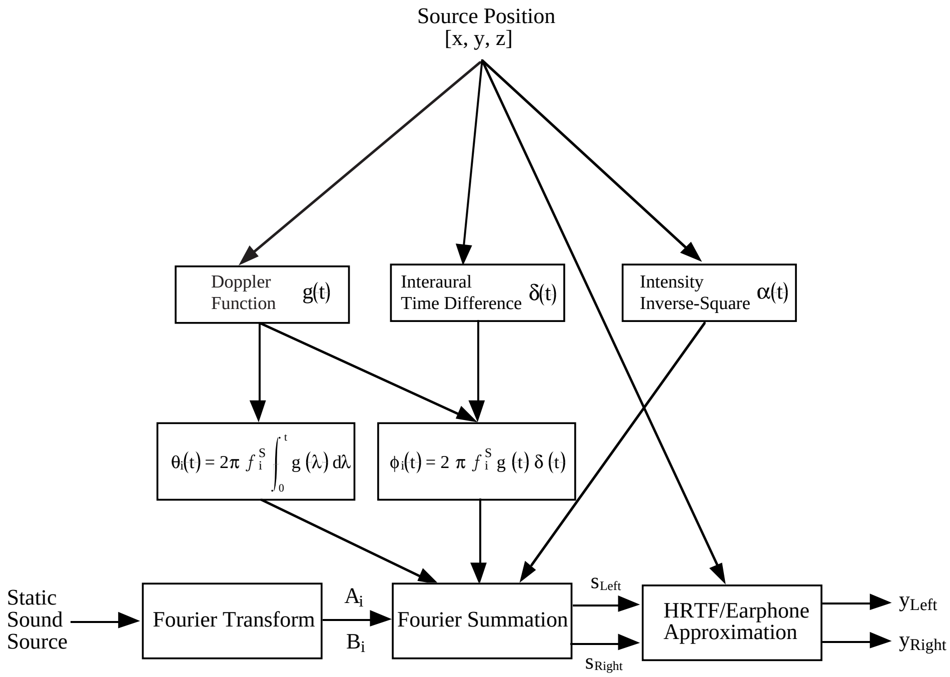

Fig.4. Schematic of virtual 3D auditory motion synthesis.