Spherical Harmonic Coding of Sound Objects - the Ambisonic ‘O’ Format

D.G. Malham
Department of Music
University of York, York YO10 5DD, U.K

Abstract

Spherical harmonics can be used to encode the radiation patterns of individual sound objects in a manner compatible with the ambisonic ‘B’ format. This ‘O’ format enables the construction of more realistic sound objects for use within synthesised ambisonic soundfields, whilst maintaining the simplicity and ease of use of ambisonics. This paper examines extensions to basic ‘O’ format which allow much closer approximation of the changes in sound radiation caused by differing source-listener distances.

INTRODUCTION

Developed during the 1970's, Ambisonic surround sound systems use spherical harmonics to encode the direction of sound sources within a three dimensional soundfield. Recently this ‘B' format representation has been extended from the original, four channel, first order version to the second (Furse-Malham)**[1][2]** or higher (Daniel)**[3]** orders necessary to attain higher precision over a wider useful audience area or over a wider frequency range. However, even four channel, first order B format soundfields, if recorded from real acoustic scenes using an Soundfield microphone, capture the complex and extended nature of real sound radiating bodies accurately.

On the other hand, when constructing artificial soundfields, we have been, in general, limited to placing sound sources into the image using panning algorithms which treat the sounds as simple point sources. In natural soundfields, of course, there is no such thing as a point source. Even supposedly simple ***sounding objects*** exhibit very complex relationships between their source/listener distance and the listening position frequency response.

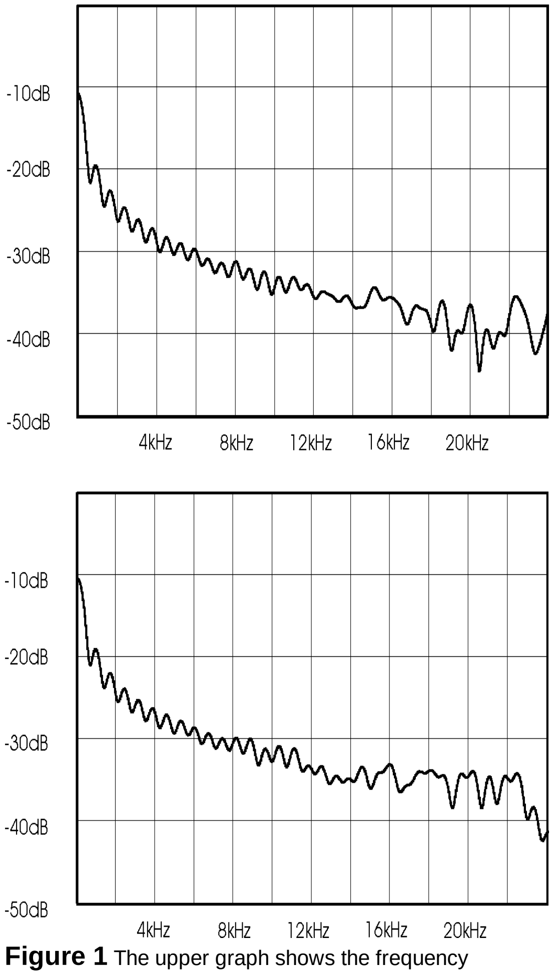

Figure 1 The upper graph shows the frequency response at a listening position 1.0m from a rectangular plane radiator 1x 1.1m in size, the lower graph, the response at 1.01m

For instance, **figure 1** illustrates how changing the source/listener distance of a simple sound radiator, in this case a flat, rectangular sheet by even very small amounts can produce major changes in perceived frequency response. Note that for a change of 1 centimetre in a metre, we find that, in the region around 2kHz, there are already deviations of around 1dB which increase to 6dB in the region of 18kHz. Clearly, in order to render sounding objects realistically within a sound image, it is insufficient to treat them as point sources.

To date, little seems to have been done in existing sound spatialization systems to address this limitation, at least so far as studio based systems are concerned. It is, of course, true that within many of the full acoustic simulation systems which are available, it may be possible to describe sounding objects in sufficient detail to avoid these problems. Such systems, however, tend to place very heavy demands on computing power and are generally limited to non-real time research applications.

On the other hand, the sort of synthesised soundfields produced by studio systems, whether they have stereo, binaural or even full surround sound outputs, have largely been limited to the use of either idealised point sources or to those giving a simplistic impression of having some sort of simple radiation pattern. For instance, in Microsoft’s DirectSound, the sound source is given a limited directional variability by defining cone shaped volumes where the sound changes character so as to appear to be facing towards or away from the listener’s position. **(fig. 2)**

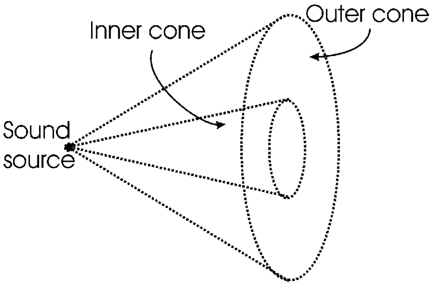

Figure 2 Sound cones in DirectSound acoustic images

Typically, only simple amplitude differences distinguish the inner (louder) cone, the transition zone between the inner and outer cones and the quiet outer zone. Despite the simplicity of this approach, it has been shown to give a greater sense of realism in computer games, particularly in conjunction with the visual clues.

Several investigations have been conducted into simulating musical instrument radiation patterns using arrays of loudspeakers **[4][5]** but these do not produce synthesized soundfields directly within the electronic domain, as is need for studio applications. Except for the output of full acoustic modelling systems, which necessarily involve considerable computing power, current systems do not provide good, realistic synthetic images of complex, extended sound radiating surfaces.

Sounding Objects in Ambisonic Systems

In the case of Ambisonic systems, the extended nature of sound sources has, in the past, been dealt with in a minimalistic way by either simply exaggerating the non-directional zeroeth order spherical harmonic component or by the use of phase shift based 'spreader' controls **[6]**. Yet this is exactly the kind of problem which spherical harmonics are most commonly used for. Since the nineteenth century they have been used in, for instance, describing the distribution of electrostatic charge on a surface, or the pull of gravity on a planetary surface or the radiation of heat from a hot body.

Following a suggestion made by Dylan Menzies-Gow in his PhD thesis **[7]**, a spherical harmonic coding system for the description of sound radiation has been implemented in Ambisonics. This encoding format, known as ***O format*** consists, at its simplest, of just the directional radiation pattern, as in Menzies-Gow's original suggestion. It can, however, be extended to include information about both the surface shape and it's frequency dependent emission characteristic.

Even simple modelling of the radiation pattern of the object provides noticeable improvement. Since this uses spherical harmonics in a manner analogous to the coding of soundfields in ordinary B format Ambisonics, the object can be rotated so that it may be oriented correctly to the listening position. The simple model, does not, however allow the effects of variation of the sound at the listening position with to be simulated properly. The variation in the frequency domain is largely due to changes in the impulse response at the listener's position. The impulse response changes with differing distances in two ways, as shown in **Fig. 3**.

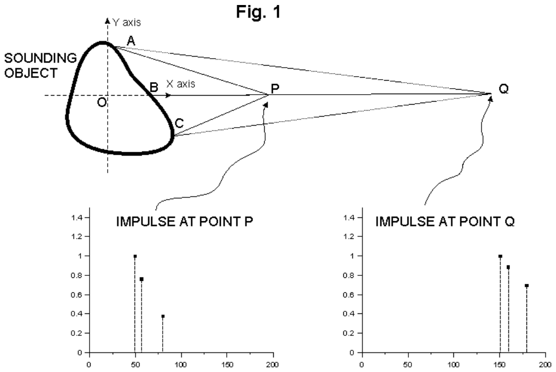

Figure 3 Impulse responses at different distances

Here the impulse response is, for simplicity, shown as being provided by three points, **A, B** and **C** (although in reality all points on the surface would contribute) and for two listener positions, **P** and **Q**. Both the position of the impulses in time and the differences in their amplitudes change with distance. Note that as the distance increases between the object and the listener the extra distance contribution of the displacement away from the origin along the Y axis decreases leading eventually, in the far field, to the situation where only distances along the X axis count

Surface Shape

If, as well as the basic radiation pattern of the sounding object, a description of the object's three dimensional shape is provided, the impulse response can be calculated for each source/listener distance as well as for each direction. Within a full acoustic modelling system, this is usually part of a complex ray-tracing algorithm which deals with a fully detailed description of the object. However, it can be easily seen that there is considerable redundancy in this since, in contrast to the situation in visual systems, the wavelengths of sound are usually significant compared to the size of the structures in the sort of sounding objects we want to model. It is possible, therefore, to use a simplified surface shape description and still obtain significant improvements in terms of the modelled impulse response of a sounding object.

If a description of surface shape based on spherical harmonics is used, this O format image can be manipulated spatially at low computational cost, although this can only be done prior to embedding the sound object in a B format soundfield, allowing orientation of the image to be varied prior to embedding. After embedding, only the normal manipulations of the soundfield as a whole are possible. Note, however, that although it is possible to use spherical harmonics directly to model surfaces and to calculate the impulse responses from the shape, in this application it is more efficient to use sets of impulse responses, one for each spherical harmonic, as in **Figures 4** and **5**.

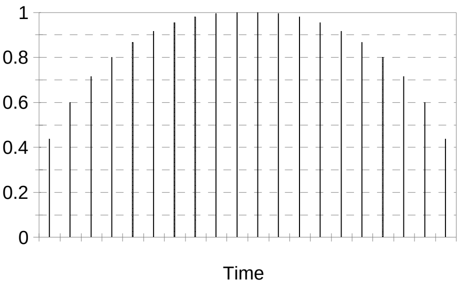

Figure 4 0th Order Impulse Response

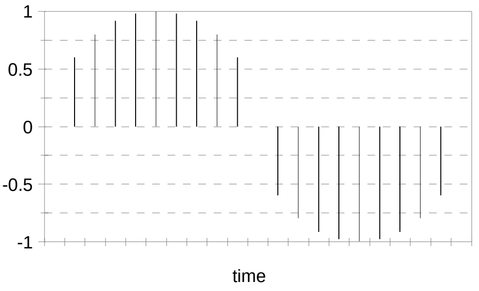

Figure 5 1st Order Impulse Response

These impulse responses represent the far field response of the spherical harmonic primitives. The length of each is set by the overall size of the object and the height of each time point is determined by the area radiating at that distance. There is no additional effect resulting from the distance off axis of each radiating point, nor is the inverse square law applied. A modified form of impulse response such as this will be referred to as a ***non-distance weighted impulse response (ndw)***.

Such **ndw** responses can either be calculated, for instance from a model generated in a standard CAD program or they can also, as we shall see later, be measured directly from real objects.

Once the modified impulse responses have been produced in spherical harmonic form, the sounding object is then oriented in the acoustic scene in accordance with its relationship to the listener, for instance by applying rotational transforms such as an angular rotation to the left by an angle of β from the centre front coupled with a tilt by an angle α from the horizontal, which requires the following transformation, assuming a first order description of the object;

$$W' = W$$

$$X' = X * \cos\beta - Y * \sin\beta$$

$$Y' = X * \sin\beta * \cos\alpha + Y * \cos\beta * \cos\alpha - Z * \sin\alpha$$

$$Z' = X * \sin\beta * \sin\alpha + Y * \cos\beta * \sin\alpha + Z * \cos\alpha$$

where W', X', Y', Z' form the rotated and tilted spherical harmonics describing the reoriented sounding object. Following this transformation, a weighted sum of the **ndw** spherical harmonic coded impulse responses can be produced corresponding to the **ndw** impulse response required for the relationship of the sounding object to the listening position. The effects of distance on the amplitude of each impulse can then be applied by weighting the value of the impulse at each time points according to the inverse square law, derived by using the formula

$$\left(\frac{T_S}{T_C}\right)^2$$

where Ts is the time of appearance of the first component in the impulse response and Tc that of the current component. This can be used as the final impulse response, provided that the sounding object is effectively in the far field of the listener. The accuracy of its match to reality can be chosen, in accordance with the computing power available and the quality of effect desired, by varying the number and maximum order of spherical harmonics used. As in Ambisonics B format, if the number of spherical harmonics used, and hence the overall order of the system, is reduced by removing the higher orders first using a suitable windowing function, the degradation caused is gradual and smooth.

The order of the O format object does not have to match that of the soundfield it is embedded in, since it is passed through a matrix akin to that used for speaker decoding prior to being added to the B format soundfield and only the output of the matrix need be of matching order. This means that high order descriptions of sound objects can be embedded in standard low order soundfields, allowing very rich acoustic behaviour to be implemented without necessarily impacting on the final channel numbers and hence the storage required.

When the sounding object is in the near field of the listener, the impulse response has to be modified. In this case, the distance between the sounding body and the listener is such that the effect of the distance off axis becomes significant and so has to be incorporated. In this case, the time axis is be warped to model the extra delay imposed by the distance point is from the Y axis. A typical warping factor is represented by that for the zero order spherical harmonic

$$\sqrt{\left(\sin\left(\cos^{-1}(n)\right)\right)^2 + (1-n)^2}$$

where ***n*** is the number of the sample and all distances are expressed in terms of multiples of the size of the object. The effect of sound diffusion from those areas of the sounding object which are either facing away from the listener or whose acoustic path to the listening position is otherwise obstructed may need to be modelled, since sounds of wavelengths small compared to the object are delayed more than others, but this is only needed where the model has to be of extreme accuracy.

Following computation of the final impulse response, the actual sound source is processed by the impulse response so generated, via convolution. This produces the appropriate frequency domain corrections such that it will sound as if it was emitted by the sounding object at the desired distance and orientation from the listening body. Further processing by the standard Ambisonic panning processes, or indeed by any other form of sound spatialization, can then be used to produce the final image.

Object Size

As indicated, the duration of the **ndw** impulse response is dependent on the overall size of the sounding object. This can, however, be looked at in another way. Since the shape of the spherical harmonic components of the **ndw** responses remain constant from object to object, with only the actual length varying, it can be seen that only one set of responses needs formally calculating, since those for any other size of object can be generated by re-sampling the basic set. This also means that the apparent size of an object can be dynamically varied by changing the re-sampling rate

Measurement of Real Objects

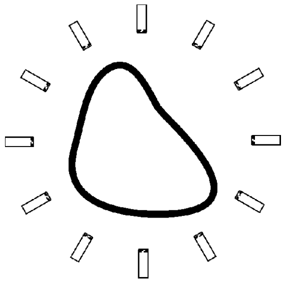

Figure 6 Microphone array for capturing object shape

The radiation pattern of a real sound object can be captured, within certain limitations, by sampling the soundfield around the object using a suitable array of microphones **(Fig. 6)**. The sampled soundfield has then to be re-coded into spherical harmonics to produce an O format file which can again be processed (rotated, tilted, tumbled, etc.) prior to final embedding into the B format soundfield.

With this sort of array, the time of arrival of the first sound at each microphone can be used to determine the distance to the nearest point to that microphone. Any other measureme0.nt technique can be substituted for this, of course, prior to deriving the weighting of the spherical harmonics encoding the shape with a Fourier series analysis. This yields the following formulae for the weights of each spherical harmonic component.

$$P_{mn} = \int_{\phi=0}^{\pi} \int_{\theta=0}^{2\pi} f(\theta,\phi)\, p_{mn}(\theta,\phi) \sin\phi\, d\phi\, d\theta, \quad 0 \le m \le n$$

$$Q_{mn} = \int_{\phi=0}^{\pi} \int_{\theta=0}^{2\pi} f(\theta,\phi)\, q_{mn}(\theta,\phi) \sin\phi\, d\phi\, d\theta, \quad 1 \le m \le n$$

Since the measurements will, in general, be taken on a discrete grid, we may approximate this using a formula such as;

$$\int_{\phi=0}^{\pi} \int_{\theta=0}^{2\pi} f(\theta,\phi) S_{mn}(\theta,\phi) \sin\phi\, d\phi\, d\theta \approx \sum_{i=1}^{N} f(\theta_i,\phi_i) S_{mn}(\theta_i,\phi_i)$$

where N is the number of points used.

Rather than making measurements of size and then computing the **ndw** responses, they can be measured directly. To a reasonable degree of accuracy this can be done using microphones, provided they are placed far enough away from the sounding body **(Fig. 7)**. The distance between sounding body and measuring microphone must be such that the angles subtended by any point on the surface of the sounding body away from the microphone's axis is so small that there is an insignificant extra time difference between points the microphone axis and those off it. If this condition is met, the impulse response becomes an **ndw** response containing all the spherical harmonic components in their weightings for that direction. Measurement of a sufficient number of these impulse responses over an appropriate grid of measurement points enables the individual spherical harmonic components to be obtained, via a similar process of approximation to that discussed above .

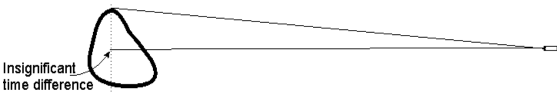

Figure 7 Measurement of **ndw** response using distant microphone

Surface Radiation Characteristics

So far the sounding objects considered have been treated as if they radiate sound homogeneously from all points on their surfaces. For real objects, this is not usually the case. Indeed, an impulse response measured by the approach shown in **Fig. 7** contains not just information related to the object's shape, but also information about the surface radiation pattern, that is to say, the way in which sound is emitted from each point on the surface. This may, and, in real objects, usually does, vary in level or in spectral content or both. It is therefore worthwhile investigating how this can be incorporated in O format.

Surface radiation characteristics can be encoded using spherical harmonics in exactly the same manner as the surface shape is encoded in O format. In the case of surface radiation characteristics, what needs to be encoded is essentially a filtering function. The filtering can be as simple as a gain function, reflecting the radiation efficiency of each part of the surface. The sound cones shown in **Fig. 2** could easily be implemented this way, although if it was desired to retain sharp edges a high order spherical harmonic encoding would be required.

More detail can be provided in the model of surface characteristics by incorporating frequency dependence This can be accomplished by defining another spherical harmonic encoded impulse response set. For flexibility, the surface radiation set and the shape set are normally kept (and applied) separately, allowing considerable creative freedom. Where computational power is at a premium, however, they can be combined as a single set, which is then equivalent to the measured **ndw** responses discussed earlier.

Conclusions

O format as a means of modelling the three-dimensional radiation characteristics of a sounding body has been discussed. It has been shown to be an economical and easily manipulated means of modelling the complex behaviour of real sound sources. It provides a significant improvement over existing methods of dealing with sound sources in artificial soundscapes, where the sound sources are often modelled only as points. The quality of the image produced can be selected depending on the available computing power, degrading in a graceful manner with resource availability to the limit at which it is the same as existing point source models.

References

1. Malham, D.G. “Higher order Ambisonic systems for the spatialisation of sound” ICMC Proceedings, Beijing, 1999, pp484-487

2. http://www.york.ac.uk/inst/mustech/3d_audio/secondor.html

3. Daniel, J., “Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia” PhD thesis, 1996-2000 Université Paris 6

4 Caussé, R. Bresciani, J., and Warsufel, O., “Radiation of Musical instruments and control of reproduction with loudspeakers” Proceedings of the International Symposium on Musical Acoustics, Tokyo, 1992

5. Cook, P. R., and Trueman, D., BoSSA: “The Deconstructed Violin Reconstructed” ICMC Proceedings, Beijing, 1999, pp232-239

6. Gerzon, M.A. “Artificial Reverberations and Spreader Devices”' NRDC Ambisonic Technology Report no. 4. August 1975.

7. Menzies-Gow, D. “New Electronic Performance Instruments for Electroacoustic Music” PdD. thesis, University of York, 1999 pp.99-101