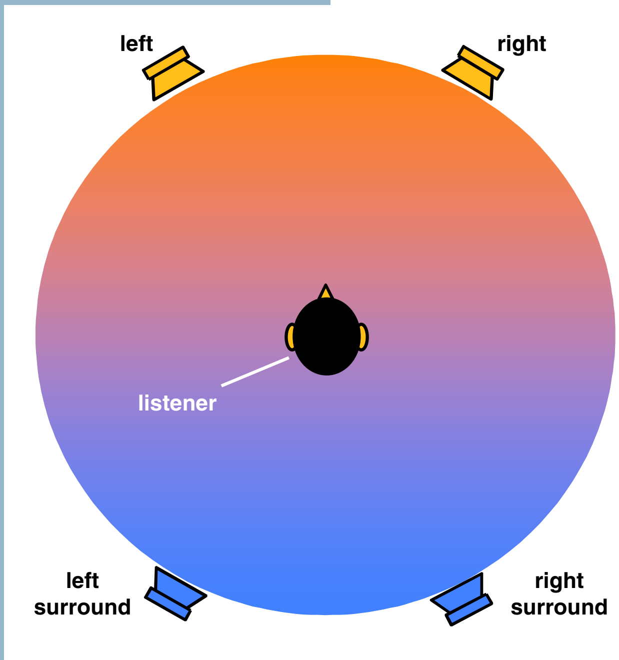
# MultiDrive™ Technology

by Alastair Sibbald & Richard Clemow




**Sensaura 3D Positional Audio technology now incorporates a patented multi-speaker capability for creating 3D sound fields using four or more loudspeakers. This is achieved by using each speaker pair to create a complementary sound field, each addressing the front and rear hemispheres respectively, as shown here on the left. This creates robust 3D audio, even in difficult listening environments.**

**The frontal and rearward sound hemispheres are seamlessly integrated together such that any number of virtual sound sources can be made to move around the listener, controlled by DirectSound3D. In addition, Sensaura MultiDrive technology uses patented Sensaura transaural crosstalk cancellation algorithms, which ensure optimal performance. Now, multi-speaker system owners are not restricted to flat, 2D sound panning, but can enjoy the full benefits of immersive 3D sound.**

## 1 Spatial audio for multimedia

At present, the two major formats of multimedia spatial sound technology are 3D Positional Audio, which plays over two loudspeakers, and multi-speaker surround sound, which requires four (or more) loudspeakers. (The term *spatial sound* here excludes the *stereo-effects* families of technologies, such as stereo enhancers and stereo expanders.)

## 2 Multi-speaker surround and the Haas effect

There are several systems available which create 2D 'surround' sound effects using a pair of frontal speakers (and, sometimes, a centre one) together with a pair of rearward loudspeakers. Such systems rely on panning the signal intensities between the various speakers to position the sound image. Prior to the introduction of the new 3D Positional Audio systems, this was the only real option for creating a spatial effect. However, the optimum listening position (sweet spot) of such intensity-based systems is very narrow because of the *Haas* effect.

Sensaura™
3D POSITIONAL AUDIO

The Haas (or *Precedence*) effect[1] is the phenomenon that, when presented with several similar pieces of audio information to process at slightly differing times, the brain uses the *first* information to arrive from which to compute directional information. It then attributes the subsequent, similar information packets with the same directional information.

For example, if several loudspeakers are playing music at exactly the same loudness in a room, *all* the sound appears to come from the *nearest* loudspeaker and all the others appear to be silent. The first signals to arrive are used to determine the spatial position of the sound source and the subsequent signals simply make it sound louder. This effect is so strong that the intensity of the second signal could be up to 8 dB greater than the initial signal and the brain would still use the first (but quieter) signal to decide from where the sound originated. This effect is known also as 'the law of the first wavefront', 'the auditory suppression effect', 'the first arrival effect' and 'the threshold of extinction', and is used as the basis of sound reinforcement used in public address systems.

The brain attributes greater relative importance to *time* information as opposed to *intensity* information. For example, an early paper of Snow[2] describes experiments on compensating differences in left-right intensity balance using relative left-right time delays. It was reported that a 1 ms time delay would balance as much as 6 dB of intensity misbalance. It seems likely that the Haas effect has evolved as a valuable survival mechanism so that the brain can cope with multiple reflections in an enclosed space without confusion, thus rapidly distinguishing the true sound source from the array of first-order sound images.

As a direct consequence of this phenomenon, the sweet spot for intensity-based systems is quite small. If the listener were to sit slightly closer to one loudspeaker than the other, then the intermediately panned sources would appear to originate from the closer loudspeaker, rather than being spread over an area between the two. A centrally panned sound would collapse onto the nearest loudspeaker (Figure 1). Although a sweet spot is also an intrinsic feature of 3D Positional Audio, it is more robust, especially at lower frequencies.

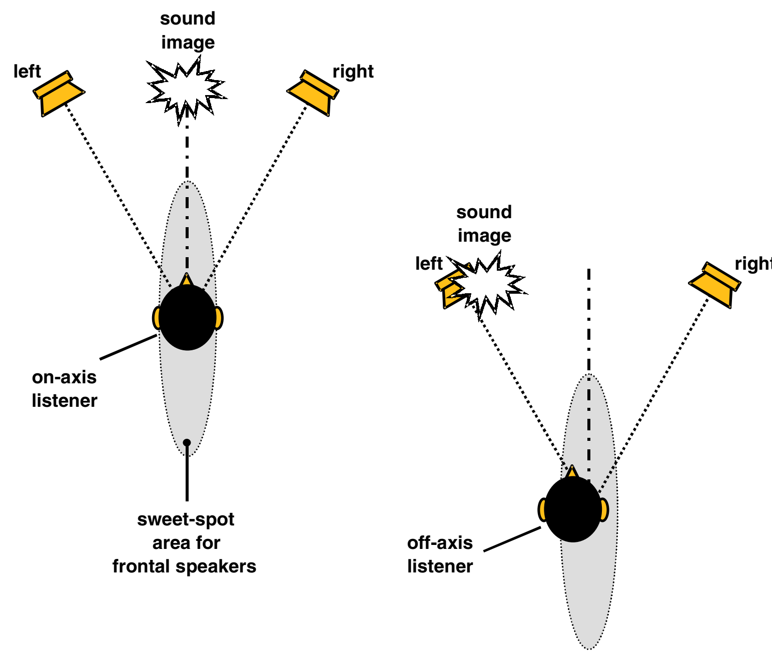



**Figure 1: Dependence of frontal image on listener position for an intensity-based system**

If we consider a four-speaker intensity-panned system, rather than a two-speaker arrangement, then the sweet spot constraints are even greater because it is limited in both lateral and front-back directions. The Haas effect always *pulls* the panned sound image to the closest loudspeaker. For example, if a sound has been panned midway between the front-left and rear-left speakers, then it might appear so if the listener is exactly equidistant

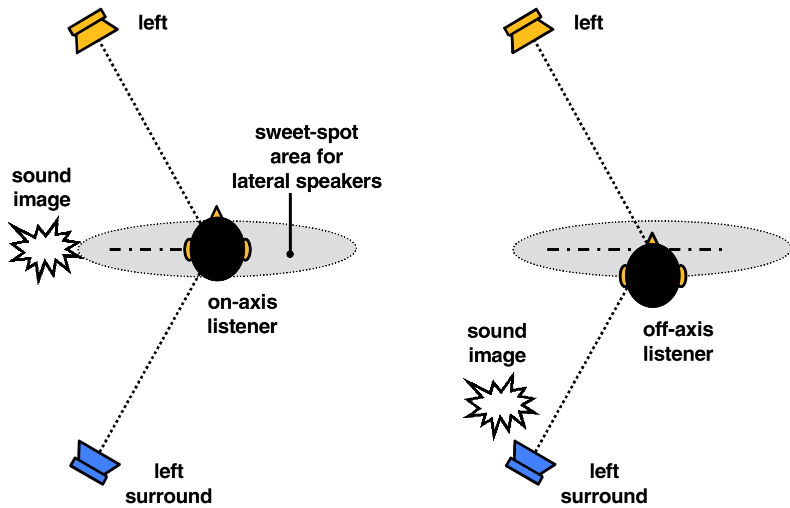



**Figure 2: Dependence of lateral image placement on listener position**

from the speakers. However, if the listener were to move, say, backwards by a foot, then the sound image would tend to collapse onto the left-rear speaker (Figure 2). Whereas *hard-panned* sounds (i.e. where virtually all the sound energy originates from a single loudspeaker) correctly appear to come from their respective speakers, all of the intermediately panned sounds appear to originate from the *closest* of their respective speakers, which fragments the overall sound picture.

In short, the sweet spot of a four-speaker intensity-based system is defined by the overlap between its frontal and lateral sweet spot regions (Figure 3).

Some final points worthy of note are that the use of intensity panning between four loudspeakers does not allow accurate placement of sound sources and prevents smooth, linear movement. (This can be tested using a panned system either by attempting to move a sound source smoothly around the listener in the horizontal plane or by placing the sound source directly on the left- or right-hand side.)

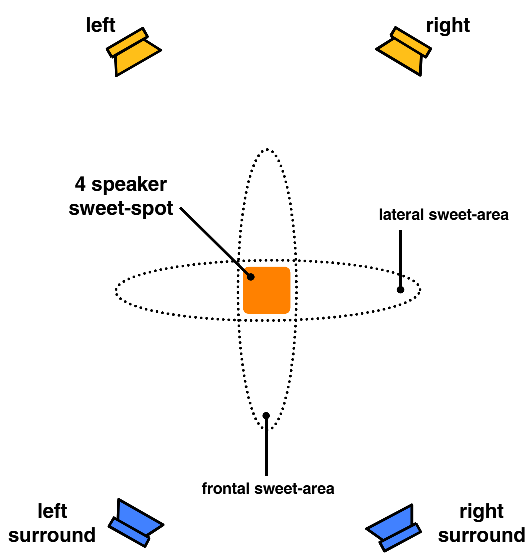



**Figure 3: Sweet spot locus for intensity-based four speaker system**

## 3 3D Positional Audio

True 3D Positional Audio systems are based on the processing of mono sound sources via *Head-Related Transfer Function* (HRTF) filters, such that the resultant stereo pair signal contains natural 3D sound cues. (These cues are more fully described in another technical white paper in the present series *An introduction to sound and hearing.)* In real life, these sound cues are introduced naturally by the head and ears when we listen to sounds. They include the inter-aural amplitude difference (IAD), inter-aural time difference (ITD) and spectral shaping by the outer ear. When this binaural signal pair is introduced directly 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.

If loudspeaker playback is used (rather than headphones), then transaural crosstalk occurs. This means that the left ear hears a little of what the right ear is hearing (after a small, additional time delay of around 0.25 ms) and *vice versa*. In order to prevent this happening, appropriate *crosstalk cancellation* signals can be created 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. More advanced schemes exist which anticipate and prevent the effects of the cancellation signals themselves contributing to secondary crosstalk. This topic is described in some detail in another technical white paper *Transaural acoustic crosstalk cancellation*.

A high quality 3D Positional Audio technology, such as Sensaura, will reproduce virtual sound sources in all three dimensions, providing a stunning, immersive listening experience from only two conventional loudspeakers (or headphones).

## 4 Why use more than two speakers for 3D Positional Audio?

A high quality 3D positional system requires only two loudspeakers so why bother using two more? There are several reasons why this can be worthwhile:

(a) reinforcement of *weak* cue areas;

(b) allowance for limitations of transaural crosstalk cancellation;

(c) support for existing multi-speaker owners.

### Reinforcement of *weak* cue areas

When the HRTF processing and crosstalk cancellation are carried out sequentially 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.

However, it is more difficult to synthesise sounds in certain spatial positions than in others, some for psychoacoustic reasons, we believe, and some for practical reasons. For example, the effectiveness of sound sources moving directly upwards and downwards is greater at the sides of the listener (azimuth = 90°) than directly in front (azimuth = 0°). This is probably because there is more left-right difference information for the brain to work with. Similarly, it is difficult to differentiate between a sound source directly in front of the listener (azimuth = 0°) from a source directly behind the listener (azimuth = 180°). This is because there is no time domain information present for the brain to operate with (ITD = 0). The only other information available to the brain (spectral data) is similar in both of these positions. In practice, sounds appear *brighter* when the source is in front of the listener, because the high frequency sounds from frontal sources are reflected into the auditory canal from the rear wall of the concha, whereas from a rearward source they are occluded by the pinna.

Front-back discrimination can be improved by the use of a rearward pair of loudspeakers to reinforce the weak cue area at 180° azimuth.

### Allowance for limitations of crosstalk cancellation

In practical terms, a limiting feature in the reproduction of 3D sound from two loudspeakers is the adequacy of the transaural crosstalk cancellation, especially at higher frequencies.

In practise, of course, it is not possible to carry out absolutely perfect crosstalk cancellation. Setting aside the differences between individual listeners and the artificial head from which the algorithm's HRTFs derive, the difficulties relate to the HF components above several kHz. When optimal cancellation is arranged to occur at each ear of the listener, the crosstalk wave and the cancellation wave combine to form a node. However, the node exists only at a single point in space and, as one moves further away from the node, the two signals are no longer mutually time aligned and so the cancellation is imperfect. Consequently, if the listener were to move away from the sweet spot, then the crosstalk cancellation would become less effective.

When the listener's head is not perfectly positioned, any imperfectly cancelled HF components become localised on the speakers themselves, which makes them appear to originate *in front* of the listener, making rearward synthesis more difficult to achieve.

Previously, there have been several attempts to limit the spatial dependency of crosstalk cancellation systems at the higher frequencies. One reported scheme introduces a high-cut filter into the crosstalk cancellation path, so that the HF components (>8 kHz or so) are not actually cancelled at all, but fed directly to the loudspeakers, just as they are in ordinary stereo. The problem with this approach is that the brain localises the HF sounds on the loudspeakers themselves, because both ears hear correlating signals from each individual speaker. Although these frequencies can be difficult to localise accurately, the overall effect, nevertheless, is to create HF sounds of *frontal* origin for *all required spatial positions*. Once again, this inhibits the illusion when trying to synthesise rearward positioned sounds.

### Support for existing multi-speaker owners

Multi-loudspeaker 3D sound systems are at a disadvantage over two-speaker systems for the obvious reasons of cost, wiring difficulties and the need for extra audio drivers. Nevertheless, a proportion of multimedia enthusiasts will already possess, or will buy, a four (or more) speaker configuration to cater for alternative formats, such as Dolby Digital™. These formats, of course, are 2D 'surround' systems that are incapable of true 3D source placement. Sensaura MultiDrive has been devised to provide a true 3D listening experience for multi-speaker users.

Many sound sources which are used in PC games contain predominantly low-frequency energy (explosion sounds, for example, and crash effects) and so the limitations referred to earlier are not so important because the transaural crosstalk cancellation is adequate for these long wavelength sources. However, if the sound sources were to contain predominantly higher-frequency components, such as bird song, and especially if they comprise relatively pure sine wave type sounds, then it would prove more difficult to provide effective crosstalk cancellation. Bird song, insect calls and the like, can be used to great effect in a game to create an effect of ambience and it is often required to position such HF-rich sounds in the rearward hemisphere. This can be achieved to great effect using Sensaura MultiDrive.

## 5 Sensaura MultiDrive

Sensaura 3D Positional Audio technology now incorporates a patented multi-speaker capability for creating 3D sound fields using four or more loudspeakers. This enables the playback of DirectSound3D material and provides improved effectiveness of rearwardly placed virtual sound sources. This has been achieved by using each speaker pair to create a complementary sound field, addressing the frontal and rearward hemispheres respectively.

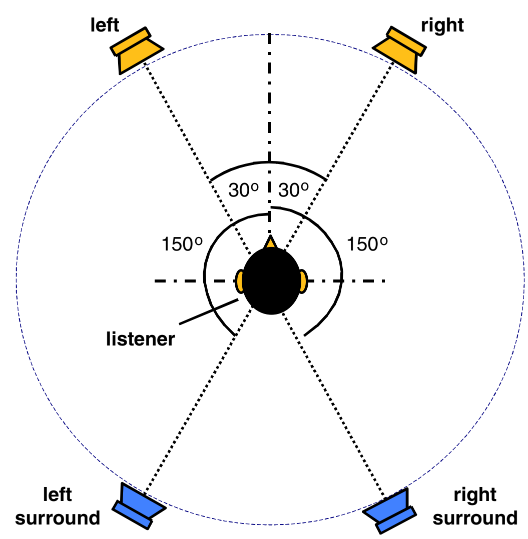



**Figure 4: Typical four loudspeaker configuration**

This creates very robust 3D audio, even in difficult listening environments, and it is especially valuable in rendering rearward virtual sound sources which are rich in HF. Figure 4 shows a typical four speaker setup.

The angles subtended at the listener's head by the four loudspeakers are usually taken to be

±30° (and their complements, ±150°), although they can be adjusted in software for various other angles. The volume levels are set so that all four loudspeakers sound equally loud to the listener. Front-back balance is not critical, but it is important to set a good lateral balance. The rear speakers are driven via a special crosstalk cancellation algorithm, appropriate for their angular position. The frontal sound field is created in the usual manner, but it is configured so as to be frontally-biased, as indicated in Figure 5.

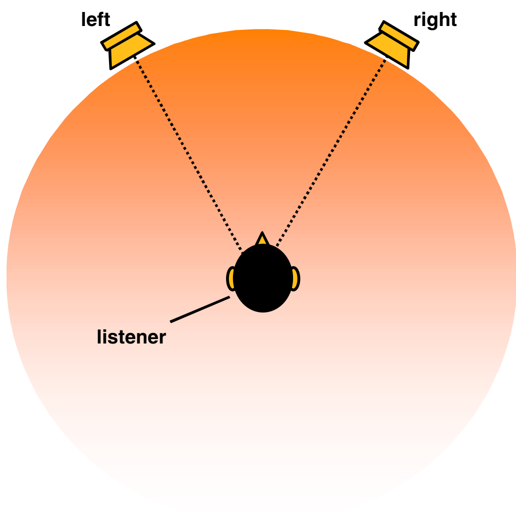



**Figure 5: Frontal sound-field creation**

The frontal biasing is carried out using a proprietary algorithm. Similarly, the rearward sound field is created using the rear loudspeakers and it is configured to be rearward-biased, as indicated in Figure 6.

Because each of the Sensaura sweet spots is tolerant laterally and has high tolerance for front-back movement for frontal (and rearward) speakers, the sweet areas for frontal and rearward speakers are quite large and *have a high degree of overlap*. Consequently, the final sweet spot covers a much greater area (Figure 7) than does the intensity-based sweet spot (Figure 3).

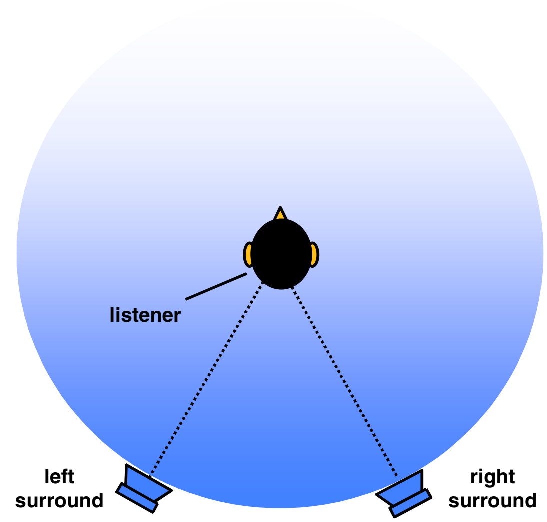



**Figure 6: Rearward sound-field creation**

The frontal and rearward sound hemispheres are seamlessly integrated together (Figure 8) such that any number of virtual sound sources can be made to move around the listener, travelling smoothly between the two, controlled by DirectSound3D.

Sensaura MultiDrive uses patented Sensaura TCC technology, which ensures optimal crosstalk cancellation. Sensaura TCC is also user-configurable for the best possible

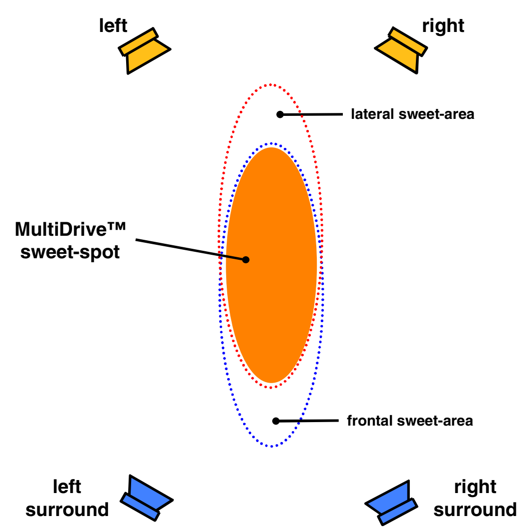



**Figure 7: Sweet spot locus for Sensaura MultiDrive**

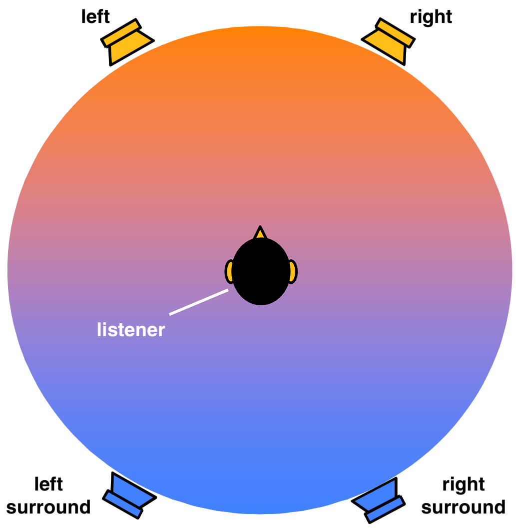



**Figure 8: Integrated duplex sound-field of Sensaura MultiDrive**

listening experience. Now, multi-speaker system owners are not restricted to flat, 2D sound panning, but can enjoy the full benefits of immersive 3D sound.

## 6 Benefits of MultiDrive compared with other systems

The main benefits of Sensaura MultiDrive, compared to multi-speaker systems, are as follows:

- ❑ completely solid rearward positioning, including HF-rich sound sources;
- ❑ large sweet spot (much better than intensity-based multi-speaker systems);
- ❑ true, immersive 3D effects (not just surround), including source elevation and depression ('height');
- ❑ insignificant additional processing capacity required;
- ❑ crisp, accurate sound images (not vague spatial effects);
- ❑ smooth, dynamic movement (rather than corner-to-corner panning);
- ❑ enables use of Sensaura MacroFX for close-up sound effects;
- ❑ enables use of Sensaura ZoomFX™ for large area sound sources and scaling;
- ❑ supports Environmental Audio extensions (EAX).

## 7 Suggested tests for multi-speaker systems

How is it possible to compare one multi-speaker system with another? Here, below, are some suggestions for tests that can easily be applied using a spatial *exerciser* program. If your system was not supplied with one, the Sensaura *Athene* program is freely downloadable from **www.sensaura.co.uk**. This can be used to move a sound source (in the form of a .WAV file) around spatially, either under mouse control or on a pre-set trajectory. For the sound source, use music with rich high frequency content, rather than a sound effect .WAV (such as the ubiquitous helicopter). Music will show up any position-dependent tonal changes, will test the crosstalk capability properly and will be a good test of front-back separation.

### Test #1: horizontal circle

Move the sound source around you in a complete horizontal circle, starting at a frontal position. Does the image go behind you satisfactorily? Are there any tonal changes?

### Test #2: vertical circle

Move the sound source around you in a complete vertical circle, starting in a lateral position, say on your right-hand side, travelling over your head to your left side, and then underneath you back to the right-hand side. Does the image go above you and underneath you satisfactorily?

### Test #3: lateral positioning

Try and position the sound source on, say, your left-hand side (azimuth -90°, with no elevation or depression). Is this easy to do or does the image tend to *flip* from front to back? Does it sound as though the image is in one position or is it smeared out? Is the image present both at the front and back simultaneously?

### Test #4: sweet spot lateral stability

With the sound source central (azimuth 0°, no elevation or depression), move your head to the left and right about one foot in either direction (keep facing forwards). When you move, does the sound source stay in one position or does it *collapse* onto the nearest frontal loudspeaker?

### Test #5: sweet spot front-rear stability

With the sound source on your left-hand side (azimuth -90°, with no elevation or depression), move your head forwards and backwards about one foot in either direction (keep facing forwards). When you move, does the sound source stay in one position or does it *collapse* onto the nearest left-hand (front or rear) speaker?

## 8 References


1. Historical Background of the Haas and/or Precedence Effect.
   M B Gardner
   J. Acoust. Soc. Am., 43, (6),
   pp.1243-1248 (1968).

2. Effect of Arrival Time on Stereophonic Localization.
   W B Snow
   J. Acoust. Soc. Am., 26, (6),
   pp.1071-1074 (1954).



**For further information please contact:**

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

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

A Scipher Company





DEVPC/006/0001/1