

Synthesising Moving Sounds

Michael J. Evans

School of Computer Science, Cybernetics and Electronic Engineering, University of Reading
Reading, United Kingdom

Abstract

Auditory display designers are making increasingly effective and creative use of our ability to localise sound; to particular auditory events as occurring at particular locations. Many applications in which spatial audio has been applied could also benefit from exploiting another important ability of the auditory system; the detection and identification of sound source motion. The display of moving sources could improve usability, provide additional variables in sonification, make virtual environments more perceptually realistic and provide new creative possibilities for designers. Transaural cancellation allows the creation of spatial audio with just two loudspeakers. These techniques are now extended to create the illusion of a sound source moving along an arbitrary trajectory at an arbitrary rate.

This paper discusses the application of synthesised sound source movement to a number of practical applications in auditory display. We seek to extend the use of Head-Related Transfer Functions (HRTFs) in stationary sound spatialisation to encompass movement synthesis. The detection of moving sources is not time-invariant so we propose and demonstrate the use of time-frequency spectrograms as a mechanism for characterising source movement. There are an infinite number of such trajectory-related spectrograms and we address the need for a continuous directional model to accommodate this.

1 Introduction

In seeking to maximise the information that sound can convey, auditory display designers are making increasingly effective and creative use of the human ability to localise sounds. Despite a modest sensory stimulus, the human auditory system is able to extract a great deal of information from the varying sound pressure levels at the eardrums. As Bregman [1] has demonstrated, our acuity is particularly highlighted by our spatial perception of sound; akin to deducing the movements of boats on a lake merely by observing the perturbations of a pair of handkerchiefs pinned at one end of the lake. In recent years, applications of spatial audio have become more widespread, generally using loudspeakers or headphones to convey localisation information and thus exploit our spatial acuity [2]. The inclusion of sound spatialisation in an auditory display can

- use spatially distinct presentation to increase usability in interfaces
- alleviate ambiguity in teleconference and videophone systems by increasing discrimination between participants
- map a region of sound onto a set of data thus providing an additional independent variable for sonification
- help render more perceptually realistic virtual environments
- provide attractive new creative possibilities for producers of music, film and interactive entertainment

However, Bregman's lake analogy does not merely concern the identification of the static positions of boats and other events in the lake. Bregman used the analogy of the motion of the boats on the river to illustrate our ability to perceive the *movement* of auditory events from binaural information. Since we are able to identify the spatial motion of sounds relative to our position then the artificial creation of moving sources of sound could also be beneficial in the applications listed above. An important additional application of movement synthesis could also be found in systems which employ head position tracking in order to render spatialisation tolerant of changes in listener position [3], since the relative position of source relative to the listener is generally in continuous change. Spatial audio reproduction aims to render a virtual auditory event by ensuring that the ears of the listener receive the acoustic cues which would result from a real source at the location of interest [2]. Extending this paradigm allows us to note that, as a sound source changes its azimuth or elevation relative to a listener, some cues in the signals received at the ears during this motion allow the listener to form an estimate of the trajectory. (In general, the trajectory might also include variation in source distance, which would be cued by sound level variation, as well as frequency shift as predicted by Doppler.)

In terms of creating the perception of a sound source changing its azimuth and elevation, it seems evident that synthesising sound movement is a tractable problem. Notionally, we are seeking a technique which emulates the concept of a loudspeaker moving along the rails of a spherical cage which surrounds the listener. In terms of a wavefront reconstruction approach, borrowing from such spatialisation techniques as ambisonics [4], we could surround our listener with an array of loudspeakers and progressively mix a sound signal across a set of them to generate the desired trajectory. However, as in static sound spatialisation making use of localisation cues allows spatial illusions to be rendered by means of perceptual synthesis, and requiring just two channels of reproduction for a single listener.

In this paper we seek to extend the two channel synthesis of spatial audio, as exemplified by binaural reproduction using headphones and pairs of Head-Related Transfer Functions (HRTFs) [2], and transaural techniques using a pair of frontally placed loudspeakers, HRTFs and the necessary processing to ensure that left-right crosstalk is cancelled [5]. The potential benefits of such a solution are similar to those already encountered in static spatialisation; the cost and bandwidth economies of only requiring two channels of transmission or recording, rather than the multiple channels of ambisonics and wavefront reconstruction, and the logistic flexibility afforded by not requiring rear loudspeakers. This makes this approach particularly well suited to single-user applications using a head mounted display, or for desktop applications using conventional monitor-side loudspeakers [6]. In the remaining sections of this paper we will begin to develop a new framework for representing and implementing movement synthesis using two channels, by the extension of HRTF techniques. We examine the psychoacoustics literature with particular reference to sensory mechanisms for detecting auditory movement, considering both movement detection and inferring movement from a change in perceived location. Time-frequency spectrograms are introduced as an effective method of characterising sound source movements and the problems of spatial sampling in practical HRTF measurements sets is illustrated. A solution is presented based on a continuous functional representation of the amplitude of particular frequency components reaching the ear.

2 Perceiving Moving Sources

2.1 Minimum Audible Movement Angle

Human sound localisation ability is effective and robust. Most listeners can make accurate and reliable localisation judgements in a variety of listening environments, ranging from small rooms to open spaces, and where widely different acoustic conditions, such as background noise, exist. Nevertheless, the spatial resolution of sound localisation is not infinite. The term Minimum Audible Angle (MAA - [7]) is usually used to refer to the smallest angular separation of source direction discernible by a listener. MAA is a function of the audiological characteristics of individual listeners, as well as the scale of acoustic effects caused by listener physiognomy. Therefore, MAA varies between listeners, as a function of source direction and also as a function of frequency. In general, localisation is at its most sensitive for low frequency sources directly in front of the listener. Under these conditions MAA can be as low as 1° - 2° [7].

The MAA concept can be extended to encompass our ability to perceive source movement. Minimum Audible Movement Angle (MAMA) may be defined as the smallest net change in location of a moving stimulus that can be detected under some specified set of conditions [8]. In practical experiments MAMAs can be measured using moving sound sources and establishing the smallest angle of movement required for a subject to discriminate the direction of the source's trajectory [9]. As with the MAA of static source localisation, MAMA varies as a function of direction, listener and the spectral structure of the moving sound signal. Critically, MAMA also varies significantly depending on the rate with which a moving source traverses its trajectory, with rapidly moving sources exhibiting a generally higher MAMA than equivalent slower moving ones. Grantham notes that, typically, MAMA values are significantly (generally at least twice as large) as MAAs. Motion perception is at its most sensitive for slow moving sources directly in front of the listener, under which conditions MAMAs as low as 2° - 5° have been measured [9].

2.2 Motion Detection vs. Stationary Snapshots

Researchers examining human responses to moving visual stimuli indicate that we possess mechanisms sensitive to stimulus movement. There is considerably less evidence indicating that there are equivalent mechanisms in

audition. Our ability to identify moving sources and estimate their trajectories might arise from particular parts of the auditory system which act as motion detectors and are stimulated by moving stimuli. However, an alternative explanation might simply state the perceived location of a sound is different at one moment in time than it was at an earlier moment. The movement of the source between the two locations is then inferred by the auditory system creating the movement percept. Grantham, and others, have called this second explanation the *snapshot* theory. Movement perception based on snapshots requires no systems in human audition which are directly sensitive to motion. Trajectory perception merely used stationary localisation cues available at the start and end of a movement.

Grantham has carried out a number of detailed investigations into snapshot perception. In [10] he implements a series of listening tests in which subjects were presented with a 500Hz source moving in front of them (by means of amplitude panning between a pair of loudspeakers). A number of test methods were used but the general task for listeners was to discriminate between rapidly moving and slow moving (or completely stationary) stimuli. These results tend to indicate that the angle subtended by a moving source is a stronger movement cue than the angular velocity of movement. Although these early results seem to confirm that movement perception obeys the snapshot theory it should be noted that the source movement here is simulated in a rudimentary manner. However, further and more detailed evidence was gathered [11] for simulated movements at a wide range of angular velocities. Excepting at very slow movement rates (where the separate spatial localisations of the notional end points became disassociated from one another) the snapshot theory was generally confirmed by establishing equivalent perceptions for presentation merely of the end points as achieved for more continuous movements.

Movement perception does indeed seem possible using static localisation mechanisms alone. However, the snapshot theory must be weighed against results such as [12]. This research found that listeners were able to discriminate between moving sounds which accelerated and decelerated over time, and equivalent sounds with constant angular velocity but the same temporal and spatial end points. Such discrimination between moving sounds may indicate the existence of motion detection mechanisms in the auditory system. Similarly, [13] identified that it was possible to measure an MAMA which was (modestly) smaller than the corresponding MAA. This result alone implies the existence of a motion detection mechanism, providing the additional directional information available during perception of the moving source. In [9] Grantham has observed a similar result and proposed the existence of a combined and adaptive movement perception system, in which dedicated motion detectors do exist and provide some information regarding the dynamic behaviour of a sound source in the midpoint of its motion. This additional information can serve to enhance the estimate of motion made by static, 'snapshot' localisation of the end points of the trajectory.

2.3 Implications

The apparently conflicting experimental evidence regarding the existence of a motion sensitive mechanism exemplifies the fact that our understanding of sound movement perception is not very strong. The adaptive model proposed by Grantham is an attractive one as it accounts for the robustness of our auditory sense across a range of conditions. However, there is still no experimentally established relationship between particular physiological features in the listener and the precise acuity of that listener to movement. Similarly, the precise performance parameters of movement perception, in terms of required duration, extent of angular movement and frequency composition can not yet be attributed to particular human factors. We simply do not know which features in the sound reaching the ear form movement cues, and which are superfluous to movement and spatial perception. Neither do we know the relative salience of particular cues.

In the spatialisation of static sounds HRTFs are generally used to create the appropriate sound pressure levels at the ear canals. Although a number of localisation mechanisms are well understood [14], such as head shading and interaural delay, there is still very little research available to tie particular features within a set of HRTFs to particular percepts. Acoustically, a set of HRTFs will be over-specified and potentially inefficient, containing an array of acoustic information some of which will be crucial to localisation but also some that plays no role in localisation, or is even not perceived. Nevertheless, HRTFs are a convenient form for encoding sound localisation cues, allowing them to be imposed on a set of signals and ensure that the appropriate information, including this redundancy, reaches the ears. It is likely that effective spatial audio is possible without the full complexity of a global set of HRTFs, by exploiting particular cues and particular hearing mechanisms. Similarly, in synthesising movement, the work of Grantham and others in establishing limitations in the complexity of movement perception may lead to efficient movement synthesis by directly exploiting particular

psychoacoustic mechanisms. However, in the absence of a more detailed understanding of movement perception we seek to establish an all-encompassing and potentially overspecified analogue of HRTFs, representing the acoustic information reaching the ears of a listener as a result of a sound source transcribing an arbitrary trajectory at an arbitrary rate.

3 Approaches to Synthesis

3.1 Time-Frequency Spectrograms

Head-Related Transfer Functions (HRTFs) are generally measured in the form of a set of impulse responses, usually using approximate impulse excitation via pseudo-random sequences [15]. An HRTF impulse response, or its Fourier transformed equivalent magnitude and phase spectra, can unambiguously characterise the acoustic modification between a particular location and one of the ears. An impulse response is suitable for this representation because acoustic propagation to the ear forms a *Linear, Time-Invariant System*: The system obeys the principles of proportionality and superposition, and the input-output relationship is independent of time.

In its simplest form, binaural spatialisation can be achieved by filtering a sound signal with the HRTF pair appropriate for a given location, creating the spatial illusion. It would be highly desirable if the illusion of trajectory could be created in the same way. However, impulse responses cannot characterise the movement of a source along a trajectory, and there is no simple pair of filters which can be used to create the illusion of movement. The reason for this is that the propagation of sound from a moving source to the ear is not a linear, time-invariant system. This is fundamental and unavoidable, because at different moments in time the source will generally be at different points along its trajectory and thus the spectral filtering is time variable. Therefore, a means other than impulse responses or magnitude/phase responses is required to characterise the propagation of sound from moving sources.

The sound arriving at the ear from a moving source experiences spectral filtering due to physiological features such as head, torso and pinna; filtering which varies over time. Therefore, a simple method of characterising a moving source would be as an amplitude response which changes over time. This amplitude versus frequency versus time representation forms a *spectrogram*. Spectrograms will be familiar to researchers who have worked with time-varying audio signals, such as speech. A spectrogram is a three dimensional representation; amplitude values of particular frequency component over particular spaces of time. In terms of graphical plots, amplitude can be represented as greyscale or colour temperature. Audio frequency values usually form a vertical axis. Time is plotted along the horizontal (often as a rolling variable for ongoing analysis of continuous signals as in [16]), with the duration of a sound divided into a sequence of time domain windows (e.g. of duration 10ms) and the magnitude spectra of the signal during these intervals plotted

Let us define the Head-Related Spectrogram, or HRS. A pair of HRSs encompass the modification imposed on a sound signal between its source transcribing a particular trajectory and its detection at the ears. An HRS is essentially a continuous sequence of HRTF magnitude responses, corresponding to the points on the trajectory. When the trajectory is null - i.e. the source is stationary - the HRS pair collapses into the appropriate HRTF pair.

3.2 Spatial Sampling

In theory we could measure a pair of HRSs using similar techniques that are used in the measurement of static HRTFs: We could configure an anechoic chamber with a movable loudspeakers and an acoustic dummy with microphones for ears, then record those microphone signals as the loudspeaker moved along the trajectory whilst emitting a broadband signal. Although, like much acoustic measurement, the procedure is relatively laborious, such a technique might be practical for the measurement of a small number of HRSs. If we require full freedom of movement synthesis, however, we need the HRS pair for any conceivable trajectory and it is evident that this corresponds to an infinite data set. Given that HRSs are comprised of sets of HRTFs it is perhaps reasonable to consider compiling the required spectrogram values using a global set of static HRTF measurements. However, we have said that we seek to create the illusion of source moving along an arbitrary trajectory at an arbitrary rate. It is impossible to measure an exhaustive set of HRTFs (even 10° spacing of directions across the surface

of the sphere requires 648 separate measurements to be made for each ear) so there will always be areas between measurements where the precise acoustic information is not known. If we are synthesising sources with rapid angular movement that faithfully transcribe patterns of HRTF measurements then the spatially discrete HRTFs may be imperceptible from a continuum. However, in general, for fully flexible movement synthesis sources may be required to move relatively slowly between measurement positions, potentially yielding discontinuous movement, and along trajectories which do not map conveniently onto the spatial patterns formed by the distribution of measured HRTFs in a set.

Straightforward linear interpolation could be used to estimate magnitude response amplitudes for any directions between measured HRTFs. However, we are not considering a one-dimensional variable but the two-dimensional spherical surface formed by the fully set of directions relative to the head. Given that a set of HRTFs contains magnitude spectra corresponding to a discrete set of directions, we need to compute an estimate a spatially continuous representation of the amplitude with which each audio frequency will arrive at the ears. Provided that this model is sufficiently accurate, it should then be possible to calculate HRTFs at any spatial resolution, and thus compute spectrograms for any conceivable trajectory and rate of movement around the listener.

4 Continuous Modelling of HRTFs

4.1 Spherical Harmonic Analysis

A finite set of HRSs can be formed from experimentally measured HRTFs, but, in general, a moving source might move into the spaces between sampled directions. In order to alleviate any perceived discontinuities in movement and allow HRSs to be deduced with complete directional freedom we need a continuous HRTF model. In other words, the amplitude with which a given frequency component reaches the ear should be calculable as a function of azimuth and elevation. Since moving along a trajectory implies that azimuth and elevation are a function of time, for a given trajectory and with a functional model the amplitude of each frequency component becomes a function of time.

In [17] the author develops a technique for analysing HRTF data from a set of sampled directions to from a representation which is a weighted sum of surface spherical harmonics [18]. Spherical harmonics form an orthogonal set of basis function on the surface of a sphere and, therefore, are suitable for the development of a Fourier series for spherical functions in the same way as orthogonal sinusoidal harmonics can be used in the Fourier analysis of 1-D signals. The extent of angular frequency variation modelled by a particular spherical harmonic component is determined by its degree. Higher degree spherical harmonics are responsible for subtle variations from direction to direction whilst lower degree harmonic are responsible for significant directional differences, such as front-back or left-right variations. The higher the maximum degree components included in a spherical harmonic representation, the more accurate the interpolated approximation becomes. The spherical harmonic representation is continuous since each individual spherical harmonic is a continuous function of azimuth and elevation. Therefore, the weighted sum of components up to a particular degree is also a continuous function of direction.

4.2 Spectrogram Computation

In [17] an set of 648 HRTFs (corresponding to 10° spacing across the surface of the sphere) were measured in an anechoic chamber using an acoustic dummy. This data was used to form 48 sampled spherical functions; corresponding to 48 frequency bins from 210 Hz to 10101Hz. Thus, for each of these audio frequency the relative amplitudes with which sound of that frequency could reach the ear across the 648 directions was available. Spherical harmonic analysis was applied to the data for each frequency in turn yielding 48 continuous functions in azimuth and elevation. Spherical harmonic components up to and including degree-17 were calculated, leading to approximations with a mean square error of approximately 1% or better.

The 48 weighted sums of spherical harmonic components allow spectrograms with a frequency resolution of 210Hz to be computed. For the time axis, let us choose a windowing interval of 10ms. Therefore, the HRS for a particular trajectory and particular rate of travel can be compiled using a 10ms sampling of the source's direction. At each instant the azimuth and elevation of the source are used to calculate 48 spectral amplitude values.

4.3 Example Head-Related Spectrograms

Consider a source moving with a constant angular velocity from a position directly opposite the right ear (azimuth of 90° , elevation of 0° or $(90^\circ, 0^\circ)$) to an equivalent position opposite the left ear ($270^\circ, 0^\circ$). The trajectory goes round the rear of the head and takes 1.8 seconds for the source to traverse. With a temporal resolution of 10ms, the spectrogram consists of 180 sets of spectral amplitudes, corresponding to a 1° spacing. The HRS characterising propagation from this source to the left ear is shown in Figure 1.

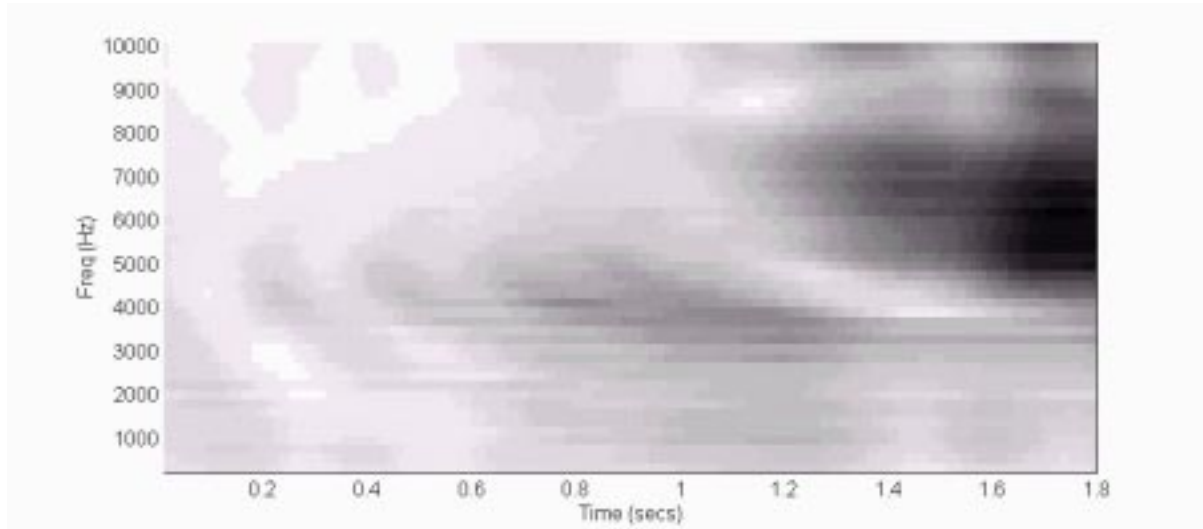


Figure 1: Left ear head-related spectrogram for right ear to left ear trajectory (dark shading indicates high amplitude)

Assuming a symmetrical head, the equivalent HRS for the right ear is simply a reflection in the median plane of the head, for this trajectory this is shown in Figure 2.

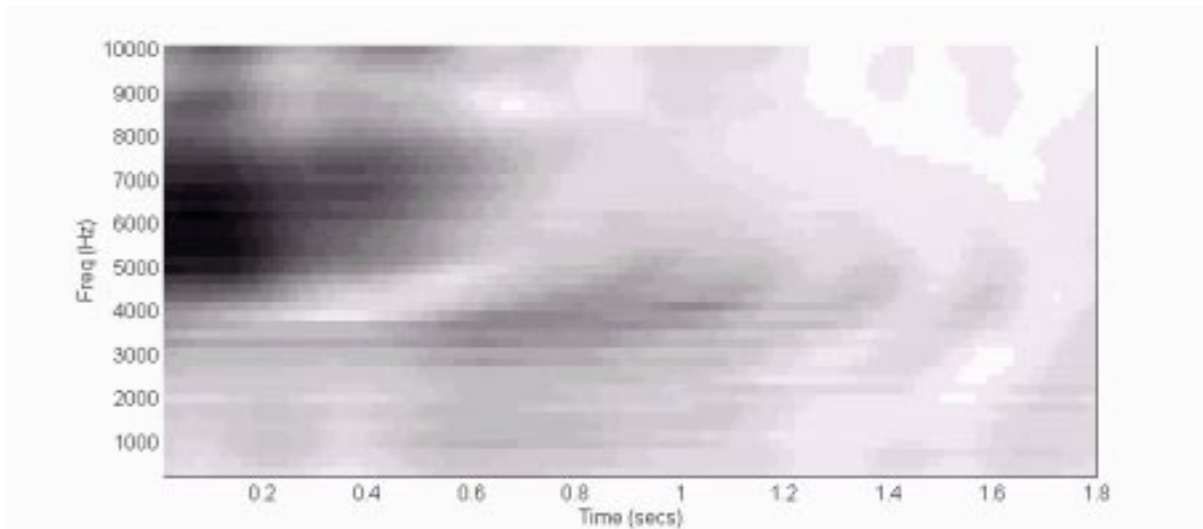


Figure 2: Right ear head-related spectrogram for right ear to left ear trajectory (dark shading indicates high amplitude)

The HRS pair shown in Figures 1 and 2 could conceivably have been computed without the need for the continuous HRTF model. (It would have required an extremely fine resolution set of HRTFs to be measured; 1° spacing on the horizontal plane.) However, the power and value of the continuous model can be demonstrated by considering a second trajectory: Once again this a horizontal plane trajectory of duration 1.8 seconds. However, in this trajectory the source subtends a smaller azimuth; starting at $(120^\circ, 0^\circ)$ and finishing at $(156^\circ, 0^\circ)$. With 10ms temporal resolution this corresponds to a new HRTF magnitude spectrum every 0.2° . It is impractical to consider making a set of experimental measurements at such fine resolution. However, using the

continuous directional model, the HRS pair characterising this trajectory can easily be computed. Left and right HRSs are shown in Figure 3.

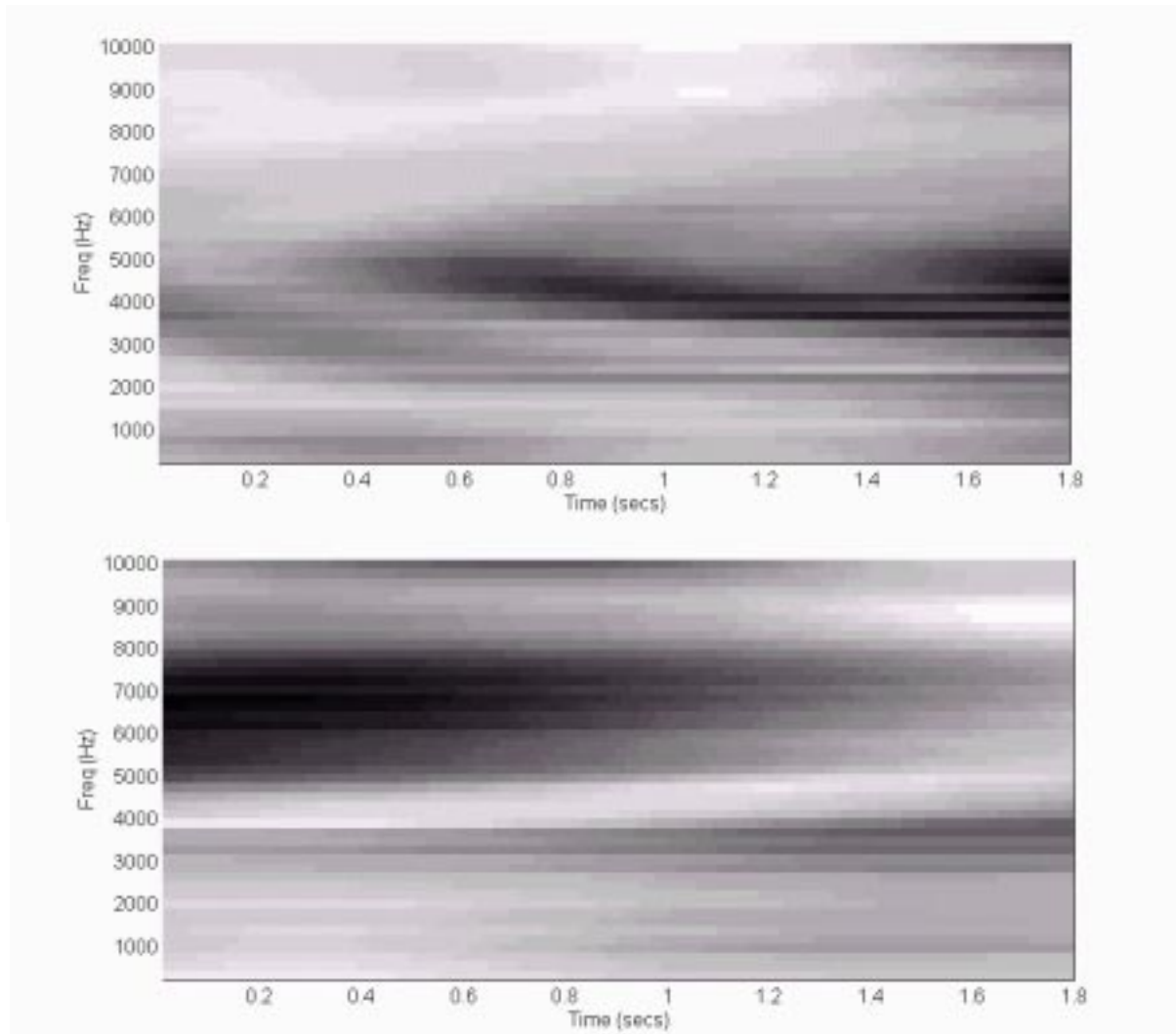


Figure 3: Left (top) and right (bottom) head-related spectrograms for shortened trajectory (dark shading indicates high amplitude)

Perhaps an even more compelling example is to consider a more erratic trajectory (such as might correspond to listener position movements as detected by a head position tracker). The trajectory described below changes direction and angular velocity on a number of occasions.

Start directly above the head ($0^\circ, 90^\circ$) and move down to ($0^\circ, 45^\circ$) in 300ms. Over the next 500ms move leftwards to ($270^\circ, 45^\circ$) and then move back and down for 300ms finishing directly behind ($180^\circ, 0^\circ$). Then move upwards and to the right, towards ($60^\circ, 30^\circ$) for 400ms and conclude by moving to the centre-front direction ($0^\circ, 0^\circ$). Total trajectory duration 1.8 seconds.

If we were trying to compute the HRS pair corresponding to this trajectory from a finite set of HRTF measurements, we would almost certainly face the prospect of irregularity and possible discontinuity due the changes in direction and velocity. The irregularity of the trajectory would not, in general, map onto a regularly spaced set of HRTF measurement directions. Similarly, the changing rate of travel would present problems for the regular time windowing required in the spectrograms. However, provided that the trajectory is continuous and that azimuth and elevation values are available at the 10ms intervals required for our HRSs any conceivable trajectory around the head can be synthesised. Figure 4 below illustrates the left ear and right ear HRSs for the irregular trajectory described above. Imposing this HRS pair on a sound signal (assuming compatible bandwidth) should yield the binaural signals that would reach the ears as a result of the source of that signal travelling the synthesised trajectory.

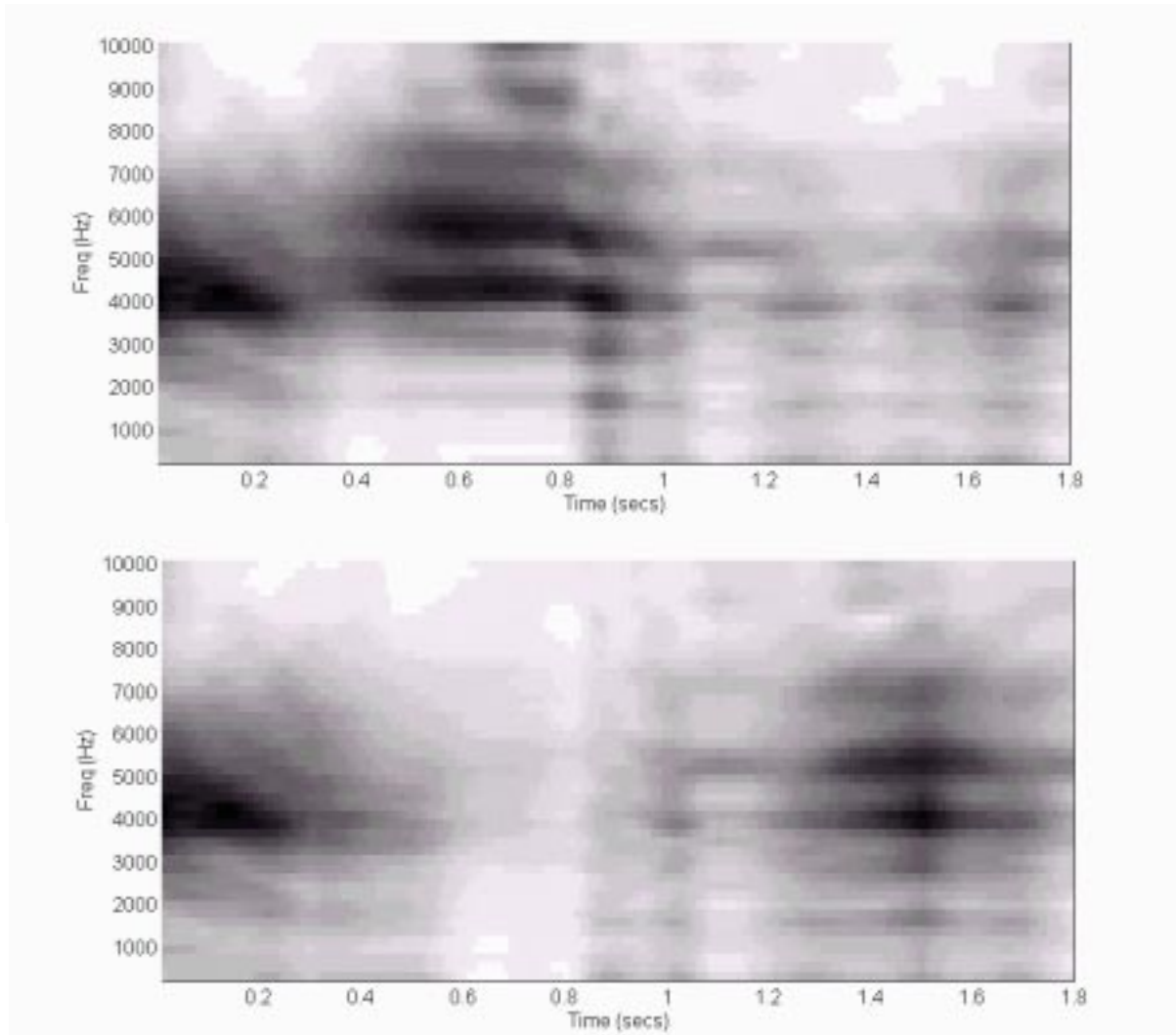


Figure 4: Left (top) and right (bottom) head-related spectrograms for irregular trajectory (dark shading indicates high amplitude)

5 Conclusions and the Future

5.1 Redundancies and Limitations

This research is still in its early stages. In particular, the performance of the HRSs in creating the illusion of moving sound sources to listeners will have to be verified across a range of stimuli and trajectories using subjective listening tests. There are currently no agreed standard testing techniques or metrics for localisation evaluation of static spatialisation let alone movement synthesis. However, the small number of informal listening trials that have been carried out seem to confirm the potential for this technique in synthesising source movement to listeners. Indeed, it seems possible that this technique of modelling source movement can encode excessive amounts of information. One serious problem which might restrict the use of this technique in real-time and interactive applications are the time overheads associated with computing the HRSs.

In general, there is evidently some redundancy in our HRS pairs. HRTFs themselves are empirical measurements which may contain features which are imperceptible to the auditory sense or irrelevant as localisation cues. However, HRTF spatialisation techniques encode and convey these features to the listener as if they were crucially important localisation cues. Moreover, we have discussed the finite and variable resolution of human localisation ability and described the existence of Minimum Audible Angle (MAA) and Minimum Audible Movement Angle (MAMA). The notionally infinite resolution of our spatial model (in practice determined by the angular velocity of the source) should be overspecified for the majority of trajectories. Processing economies might be made by exploiting MAA and MAMA and reducing the required

resolution of angular variation. However, if we are to achieve such increases in efficiency without compromising the perceived quality of the movement synthesis we must have a robust and accurate model of localisation resolution, and be confident that this model can deal with variations between listeners. Ultimately, if we have a sufficiently effective model of movement acuity it might be possible to obtain highly significant simplifications in movement synthesis by exploiting “snapshot” motion perception in many cases: For those trajectories for which a change of location alone is sufficient for the listener to infer the desired motion we can synthesise the movement as simply a transition between the discrete directions marking the end points of the motion.

5.2 Functional Descriptions

We have discussed a number of possible inefficiencies in movement synthesis using HRSs and identified computation time as a possible restriction to the use of this technique. More immediately there is a potential problem inherent in the use of spectrograms themselves. HRSs are comprised of a series of HRTF magnitude spectra, with a finite (10ms) temporal resolution between them. At these windowing instants the phase response of the resulting movement synthesised sound can exhibit some discontinuity. In [17] the phase responses of the HRTF set were analysed using the spherical harmonic technique. Although this data was dominated by the effects of interaural time delay (which can be easily modelled and compensated for [19]) some cues, generally poorly understood, to exist in the temporal structure of the responses [20].

A possible solution to this potential problem would be to find some method of describing an HRS in a genuinely continuous fashion, without the need for the finite temporal sampling generally associated with spectrograms. Once again, the spherical harmonic HRTF model may hold this solution since, in addition to being a continuous representation of sensitivity to each frequency across the range of directions, it is also a functional representation. The amplitude of any given frequency component is a function of azimuth (θ) and elevation (ϕ). For example, consider frequency components of 5050Hz. The amplitude with which these components reach the ear is expressed as a weighted sum of spherical harmonic components; each a continuous function in (θ , ϕ). Our full 17th degree representation contains 324 terms but for illustrative purposes let us consider just the four terms of the 1st degree spherical harmonic approximation.

$$H_{5050}(\theta, \phi) \approx 0.503 - 0.026 \sin \phi + 0.2 \cos \theta \cos \phi - 0.501 \sin \theta \cos \phi \quad (1)$$

Consider a constant angular velocity, 1.8 second trajectory which begins directly behind the head (180° , 0°) and finishes at a position directly opposite the left ear (270° , 0°). This is a continuous trajectory and an equation for the azimuth at time t can be written. (Note that this is a special case trajectory in which elevation is always zero. In general we could write an equation for elevation as a function of time as well.)

$$\theta(t) = 180 + \frac{90t}{1.8} \quad (2)$$

Substituting $\phi=0$ and Equation 2 into Equation 1 yields a continuous expression for the amplitude of the 5050Hz component as a function of time:

$$H_{5050}(t) \approx 0.503 + 0.2 \cos\left(180 + \frac{90t}{1.8}\right) - 0.501 \sin\left(180 + \frac{90t}{1.8}\right) \quad (3)$$

An equivalent function could be written for each frequency bin in the spectrum. Collectively they describe how spectral amplitudes should be varied to create the illusion of the given trajectory. The temporal resolution of this variation can be as fine or coarse as required, or indeed possible, given the particular situation. (Imagine as robot arm continuously moving the sliders on a graphic equaliser in accordance with these functions of time.) Even in interactive or other ‘live’ situations - such as responding to head tracker data - a function can still be written for an trajectory, by storing the last few directions and using ongoing processing to keep fitting and refitting a spline to these points. The equation of the spline will be a continuous function describing the trajectory over the last few samples of the tracker.

5.3 Summary

This paper has introduced a novel technique for representing the propagation of sound between sources moving around a human listener, and the ears. We have discussed the potential benefits of extending current binaural and transaural spatial audio techniques so that they can create the illusion of moving, as well as static, sound sources. Spectrograms have been proposed as a useful method of encoding the acoustic cues associated with a particular trajectory and we have established a platform for computing these head-related spectrograms using a continuous HRTF model. We have also discussed how a functional implementation of this model may yield more efficient sound motion synthesis by alleviating the need for excessive computation power and the ability to adapt to the finite and variable angular resolution of the human auditory system. Future research will use subjective listening tests to verify the perceived performance of sound source movement synthesised in this way.

6 References

1. Bregman A.S. Auditory Scene Analysis - The Perceptual Organization of Sound. MIT Press, 1990
2. Begault D.R. 3-D Sound for Virtual Reality and Multimedia. AP Professional, 1994
3. Wenzel E.M. Localization in virtual acoustic displays. Presence 1992; 1:80-107
4. Malham D.G., Myatt A. 3-D sound spatialisation using ambisonic techniques. Computer Music Journal 1995; 19:4:58-70
5. Cooper D.H., Bauck J.L. Prospects for transaural recording. Journal of the Audio Engineering Society 1989; 37:3-19
6. Evans M.J., Tew A.I., Angus J.A.S. Spatial audio teleconferencing - Which way is better? In: Proceedings of ICAD'97. 1997, pp. 29-38
7. Mills W. Auditory Localization. In: Tobias J.V. (ed) Foundations of Modern Auditory Theory. Academic Press, 1972
8. Middlebrooks J.C., Green D.M. Sound localization by human listeners. Annual Review of Psychology 1991; 42:135-159
9. Grantham D.W. Auditory motion perception: Snapshots revisited. In: Anderson T.R., Gilkey R.H. (ed) Binaural and Spatial Hearing in Real and Virtual Environments, Lawrence Erlbaum, 1997
10. Grantham D.W. Detection and discrimination of simulated motion of auditory targets in the horizontal plane. Journal of the Acoustical Society of America 1986; 79:1939-1949
11. Chandler D.W., Grantham D.W. Minimum audible movement angle in the horizontal plane. Journal of the Acoustical Society of America 1992; 91:1624-1636
12. Perrott D.R., Costantino B., Ball J. Discrimination of moving events which accelerate or decelerate over the listening interval. Journal of the Acoustical Society of America 1993; 93:1053-1057
13. Perrott D.R., Marlborough K. Minimum audible movement angle: Binaural localization of moving sound sources. Journal of the Acoustical Society of America 1989; 85: 1773-1775
14. Blauert J. Spatial Hearing. MIT Press, 1997
15. Møller H., Sørensen M.F., Hammershøi D., Jensen C.B. Head-related transfer functions of human subjects. Journal of the Audio Engineering Society 1995; 43: 300-321
16. Brookes T.S. A Real-Time Auditory Spectrograph. DPhil Thesis (Electronics), University of York, 1996
17. Evans M.J., Angus J.A.S., Tew A.I. Analyzing head-related transfer function measurements using surface spherical harmonics. Journal of the Acoustical Society of America 1998; in press
18. Kaplan W. Advanced Mathematics for Engineers. Addison-Wesley, 1981
19. Woodworth R.S. Experimental Psychology. Henry Bolt and Company, 1938
20. Wilde M.D. Temporal localization cues and their role in auditory perception, 95th Convention of the Audio Engineering Society 1993; AES Preprint #3708