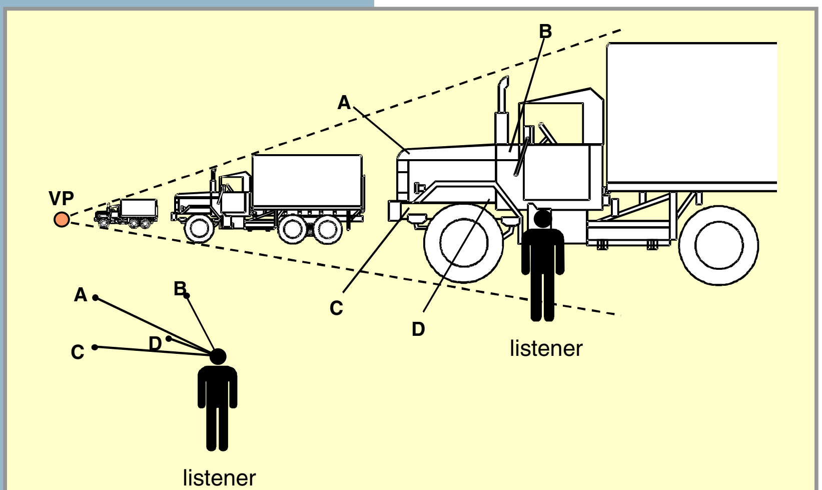
Sensaura

ZoomFX™ for 3D-sound

by Alastair Sibbald



At the present time, all 3D-sound methodology is based on the synthesis of virtual sound sources as *point source* emitters. In reality, of course, this is far from the truth, because many sound-emitting objects radiate acoustic energy from extended areas, strips or volumes. Sensaura ZoomFX has been developed to take account of this and to help create virtual 3D-sound images that are really true-to-life.

This is achieved by creating each virtual sound area from several, similar, virtual point sources, rather than from a single point source, as is commonly done. By distributing a number of virtual sources over a prescribed area or volume that corresponds to the physical nature of the sound-emitting object that is being synthesised, then much more realistic effects can be obtained. Now the synthesis is more truly representative of the real physical situation. Moreover, if the virtual sources within the array are arranged to maintain constant relative positions to one another, then when they are made to approach or leave the listener, the apparent *size* of the virtual sound-emitting object changes, just as it would in reality - Audio Zoom!

In order to create a secondary array of virtual sources from a single original sound, the secondary sounds must be decorrelated from the primary sound and from each other in order that the brain can perceive them individually. A new method has been developed to implement this, termed Dynamic Decorrelation™, which also has applications in the virtualisation of both stereo and various movie surround formats.

Sensaura™
3D POSITIONAL AUDIO

www.sensaura.com

1 Background

A monophonic sound source can be processed digitally via a pair of Head-Response Transfer Functions (HRTFs), as shown in the block diagram of Figure 1, such that the resultant stereo-pair signal contains natural 3D-sound cues. These sound cues are introduced naturally by the head and ears when we listen to sounds in real life. They include the inter-aural amplitude difference (IAD), inter-aural time difference (ITD) and spectral shaping by the outer ear[1]. When this stereo signal pair is introduced efficiently into the appropriate ears of the listener, by headphones say, then he or she perceives the original sound to be at a position in space in accordance with the spatial location of the HRTF pair which was used for the signal-processing.

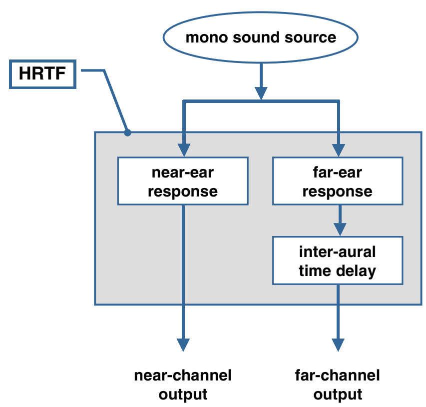


Figure 1: HRTF processing

When listening through loudspeakers instead of headphones, then the signals are not conveyed efficiently into the ears because there is *transaural acoustic crosstalk* present which inhibits the 3D-sound cues[2]. This means that the left ear hears a little of what the right ear is hearing (after a small, additional time delay of around 0.2 ms) and vice versa. In order to prevent this happening, it is possible to create appropriate *crosstalk cancellation* signals from the opposite loudspeaker. These signals are equal in magnitude and inverted (opposite in phase) with respect to the crosstalk signals and designed to cancel them out. There are also more advanced cancellation schemes which anticipate the secondary (and higher order) effects of the cancellation signals themselves, which otherwise can create secondary crosstalk.

When the HRTF processing and crosstalk cancellation are carried out correctly, using high quality HRTF source data, then the effects can be quite remarkable. For example, it is possible to move the image of a sound-source around the listener in a complete horizontal circle, beginning in front, moving around the right-hand side of the listener, behind the listener and back around the left-hand side to the front again. It is also possible to make the sound source move in a vertical circle around the listener and, indeed, make the sound appear to come from any selected position in space.

In this traditional implementation, common to many current technologies, each virtual sound source is created and represented by means of a single point source.

Each virtual sound source is representative of a sound-emitting entity such as a voice, a helicopter or a waterfall, for example. However, although it might be reasonable to represent a voice by a point source, helicopters and waterfalls are certainly not point sources, unless they are distant.

Each virtual sound source can be complemented and enhanced by the addition of secondary effects which are representative of a specified virtual environment, such as sound reflections, echoes and absorption, thus creating a virtual sound environment.

2 Real-World sound emitters

The emission of sound is a complex phenomenon. For any given sound source, one can consider the acoustic energy as being emitted from a continuous, distributed array of elemental sources at differing locations, having differing amplitudes and phase relationships to one another. If one is sufficiently far away from such a complex emitter, then the elemental waveforms from the individual emitters sum together, effectively forming a single, composite wave which is perceived by the listener. There are several different classes of distributed sound emitter, as follows.

Point source emitters

In reality, there is no such thing as a point source of acoustic radiation: all sound-emitting objects radiate acoustic energy from a finite surface area (or volume). In practise, the physical size of sound-emitting areas encompasses a wide range. For example, a small flying insect emits sound from its wing surfaces, which might be only a few square millimetres in area. In practise, the insect could almost be considered to be a point source, because, for all reasonable distances from a listener, it is clearly perceived as such.

Line source emitters

When considering a vibrating wire, such as a resonating guitar string, the sound energy is emitted from a (largely) two-dimensional object: it is, effectively, a *line* emitter. The sound energy per unit length has a maximum value at the antinodes and minimum value at the nodes. An observer close to a particular string antinode would measure different amplitude and phase values with respect to other listeners who might be equally close to the string, but at different positions along its length, near, say, to a node or the nearest adjacent antinode. At a distance, however, the elemental contributions add together to form a single wave, although this summation varies with spatial position because of the differing path lengths to the elemental emitters (and hence differing phase relationships).

Area source emitters

A resonating panel is a good example of an area source. As with the guitar string, however, the area possesses nodes and antinodes according to its mode of vibration at any given frequency and, at sufficient distance, these add together to form, effectively, a single wave.

Volume source emitters

In contrast to the insect ‘point source’, a waterfall cascading onto rocks might emit sound from a volume that is thousands of cubic metres in size: the waterfall is a very large volume source. However, if it were a great distance from the listener, and still within hearing distance, it might be perceived as a point source. In a volume source, some of the elemental sources might be physically occluded from the listener by absorbing material in the bulk of the volume.

3 Practical considerations

What are the important issues in deciding whether a real, distributed emitter can be considered to be a point source, or should it be synthesised as a more complex, distributed source? One factor that distinguishes whether or not a perceived sound source is similar to a point source is the angle subtended by the sound-emitting area at the head of the listener. In practical terms, this is related to our ability to perceive that an emitting object has an apparent significant size greater than the smallest practical point source, such as the insect. According to the literature[3], the ‘minimum audible angle’ corresponds to an inter-aural time delay (ITD) of approximately 10 µs, which is equivalent to an incremental azimuth angle of about 1.5° (at 0° azimuth and elevation). In practical terms, it is convenient to use an incremental azimuth unit of 3° in 3D-sound synthesis. This is because it is sufficiently small as to be almost indiscernible when moving a virtual sound source from one point to another and, also, the associated time delay corresponds approximately to one sample period (at 44.1 kHz frequency). However, these values relate to differential positions of a single sound source and not to the interval between two concurrent sources.

Practical experiments indicate that a good criterion for differentiating between a point source and an area source, based on the magnitude of the subtended angle at the listener’s head, is about 20°. Hence, if a sound-emitting object subtends an angle of less than 20° at the head of the listener, then it can be considered to be a point source; if it subtends an angle greater than 20°, then it is not.

4 Composite virtual sound sources

As an extension to the principle of synthesising a virtual sound source from a number of secondary sound sources (where the sources derive from a single primary source, such as a .wav computer file), an alternative approach exists where the sound sources themselves could be different to each other. This is a powerful method of creating a virtual image of a large, complex sound-emitting object such as a helicopter, where a number of individual sources can be identified.

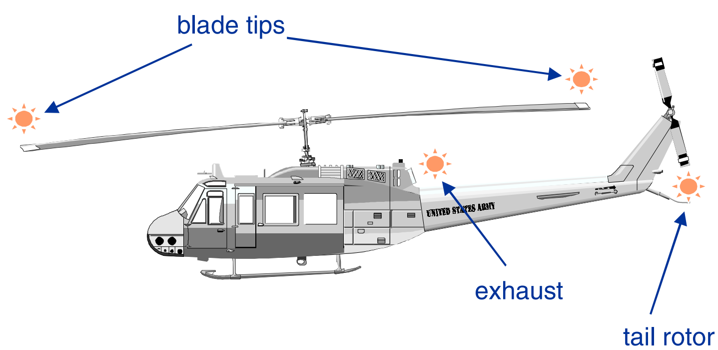


Figure 2: Individual sound source locations in a real sound-emitting object

For example, Figure 2 shows a diagram of a helicopter indicating several primary sound sources, namely the main blade tips, the exhaust and the tail rotor. Similarly, a truck comprises many sound-emitting surfaces, such as the engine block, the tyres and the exhaust. In both cases, it is more realistic to create a composite sound image of the object by means of several individual virtual sound sources: one for the exhaust, one for the rotor and so on. In a computer game application, the game itself links the individual sources geometrically, such that when they are relatively distant to the listener they are effectively superimposed on each other, but when they are close up they are physically separated according to the pre-arranged geometry and spatial positions. An important consequence of this is that a virtual sound source of this type scales with distance: it appears to increase in size when it approaches and diminishes when it goes away from the listener. Also, when this sound source approaches close to the listener, it is much more convincing, unlike systems where a single point source is used to create a virtual image of each object, irrespective of its physical size.

5 Creating a ZoomFX sound-source array

A suitable array of secondary virtual sound sources can be formed from a single primary source, such as a computer .wav file, by creating several decorrelated copies of the original. One method for doing this is to comb-filter the original using differing comb-filter parameters. The drawback with this approach is that comb filtering creates noticeable tonal changes. Also there are limits on the amount that the filter parameters can be varied and so it is possible to make only one or two secondary sources.

In order to overcome this limitation, Sensaura engineers have devised a new method for decorrelating sound sources called 'Dynamic Decorrelation'. This, tonally, is relatively neutral and, because of the random nature of the processing, there is little restriction on the number of secondary sources that can be created.

Figure 3 shows a simple block diagram of the method which is used to create an array of three (or more) secondary virtual sound sources from a single, monophonic primary sound source, S (such as a sound recording, .wav file or similar).

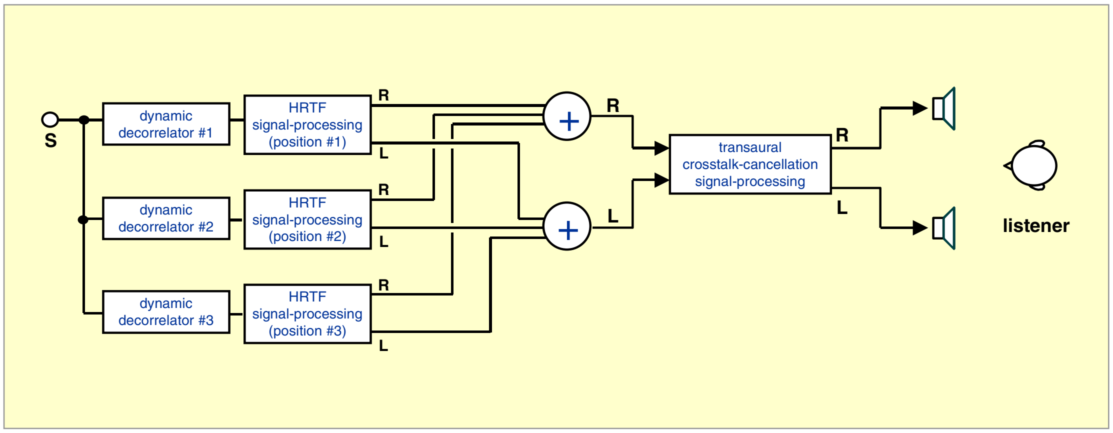


Figure 3: Conversion from single primary sound-source to secondary array using Dynamic Decorrelation

Consider, now, the situation where it is required to create the effect of a large truck passing the listener at differing distances, as shown in Figure 4. At a distance, a single point source is sufficient to simulate the truck. However, at close range, the engine enclosure panels emit sound energy from an area that subtends a significant area at the listener's head, as shown, and it is appropriate to use a number of virtual sources, as described above (Figure 3).

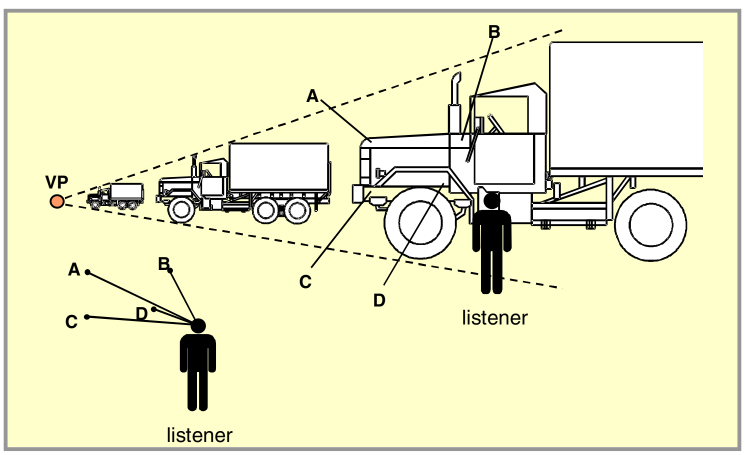


Figure 4: An area source simulated by an array of decorrelated sound sources

As mentioned previously, many real-world sound sources can be broken down into an array of individual, differing sounds. For example, a helicopter generates sound from several sources (as shown previously in Figure 2), including the blade tips, the exhaust and the tail-rotor. If one were to create a virtual sound source representing a helicopter using only a point source, it would appear like a recording of a helicopter being replayed through a small, invisible loudspeaker, rather than a real helicopter. If, however, one uses ZoomFX to create such an effect, it is possible to assign various different virtual sounds for each source (blade tips, exhaust, and so on), linked geometrically in virtual space to create a composite virtual source, such that the effect is much more vivid and realistic. There is a significant added benefit in doing this, because when the virtual object draws near or recedes, the array of virtual sound sources similarly appears to expand and contract accordingly, which further adds to the realism of the experience. In the distance, of course, the sound sources can be merged into one, or replaced by a single point source.

(Incidentally, one might ask, "Why not use a large area source for a particular HRTF measurement, if this is the final effect which is required?" The answer is that if a large loudspeaker were to be used for the HRTF measurements, then the results would be gross and imprecise. The measured HRTF amplitude characteristics become meaningless because they are effectively the averaged summation of many. In addition, it would become impossible to determine a precise value for the inter-aural time delay element of the HRTF, which is a critical parameter. The results are therefore spatially vague and cannot be used to create distinctly distinguishable virtual sources.)

6 Additional applications

There are several other 3D audio applications to which the principles of ZoomFX can be extended, primarily for headphone listening. These are the virtualisation of: (a) conventional stereo; (b) Dolby Pro-Logic™; and (c) Dolby Digital™ [Dolby Surround™].

Virtualisation of stereo

In some circumstances, when virtual sound effects are to be recreated to the sides of the listener, the HRTF processing itself can decorrelate the individual signals sufficiently such that the listener is able to distinguish between them and hear them as individual sources, rather than 'fuse' them into apparently a single sound. However, when there is symmetry in the placement of the individual sounds (say, one is to be placed at -30° azimuth in the horizontal plane and another is to be placed at +30°), then our hearing processes cannot distinguish them separately and a vague, centralised image is created. For individual left or right channels, there is no problem in creating virtualised images via headphones. However, when there is a common-mode signal present on both L and R channels, and they are virtualised symmetrically, then the common mode signal tends to occur inside the head, which detracts from the out-of-head effect that is required. By decorrelating the L and R channels prior to virtualisation, this effect can be significantly reduced, creating a pleasing and comfortable listening experience.

Virtualisation of Dolby Pro-Logic

Dolby Pro-Logic systems generate a rear surround channel that is fed into a pair of rearward placed loudspeakers (in anti-phase, so as to obviate the precedence effect and make both speakers separately discernible). For headphone virtualisation, the difficulty is that a single, rearward source sounds like a single point (unsurprisingly) and lacks 'spaciousness'. By using a pair of dynamic decorrelators, the single rear channel can be transformed into a rearward pair and then effectively virtualised.

Virtualisation of Dolby Surround

Dolby Digital systems generate a pair of surround channels which are fed into a pair of rearward/laterally placed loudspeakers. In the cinema, these are purposely engineered to create a *diffuse* sound field, rather than act as point sources. By using dynamic decorrelators on both left- and right-surround channels, 'extended' sound source arrays can be created, thus aiding the simulation of a cinema listening environment. This also has potential value for THX™-related processing.

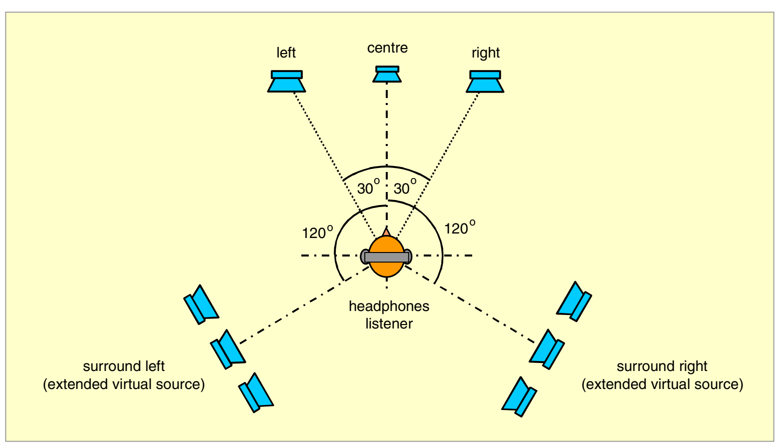


Figure 5: Virtualised extended-source rear speakers for AC3 system

7 References

1. An introduction to sound and hearing.
   A Sibbald
   Sensaura Technical White Paper,
   DEVPC/005 (1998).

2. Transaural acoustic crosstalk cancellation.
   A Sibbald
   Sensaura Technical White Paper,
   DEVPC/009 (1999).

3. On the minimum audible angle.
   A W Mills
   J. Acoust. Soc. Am., 1958, 30, (4),
   pp. 237-246.

For further information please contact:

Email: dev@sensaura.com
WWW: www.sensaura.com
Tel: +44 20 8848 6636

Sensaura Limited
Dawley Road, Hayes
Middlesex, UB3 1HH
United Kingdom

A Scipher Company

DEVPC/012/0001/1