



Audio Engineering Society Convention Paper

Presented at the 118th Convention
2005 May 28–31 Barcelona, Spain

This convention paper has been reproduced from the author's advance manuscript, without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. Additional papers may be obtained by sending request and remittance to Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA; also see www.aes.org. All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Objective Assessment of Spatial Localisation Attributes of Surround-Sound Reproduction Systems

Martin Dewhurst^{1,2}, Slawomir Zielinski¹, Philip Jackson², Francis Rumsey¹

¹*Institute of Sound Recording, University of Surrey*

²*Centre for Vision Speech and Signal Processing, University of Surrey*

Correspondence should be addressed to Martin Dewhurst (M.Dewhurst@surrey.ac.uk)

ABSTRACT

A mathematical model for objective assessment of perceived spatial quality was developed for comparison across the listening area of various sound reproduction systems: mono, two-channel stereo (TCS), 3/2 stereo (i.e., 5.0 surround sound), Wave Field Synthesis (WFS) and Higher Order Ambisonics (HOA). Models for mono, TCS and 3/2 stereo are based on conventional microphone techniques and loudspeaker configurations for each system. WFS and HOA models use circular arrays of thirty-two loudspeakers driven by signals derived from a virtual microphone array and the Fourier-Bessel spatial decomposition of the soundfield respectively. Directional localisation, ensemble width and ensemble envelopment of monochromatic tones, extracted from binaural signals, are analysed under a range of test conditions.

1. INTRODUCTION

Previous studies into the perceived spatial quality of sound reproduction systems have concentrated primarily on the “sweet spot” in the centre of the listening area (e.g. Pulkki et al. [10]). Studies that have considered spatial quality at other points have tended to use simple averaging techniques, rather than a complete rendering of the reproduced sound field. Good reproduction is ideally required at all

points in the listening area, for example with home cinema systems. Hence, it is desirable to measure spatial quality across the entire listening area for given sound reproduction systems. This allows different types of sound reproduction systems to be compared objectively, and can also assist the measurement and design of new reproduction systems and panning laws, or efficient coding of spatial audio signals.

The principal aim of the mathematical model employed here is the graphical representation of spatial sound quality in reproduced sound fields. Given a description of an acoustic scene and a sound reproduction system, the model computes and displays perceived attributes of spatial quality at any point in the reproduced soundfield. For instance, an algorithm to deduce the directional localisation of a sound source from binaural signals [7, 9] provides an objective measure of localisation.

2. METHODOLOGY

The first stage of the model simulates the soundfield synthesised by one of the reproduction systems. Mono, two-channel stereo (TCS), 3/2 stereo (also known as 5.0 surround sound), Wave Field Synthesis (WFS) and Higher Order Ambisonics (HOA) were considered for this study. This stage can be further divided into modelling the original soundfield, generating the loudspeaker feed signals for a given sound reproduction system, and modelling the soundfield that is produced by the loudspeakers.

2.1. Definition of original soundfield

The types of original soundfield that were studied consist of a plane wave or a spherical wave emanating from a point source. In either case, the source was monochromatic, i.e., a sinusoid of a single frequency f , and multiple sources constructed by superposition.

The pressure due to a plane wave P_ψ in the horizontal plane with angle of incidence ψ at time t is

$$P_\psi(\vec{r}, t) = Ae^{j(\omega t - kr \cos(\phi - \psi))}, \quad (1)$$

where $\vec{r} = (r, \phi)$ is the position vector relative to an origin at the centre of the listening space, $k = 2\pi f/c$ is the wave number, $\omega = 2\pi f$ the angular frequency, A a complex constant and c the speed of sound. Similarly, the pressure due to a spherical wave $P_{\vec{r}_0}$ emanating from a point source at $\vec{r}_0 = (\rho, \psi)$ is given by

$$P_{\vec{r}_0}(\vec{r}, t) = \frac{A}{|\vec{r} - \vec{r}_0|} e^{j(\omega t - k|\vec{r} - \vec{r}_0|)}. \quad (2)$$

2.2. Loudspeaker drives

The HOA loudspeaker feed signals were derived from the Fourier-Bessel coefficients of the original, or target, soundfield. The loudspeaker feed signals for

all other reproduction systems were obtained using computational models of microphones (virtual microphones) to record the sound pressure at points within the original soundfield.

2.2.1. Mono, two-channel and 3/2 stereo

The loudspeaker feed signals for the mono, TCS and 3/2 stereo sound reproduction systems were generated using virtual microphones. The microphones captured the simulated sound pressures in the original soundfield for each reproduction system. For mono, a single omnidirectional microphone was placed in the centre of the listening area. For TCS, two microphones with cardioid directivity were arranged in the ORTF configuration (in a Y-shape). For 3/2 stereo, an array of cardioid microphones was used, as shown in Figure 1 (see Williams and Le Dû [15]). In all of the above cases, the signal captured by each microphone was used to drive the corresponding loudspeaker in the simulated sound reproduction system.

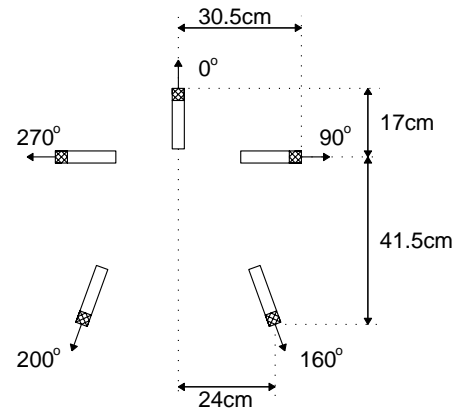


Fig. 1: Symmetrical 3/2 stereo microphone configuration.

2.2.2. Wave Field Synthesis

The mathematical basis of Wave Field Synthesis (WFS) is the Kirchhoff-Helmholtz equation:

$$p(\vec{r}) = \iint_{\partial A} \left[\vec{\nabla} p_0 \cdot \vec{n} - \frac{\vec{R}}{R} \cdot \vec{n} (1 + jkR) \frac{p_0}{R} \right] \frac{e^{-jkR}}{4\pi r} dS_0 \quad (3)$$

where A is a given volume and ∂A its boundary, $p(\vec{r})$ is the pressure at a point $\vec{r} \in A$, p_0 is the acoustic

pressure on ∂A , \vec{r}_0 is a point on ∂A , dS_0 is an area on the boundary ∂A centred on \vec{r}_0 , $\vec{n}(\vec{r}_0)$ is the unit vector perpendicular to ∂A , $\vec{R} = \vec{r}_0 - \vec{r}$, and $R = |\vec{R}|$. The Kirchoff-Helmholtz equation states that the pressure inside a closed volume A is uniquely defined by the pressure and pressure gradient on the boundary of A . WFS exploits this principle by first using a microphone array to record the pressure and pressure gradients on the boundary of a volume A and then using a loudspeaker array to recreate the pressure and pressure gradients of a similar volume A' , and thus recreate the pressure field inside A' .

Let the volume A be the infinite cylinder whose boundary is defined as $\vec{r} = r_0$ in cylindrical coordinates (r, ϕ, z) and ∂A be the boundary of the cylinder A . Let $p_0(\vec{r}_0)$ be the pressure field on the boundary ∂A , that is, $\vec{r}_0 = (r_0, \phi_0, z_0) \in \partial A$. By considering the Kirchhoff-Helmholtz integral in cylindrical co-ordinates, using the *Stationary Phase Approximation* [8, 13] and supposing the pressure field p_0 is an horizontal plane wave,

$$p_0(\vec{r}_0) = ae^{jk r_0 \cos(\phi_0 - \phi)} \quad (4)$$

where a and ϕ are the amplitude and angle of incidence of the plane wave respectively, then $\forall \vec{r} \in A$

$$p(r, \phi) = \sqrt{\frac{2\pi}{jk}} r_0 \int_0^{2\pi} \left[\frac{\partial p_0}{\partial r_0} - \cos \alpha (1 + jkR) \frac{p_0}{R} \right] \frac{e^{-jkR}}{4\pi\sqrt{R}} d\phi_0,$$

where

$$\begin{cases} R = \sqrt{r^2 + r_0^2 - 2rr_0 \cos(\phi - \phi_0)} \\ \cos \alpha = \frac{r \cos(\phi - \phi_0) - r_0}{R}. \end{cases}$$

This means that, if just the plane $z_0 = 0$ is considered, then the soundfield in an area in this plane is uniquely defined by the pressure and pressure gradient on the curve which is the boundary of the area. This makes it possible to use 2D arrays of microphones and loudspeakers for 2D realisations of Wave Field Synthesis systems, where typically the horizontal plane containing the arrays of microphones and loudspeakers is at the listener's head height.

Nicol and Emerit [8] discuss how the signals to the monopole and dipole loudspeakers are not independent, and that the soundfield can be approximated reasonably accurately using just monopole

loudspeakers and microphones. This can be achieved by using a circular array of cardioid microphones facing away from the sound source (e.g. Daniel et al. [2]), as in figure 2.

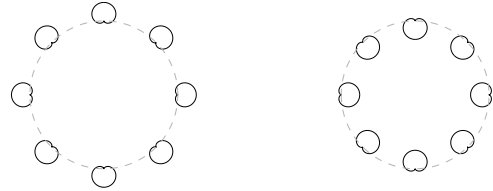


Fig. 2: Directivity patterns of an 8-microphone circular array for recording plane waves and point sources outside the array (left), and inside (right).

2.2.3. Higher Order Ambisonics

In a Higher Order Ambisonics (HOA) system the coefficients of the Fourier-Bessel decomposition are used in the reproduction of the soundfield. As the Fourier-Bessel decomposition of a soundfield has infinitely many terms, practical HOA systems truncate the decomposition after a finite number of terms, typically in the range ten to fifteen. If only a horizontal plane is considered then a cylindrical decomposition of the soundfield can be used, where the pressure is dependent only on the azimuth ϕ and the distance r from the origin, and each point in the horizontal plane is uniquely determined by an azimuth and distance from the origin, $\vec{r} = (r, \phi)$. Furthermore, if it is assumed that the area in which the soundfield is to be reconstructed is free of sources, then the truncated Fourier-Bessel decomposition is given by:

$$p(\vec{r}, t) = C_{00}^{+1}(t) J_0(kr) + \left[\sum_{m=1}^M j^m J_m(kr) \left(C_{mm}^{+1}(t) \sqrt{2} \cos m\phi + C_{mm}^{-1}(t) \sqrt{2} \sin m\phi \right) \right],$$

where $p(\vec{r})$ is the pressure at the point \vec{r} , t is time, k is the wave number, $J_m(kr)$ is the normal Bessel function (not the spherical Bessel function), C_{mm}^{+1} is the coefficient of the term involving the 2D harmonic $\sqrt{2} \cos m\phi$, and C_{mm}^{-1} involves $\sqrt{2} \sin m\phi$.

From [2] it can be shown that if N loudspeakers are arranged in a circular array of radius R around the origin of the horizontal plane, then the loudspeaker signals required to approximate the soundfield of a point source at a distance ρ from the origin and an azimuth ψ can be determined by

$$\mathbf{S}_c = \mathbf{E} \cdot \mathbf{H} \cdot \mathbf{C} = \frac{1}{N} \mathbf{D}^T \cdot \mathbf{H} \cdot \mathbf{C}, \quad (5)$$

where the n th element, $S_{c,n}(t)$, of the column vector \mathbf{S}_c is the pressure at the origin O caused by the n th loudspeaker at time t . The vector \mathbf{C} and matrix \mathbf{D} are defined as

$$\begin{aligned} \mathbf{C}(\vec{r}_0, t) &= [C_{00}^{+1}, C_{11}^{+1}, C_{11}^{-1}, C_{22}^{+1}, C_{22}^{-1}, \dots, \\ &\quad C_{mm}^{+1}, C_{mm}^{-1}]^T, \\ \mathbf{D} &= [\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_N], \text{ and} \\ \mathbf{d}_n &= \sqrt{2} \left[\frac{1}{\sqrt{2}}, \cos n \frac{2\pi}{N}, \sin n \frac{2\pi}{N}, \cos 2n \frac{2\pi}{N}, \right. \\ &\quad \left. \sin 2n \frac{2\pi}{N}, \dots, \cos mn \frac{2\pi}{N}, \sin mn \frac{2\pi}{N} \right]^T \end{aligned}$$

where

$$\begin{aligned} C_{qq}^{+1}(\vec{r}_0, t) &= \sum_{n=1}^N \sqrt{2} S'_n(t) \cos\left(\frac{2\pi qn}{N}\right) \\ C_{qq}^{-1}(\vec{r}_0, t) &= \sum_{n=1}^N \sqrt{2} S'_n(t) \sin\left(\frac{2\pi qn}{N}\right) \end{aligned} \quad (6)$$

and

$$\begin{aligned} \mathbf{H} &= \text{Diag} \left(\left[\dots \frac{F_m^{(\rho/c)}(\omega)}{F_m^{(R/c)}(\omega)} \dots \right] \right), \\ F_q^{(\rho/c)}(\omega) &= \sum_{n=0}^m \frac{(q+n)!}{(q-n)!n!} \left(\frac{-jc}{\omega\rho} \right)^n, \\ S'_n(t) &= A_n e^{j(\omega(t-R/c)+\alpha_n)}. \end{aligned} \quad (7)$$

2.3. Soundfield rendering to the ears

2.3.1. Reproduction of the soundfield

This section briefly discusses the Matlab model used to simulate the pressure fields created by arrays of loudspeakers. As has been already stated, the model currently only considers original soundfields which are monochromatic waves in a steady state. Consequently, each loudspeaker modelled in the system must also generate a monochromatic wave of the same frequency. Representing each loudspeaker n

as a point source, the pressure $P_n(\vec{r}, t)$ that it produces at the point \vec{r} and time t is given by:

$$P_n(\vec{r}, t) = \frac{A_n}{|\vec{r} - \vec{r}_n|} e^{j(\omega t - k|\vec{r} - \vec{r}_n|)}, \quad (8)$$

where A_n is the complex constant that determines the amplitude and phase of the wave, and $\vec{r}_n = (R, 2\pi n/N, 0)$ is the position of the n th loudspeaker. From this it can be seen that it is enough to determine A_n for each loudspeaker when attempting to recreate a monochromatic wave. The total pressure $P(\vec{r}, t)$ at a point \vec{r} and time t is

$$P(\vec{r}, t) = \sum_{n=1}^N P_n(\vec{r}, t). \quad (9)$$

All the pressures within the Matlab model are calculated as complex coefficients, of which only the real part corresponds to the instantaneous physical pressure. Maintaining the complex pressures allows easy calculation of phase and amplitude relations between points in the array, and the values A_n for each loudspeaker to be determined for any t .

The current version of the Matlab model does not take into account room reflections; the assumption is made that both the soundfield recording and the soundfield reproduction are occurring in anechoic conditions. Two examples of the reproduced soundfields generated at the end of this stage of the model are shown in figure 4.

2.3.2. Extracting binaural signals

Having reproduced the soundfield in the free field, we need to extract binaural signals for permissible head positions within the listening area in order to evaluate the sound's perceptual attributes. The listening area is taken as a disk in the horizontal (zero-elevation) plane with a 1.5 m radius. The listener's head can be anywhere within this disk, with any orientation. Thus, for any given head position, we can calculate the direction and distance of any source *relative to the listener's head*. Gardner and Martin's HRTF database of a KEMAR dummy head [3] was used to calculate the contribution of each virtual loudspeaker to the signals at the left and right ears of the listener. The HRTF database contains impulse responses for all azimuths at 5° intervals in the plane, from which a set of frequency responses was

calculated. However, the response at angles of virtual loudspeakers relative to the listener's ears were obtained by cubic spline interpolation of this set. The binaural signals from the reproduced soundfield were the sum of the binaural signal components from each loudspeaker.

2.4. Perceptual attributes

Three different perceptual attributes are considered, *directional localisation*, *ensemble width* and *ensemble envelopment*. The detailed definition of these attributes is presented in [11]. Other spatial perceptual attributes, such as source distance and source width, are not considered in this paper. This is because the perception of these spatial attributes relies on cues provided by room reflections and the onsets of time-variant signals, both of which are beyond the scope of the current model.

2.4.1. Directional localisation

Gerzon [4] summarises the psychoacoustics of sound localisation as depending on three different mechanisms whose interaction depends on the frequency of the sound being localised. Below 700 Hz, the listener's directional localisation is based primarily on the Interaural Time Differences (ITDs) between the signals arriving at the two ears. Interaural Level Differences (ILDs) are the primary means of localisation for the listener between 700 Hz and 5 kHz. Above 5 kHz the primary cues for directional localisation appear to depend on the colouration on the sound provided by the reflections of the pinnae (the external parts of the ears).

The Matlab model includes the imaginary parts of the signals, so values of the ILDs (in dB) can be computed directly as:

$$ILD = 10 \log_{10} \frac{\int_0^T \{p_R(t)\}^2 dt}{\int_0^T \{p_L(t)\}^2 dt} = 20 \log_{10} \frac{|p_R(t_0)|}{|p_L(t_0)|}, \quad (10)$$

where $p_R(t)$ and $p_L(t)$ are the pressure signals at time t at the right and left ears respectively, T is the integration time and t_0 can be any time at which the binaural signals are calculated. Figure 6 shows examples of the calculated ILDs.

The phases of complex pressure signals at the ears can be used to calculate the ITDs (in seconds):

$$ITD = \frac{\arg(p_R(t)) - \arg(p_L(t))}{\omega}, \quad (11)$$

where $p_R(t)$ and $p_L(t)$ are the complex pressure signals at the right and left ears respectively (so $\arg(p_R(t))$ and $\arg(p_L(t))$ are the phases at each ear) and ω is the angular frequency. It is assumed that the angle equal to the difference between the two phases is in the range $\pm 90^\circ$. Figure 5 shows examples of the calculated ITDs.

A lookup table is used to convert the ITDs to an angle of localisation in the horizontal plane. Using a similar method to the binaural signal extraction, for a given frequency the phase and magnitude responses are calculated for each impulse response with zero elevation in the Martin and Gardner HRTF database. The ITDs are calculated as the differences in phase responses of the signals at the left and right ears divided by the angular frequency, as in equation (11). This table of ITDs for angles at 5° intervals around the horizontal plane, together with cubic spline interpolation, can then be used to convert the ITDs already generated for the reproduced soundfield's binaural signals into a lateral angle of localisation. Figure 7 shows examples of the angles calculated from ITDs.

Two other methods of calculating azimuthal angles from ITDs were also implemented [6, 16], both of which are based on simplified models of the listener's head. The results from all three methods were broadly comparable, but the algorithm using the HRTF database was chosen for use in the model to ensure consistency with the earlier stage of the model where the HRTF database was used in the generation of the binaural signals.

A lookup table of ILDs is generated in a similar fashion for a given frequency, using the magnitude responses of the zero elevation impulse responses in the HRTF database and equation (10). The ILDs calculated from the reproduced soundfield's binaural signals can then be converted to a localisation angle using the lookup table and cubic spline interpolation. Figure 8 shows examples of the angles calculated from ILDs.

One limitation of the methods used in the model for converting ITD and ILD values to lateral angles is that the model can only allow angles in the range $\pm 90^\circ$. Sound sources in the rear hemisphere, i.e. behind the listener's head, have their angles erroneously located in the front hemisphere. This problem of front-back confusion is not unique to the the

present work: it also occurs with the Kuhn [6] and Woodworth [16] methods.

The front-back confusion of sound sources is much less common for real listeners, partly because the differences in the ILD and ITD cues caused by small movements of the listener's head allow the listener to disambiguate. The filtering effects of the pinnae at high frequencies (above 5kHz) also provide cues that help to reduce the front-back confusion. In future, the model may overcome the problem by simulating small movements of the listener's head.

2.4.2. Ensemble width

The term *ensemble* is defined as a number of individual sources within an acoustic environment that can be perceived by the listener as a meaningful entity, for example the instruments in a band [11]. The *ensemble width* is defined as the angle between the directional localisations of the two outer sources. The ensemble width is used to measure the perception of the width of an ensemble.

The model described in this paper currently only uses monochromatic, time-invariant waves as the source of the original soundfield. Thus, if there are two point sources both emitting monochromatic waves with exactly the same frequency, then a listener will perceive the resulting soundfield as being due to a single source. Indeed, this is one of the assumptions of multichannel sound reproduction systems. If it were possible to use more complex signals in the model, so the signals contained more than a single frequency and also varied over time, then it could be possible to decorrelate a pair of signals with nearly identical signals and localise the two signals simultaneously. As this is not possible with the current model, the ensemble width results presented in this paper have been generated by calculating the directional localisation of each of the two sound sources in isolation and then taking the difference between the two localisation angles to give the ensemble width.

2.4.3. Ensemble envelopment

Consider an ensemble of N sound sources. Let θ_n be the angle of localisation of the n th sound source in the ensemble, where $n \in \{1, 2, \dots, N\}$. Let P_n be the point on the circumference of a unit circle corresponding to the angle θ_n : $P_n = e^{j\theta_n}$. If adjacent

points P_n are connected with straight lines then the resulting polygon will have an area in the range 0 to π (see Figure 3). The *ensemble envelopment* is defined as this area divided by π , yielding a value in the range 0 to 1. Hence, when there are fewer than three different calculated directional localisations for an ensemble, the ensemble envelopment will be zero. At the other extreme, if the members of the ensemble are localised at equally-spaced angles around the head, the value of ensemble envelopment will approach 1 as N increases. Hence, ensemble envelopment provides some indication of the listener's perception of envelopment by an ensemble.

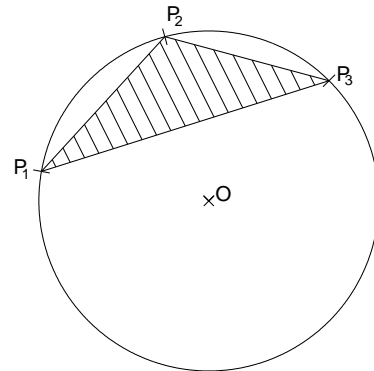


Fig. 3: Calculation of ensemble envelopment.

This crudely-defined measure of ensemble envelopment has not yet been validated with subjective listening tests. There are a few specific situations which suggest that the measure could be improved upon.¹

Similarly to the ensemble width method, the ensemble envelopment results presented here have been generated by calculating the directional localisation of each member of the ensemble in turn in isolation and then combining all the localisation angles

¹The first situation is when there are two sources located diametrically opposite each other (e.g. directly in front and directly behind the listener). The definition of ensemble envelopment defined above would give a value of zero, whereas a listener may perceive some envelopment. The second situation is when four sound sources are spaced equally in a circle around the listener. If the ensemble is increased to eight sound sources equally spaced around the listener then there will not be a great increase in the value of ensemble envelopment calculated using the definition above, whereas the listener may perceive a large increase in the perceived envelopment.

to give the final ensemble envelopment value.² As before, the ensemble envelopment can only be correctly calculated for ensembles whose members are all located in the front hemisphere. This sets an upper bound of $\frac{1}{2}$ on ensemble envelopment with the current version of the model.

3. RESULTS

All the WFS and HOA results presented in this section were created with a 32 loudspeakers placed in a circular array with radius 1.5m. The HOA results were generated using a 16th order Fourier-Bessel decomposition.

The values in figures 9 to 20 are all percentages of the listening area, and in each of these cases the listening area has been defined as the area bounded by the circle with radius 1.5m from the central listening point. This allows objective comparison, and corresponds to the area within the loudspeaker array for WFS and HOA.

The figures in sections 3.2 to 3.4 show results calculated from 100Hz to 12.8kHz for both ILDs and ITDs, although the frequencies at which listeners use the ITD and ILD cues as the principle means of directional localisation are below 700Hz and between 700Hz and 5kHz respectively. Currently the model does not attempt to fuse the ITD and ILD results to give a single angle of localisation (e.g., Supper et al. [12]).

3.1. Intermediate results

In this section, each of figures 4 to 8 shows two plots. The left hand plot is for a 3/2 stereo system and the right hand plot is for a 16th order 32-loudspeaker HOA system. In all of the figures the systems are trying to reproduce a 600Hz monochromatic wave emanating from a point source at the location $\vec{r} = (4\text{m}, 45^\circ)$. The red line on the loudspeakers shows this angle relative to the origin at the centre of the listening area. The listener's head depicted at the origin illustrates its orientation; the ITDs, ILDs and angles shown in figures 5 to 8 were calculated scanning the listener's head across all locations within the listening area.

²This method is used because of the limitations of only being able to use monochromatic, time-invariant signals in the model.

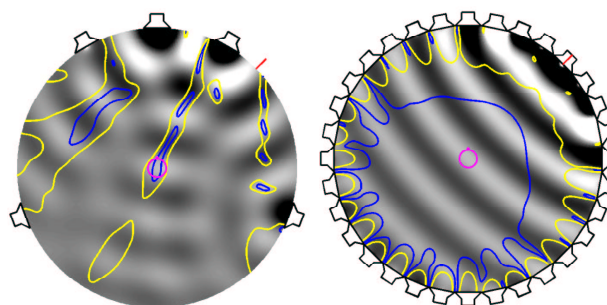


Fig. 4: Sound pressure plots for a reproduced monochromatic wave ($f = 600\text{Hz}$). The blue (dark) and yellow (light) contours enclose areas where the error between the original and reproduced soundfields is less than 20% and 50% respectively.

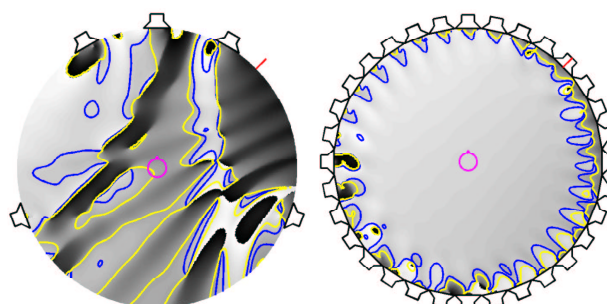


Fig. 5: ITDs for a reproduced monochromatic 600Hz wave, with 20% (blue, dark) and 50% (yellow, light) error contours.

3.2. Directional localisation

Figures 9 to 11 show the percentage of the listening area where the perceived directional localisation is within 5° of the expected angle of directional localisation. The blue lines correspond to the values calculated using ILDs and the red lines correspond to the values calculated using ITDs. In figure 9, the original sound source was a point source located straight ahead (i.e. at an angle of 0°) and 4m from the origin. The original sound source was located at $(4\text{m}, 45^\circ)$ for figure 10, and at $(4\text{m}, 90^\circ)$ for figure 11.

Across the three figures it can be seen that the ITDs generally give better localisation at lower frequencies than the ILDs. This agrees with Gerzon [4],

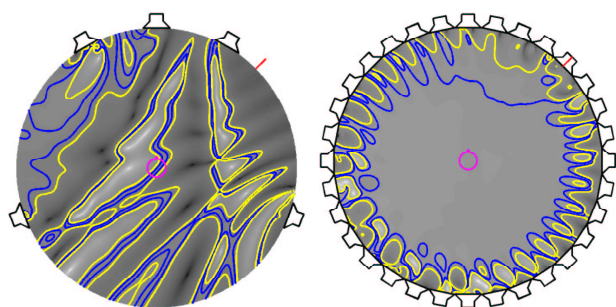


Fig. 6: ILDs for a reproduced monochromatic 600Hz wave, with 20% (blue, dark) and 50% (yellow, light) error contours.

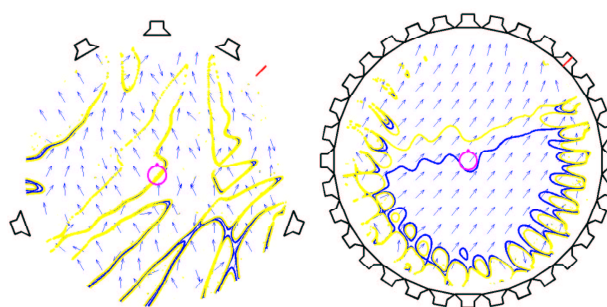


Fig. 8: Perceived angles calculated from ILDs for a reproduced monochromatic 600Hz wave, with 5° (blue, dark) and 10° (yellow, light) error contours.

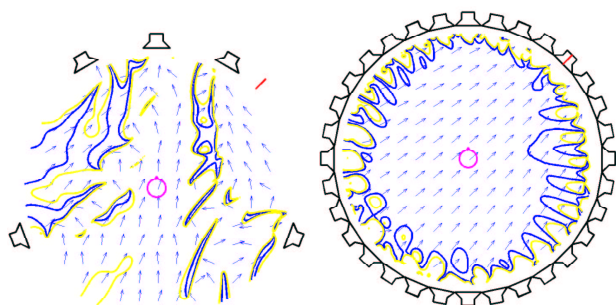


Fig. 7: Perceived angles calculated from ITDs for a reproduced monochromatic 600Hz wave, with 5° (blue, dark) and 10° (yellow, light) error contours.

who states that ITDs are the primary directional localisation cue for frequencies below 700Hz and that ILDs are for frequencies between 700Hz and 5kHz. The figures, especially figure 10, also show that the ILDs generally perform better than the ITDs at frequencies between 1kHz and 5kHz.

From figures 9, 10 and 11, it can be seen that WFS and HOA perform better than the mono, TCS and 3/2 stereo systems. When the source is directly in front of the listener (figure 9), the directional localisation is relatively good for TCS and 3/2 stereo, but as the the angle of the sound source location increases (figures 10 and 11), there is a degradation of the performance of the TCS and 3/2 stereo systems at low frequencies. The directional localisation performance of the WFS and HOA systems is more consistent than the other three systems at low frequencies as the angle of the sound source increases.

The percentage of the listening area where the calculated perceived angle within 5° of the expected localisation angle decreases with all five systems as the angle of the sound source moves toward 90°, to the side of the head. The poor performance is probably due more to the limits of the perception of directional localisation at these angles than to shortcomings in sound reproduction. This is particularly true of WFS and HOA with 32 loudspeakers, where the reproduced soundfield for a source at (4m,0°) is exactly a 90° rotation of the soundfield for a source at (4m,90°).

3.3. Ensemble width

Figures 12 to 16 show the percentage of the listening area where the perceived ensemble width is within 5° of the ensemble width (as defined in section 2.4.2, calculated using the locations of the original sound sources). Blue lines correspond to the values calculated using ILDs and red lines to those calculated using ITDs.

Figures 12 to 15 show the results for pairs of sources which have an angle of 30° between them; figure 16 shows the results for a pair of sound sources with an angle of 60° between them. The plots for the mono system in all four of these figures show that zero percent of the listening area had an ensemble width within 5° of the ensemble width based on the original positions of the two sources. This is to be expected, as the mono system will tend to give an ensemble width of zero for all signals, regardless of the original position of the sound sources. The TCS and 3/2 stereo plots are comparable across the first

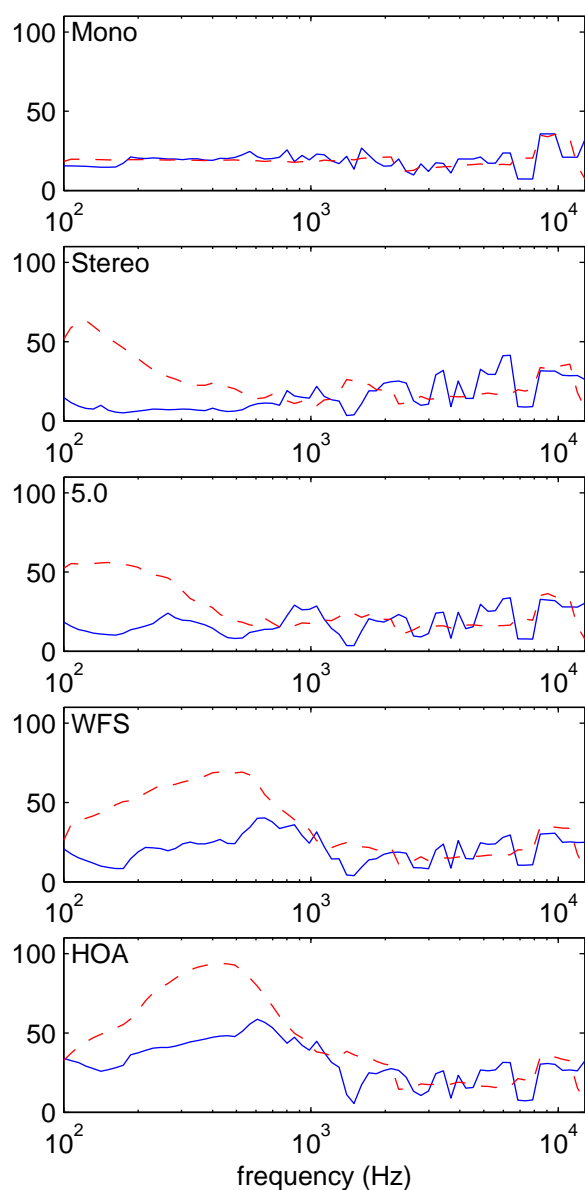


Fig. 9: Directional localisation, source at $(4m, 0^\circ)$. Percentage of listening area where the perceived angle of the reproduction sound system is within 5° of the expected angle, computed from ILDs (solid blue lines) and from ITDs (dashed red lines).

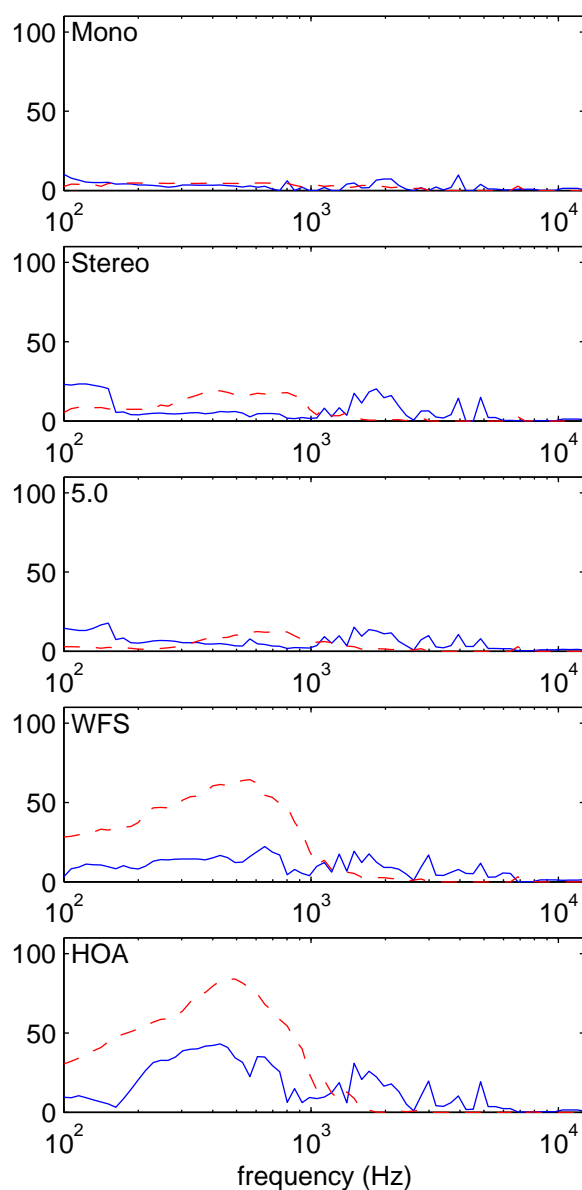


Fig. 10: Directional localisation, source at $(4m, 45^\circ)$. Percentage of listening area where the perceived angle of the reproduction sound system is within 5° of the expected angle, computed from ILDs (solid blue lines) and from ITDs (dashed red lines).

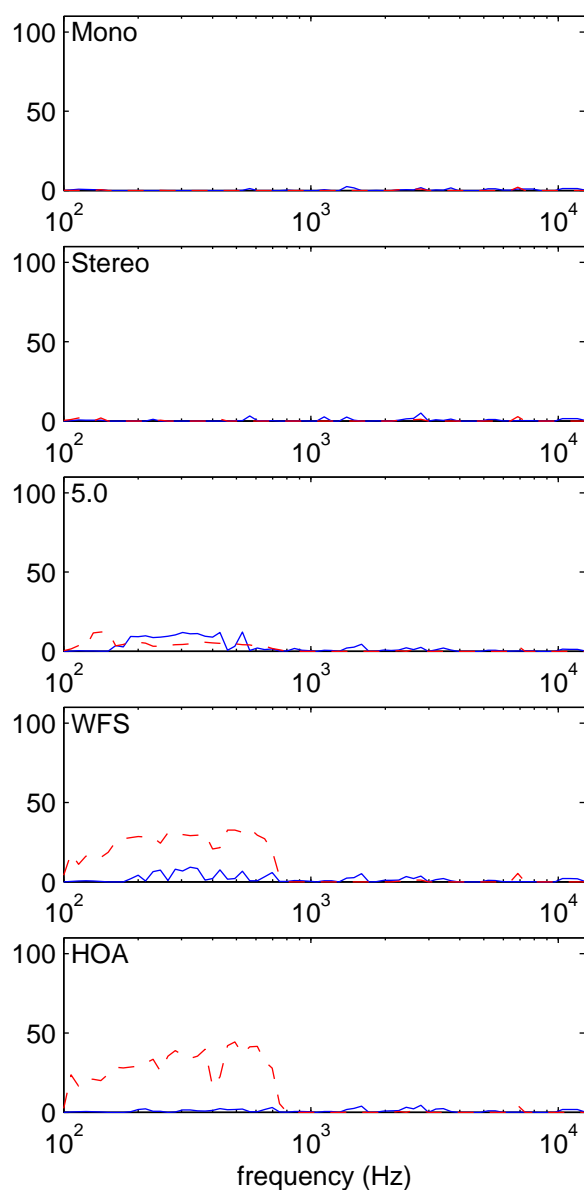


Fig. 11: Directional localisation, source at $(4\text{m}, 90^\circ)$. Percentage of listening area where the perceived angle of the reproduction sound system is within 5° of the expected angle, computed from ILDs (solid blue lines) and from ITDs (dashed red lines).

four figures, with less than 20% within 5° of the expected ensemble width, and slightly less for figure 16, where the expected ensemble width is larger.

From figures 12 to 16, it can be seen that WFS and HOA perform better than the other three sound reproduction systems, with HOA performing slightly better than WFS. When one of the pair of sound sources is at 90° then the ensemble width performance of the sound reproduction systems deteriorates, which agrees with the results for directional localisation and is probably also due to the limitations of the mechanisms of perceptual localisation at such angles. Excepting cases where one source is at an angle of 90° , the ensemble width performance of the systems appears to improve as the sources move around the head; figure 14 shows the ensemble widths for a pair of sound sources 4m away at 30° and 60° , which are the plots where the sound reproduction systems perform best.

3.4. Ensemble envelopment

Figures 17 to 20 show the percentage of the listening area where the perceived ensemble envelopment for each sound reproduction system is within 10% of the expected ensemble envelopment (as defined in section 2.4.3, calculated using the locations of the original sound sources). Blue lines correspond to the values calculated using ILDs and red lines to those calculated using ITDs. Similar to the directional localisation and ensemble width results, the HOA and WFS systems appear to perform best, with HOA slightly better than WFS.

The ensembles for figures 17 to 20 have all their members located within $\pm 90^\circ$. This is due to the inability of the current model to localise sources in the rear hemisphere, as discussed in sections 2.4.1 and 2.4.3. It is debatable whether the effect created by the ensemble containing the sources located at $(4\text{m}, -30^\circ)$, $(4\text{m}, 0^\circ)$ and $(4\text{m}, 30^\circ)$ will be perceived as a width or an envelopment by the listener. However, the results from this ensemble are shown in figure 17 for comparison with ensembles covering a larger area. The unexpectedly large values at higher frequencies shown in the plots for mono in figures 18 to 20 need further investigation.

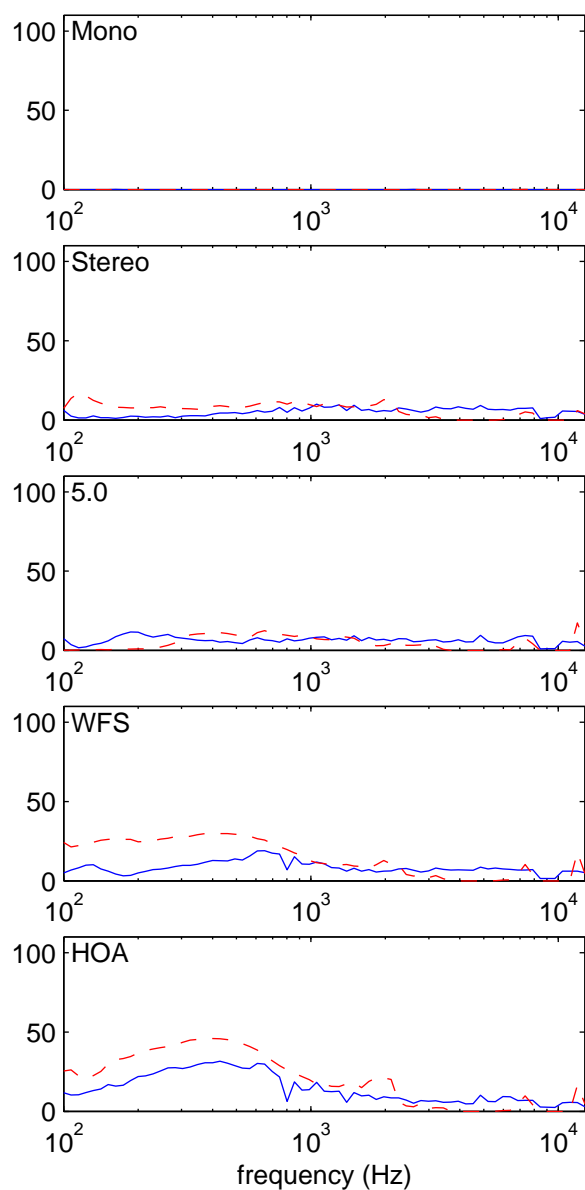


Fig. 12: Ensemble width. Percentage of listening area where the perceived ensemble width of the reproduction system is within 5° of the expected ensemble width, where the point sources are at the locations $(4\text{m}, 0^\circ)$ and $(4\text{m}, 30^\circ)$, computed from ILDs (solid blue lines) and from ITDs (dashed red lines).

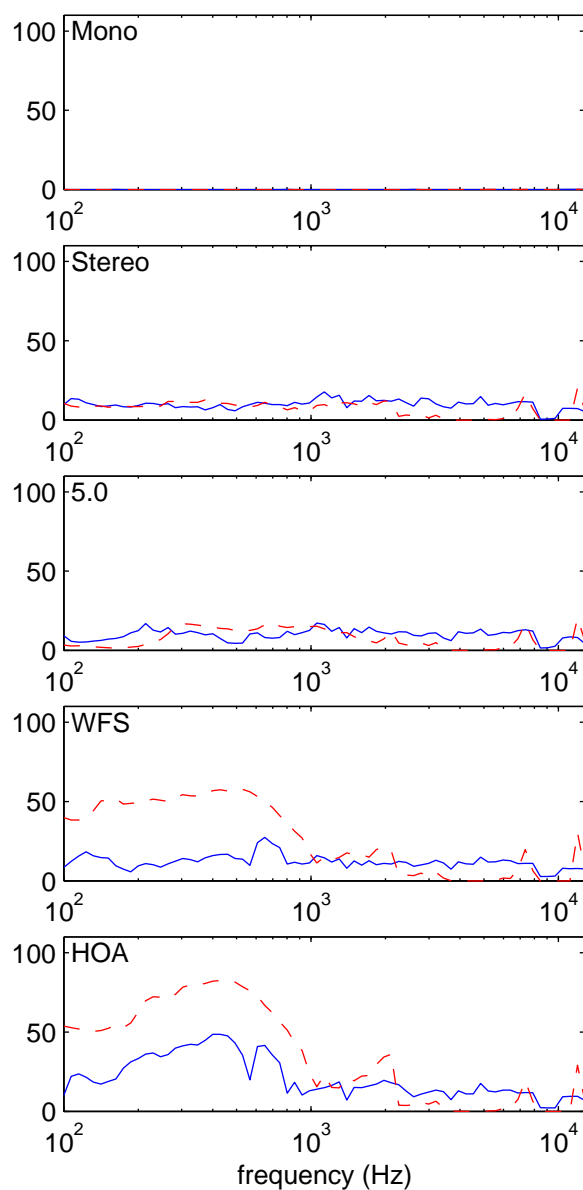


Fig. 13: Ensemble width. Percentage of listening area where the perceived ensemble width of the reproduction system is within 5° of the expected ensemble width, where the point sources are at the locations $(4\text{m}, 15^\circ)$ and $(4\text{m}, 45^\circ)$, computed from ILDs (solid blue lines) and from ITDs (dashed red lines).

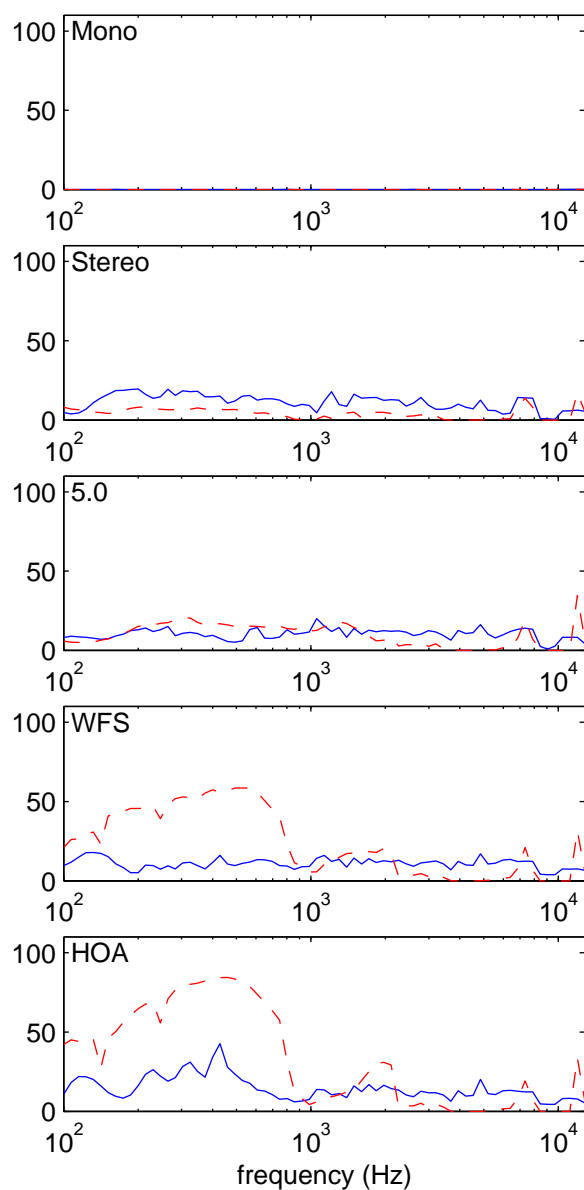


Fig. 14: Ensemble width. Percentage of listening area where the perceived ensemble width of the reproduction system is within 5° of the expected ensemble width, where the point sources are at the locations (4m, 30°) and (4m, 60°), computed from ILDs (solid blue lines) and from ITDs (dashed red lines).

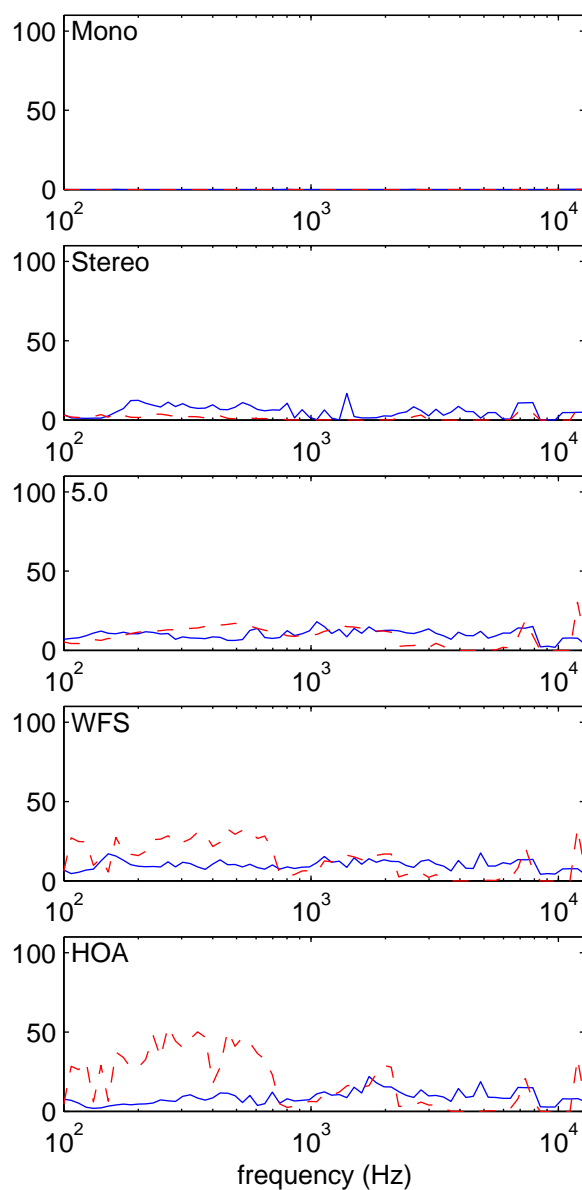


Fig. 15: Ensemble width. Percentage of listening area where the perceived ensemble width of the reproduction system is within 5° of the expected ensemble width, where the point sources are at the locations (4m, 60°) and (4m, 90°), computed from ILDs (solid blue lines) and from ITDs (dashed red lines).

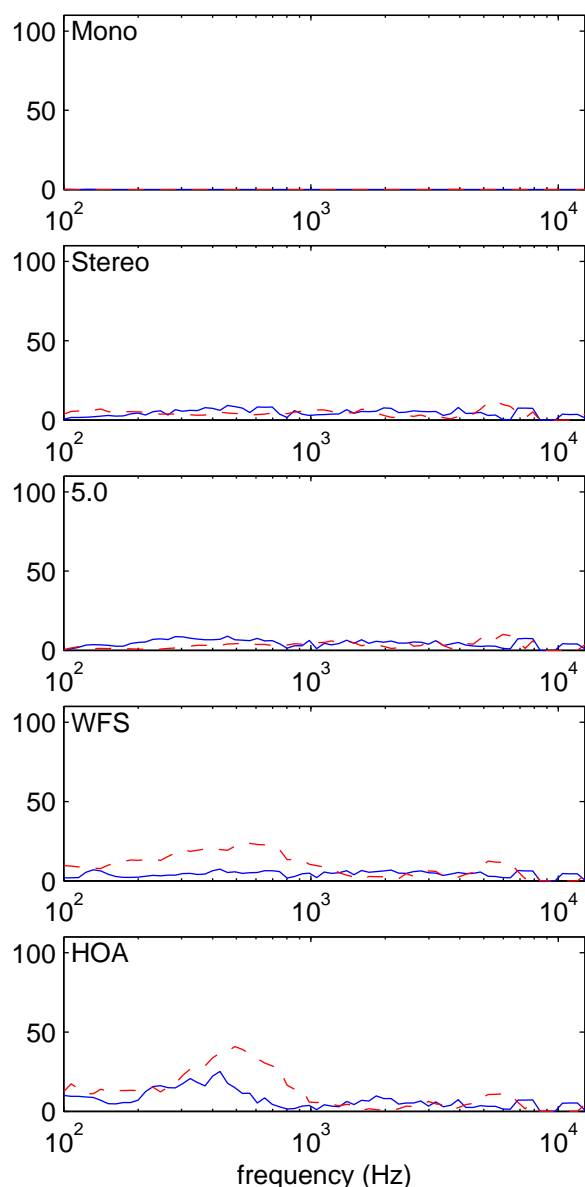


Fig. 16: Ensemble width. Percentage of listening area where the perceived ensemble width of the reproduction system is within 5° of the expected ensemble width, where the point sources are at the locations $(4\text{m}, 0^\circ)$ and $(4\text{m}, 60^\circ)$, computed from ILDs (solid blue lines) and from ITDs (dashed red lines).

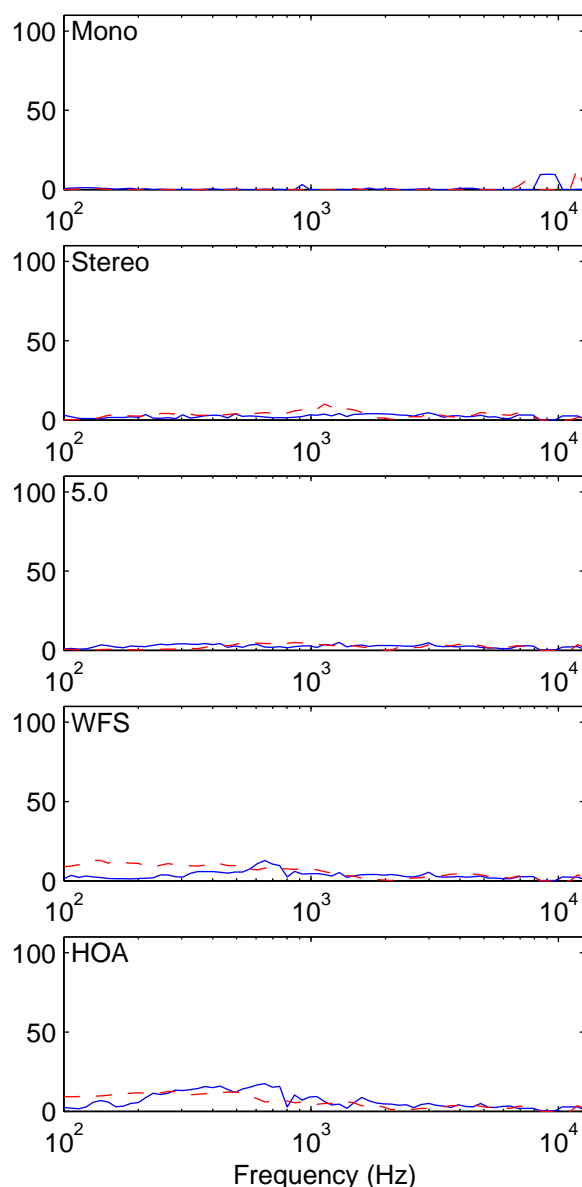


Fig. 17: Ensemble envelopment, sources at -30° , 0° and 30° . Percentage of listening area where the perceived ensemble envelopment of the reproduction system is within 10% of the expected ensemble envelopment for point sources 4m away, computed from ILDs (solid blue lines) and from ITDs (dashed red lines).

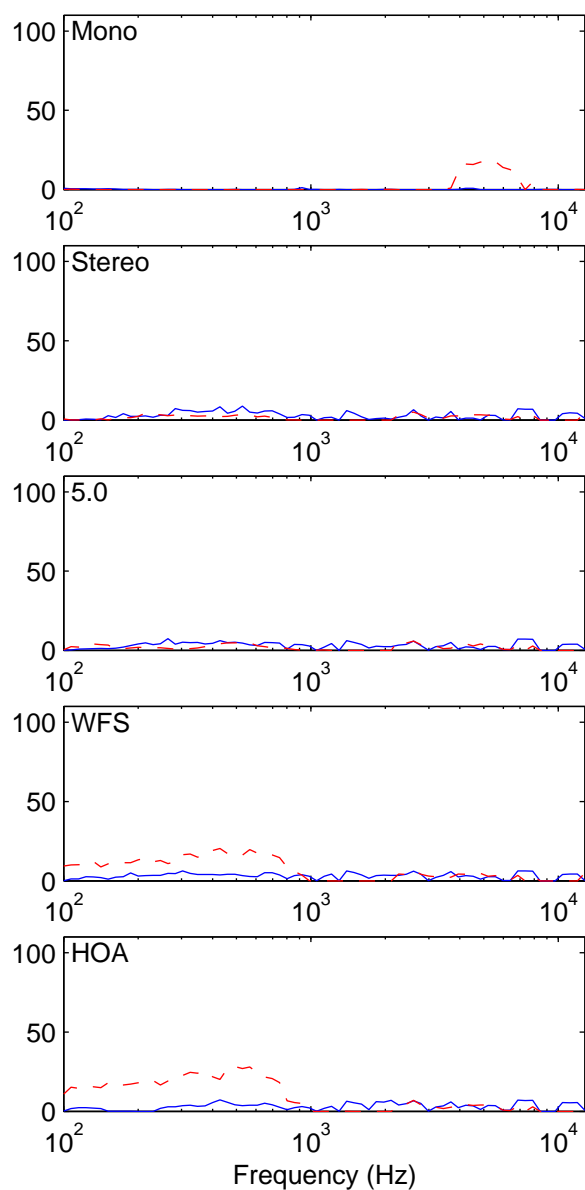


Fig. 18: Ensemble envelope, sources at -60° , 0° and 60° . Percentage of listening area where the perceived ensemble envelopment of the reproduction system is within 10% of the expected ensemble envelopment for point sources 4m away, computed from ILDs (solid blue lines) and from ITDs (dashed red lines).

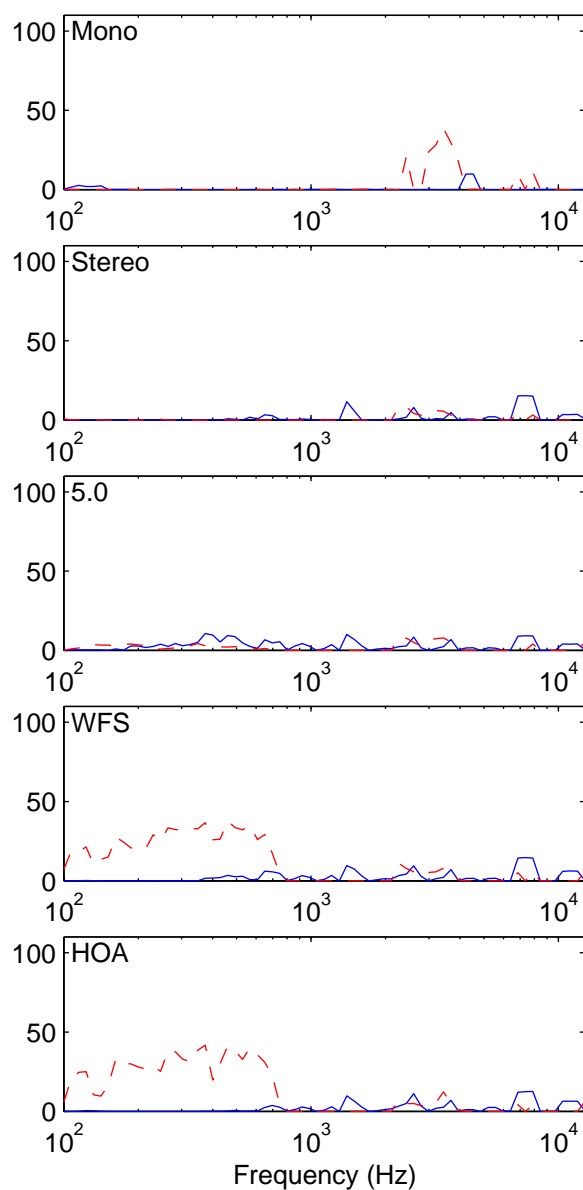


Fig. 19: Ensemble envelope, sources at -90° , 0° and 90° . Percentage of listening area where the perceived ensemble envelopment of the reproduction system is within 10% of the expected ensemble envelopment for point sources 4m away, computed from ILDs (solid blue lines) and from ITDs (dashed red lines).

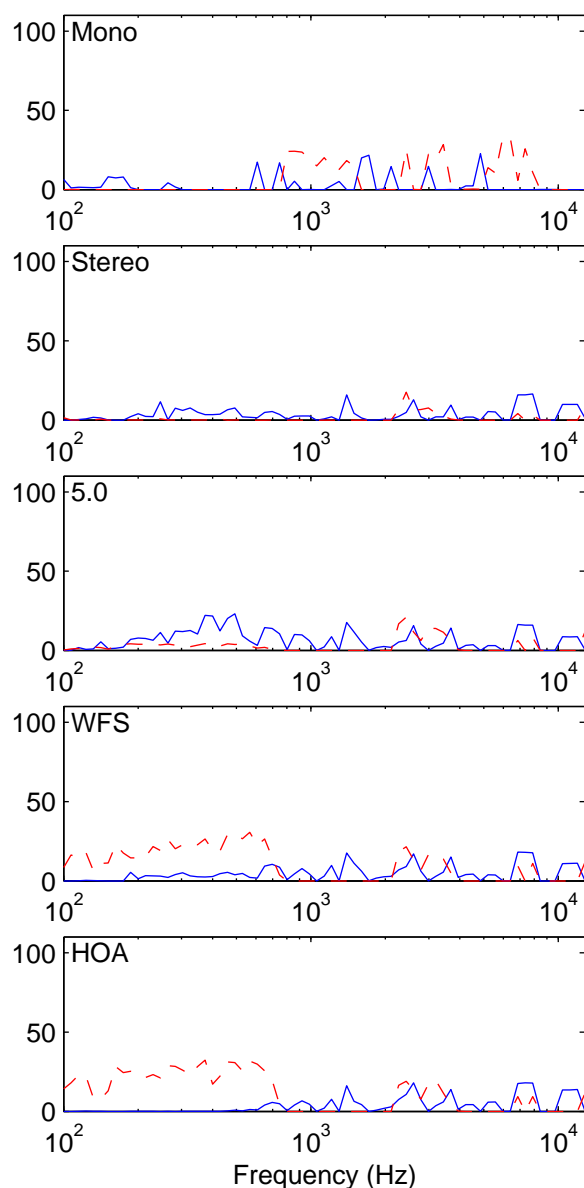


Fig. 20: Ensemble envelopment, sources at -90° , -45° , 0° , 45° and 90° . Percentage of listening area where the perceived ensemble envelopment of the reproduction system is within 10% of the expected ensemble envelopment for point sources 4m away, computed from ILDs (solid blue lines) and from ITDs (dashed red lines).

4. CONCLUSION

The model presented in this paper allows the graphical representation of three different objective measures of spatial sound quality in reproduced sound fields. The model, and the ensemble envelopment measure in particular, still requires verification with subjective listening tests. Examples of the output of the model have been given (figures 4 to 8) and plots summarising the results have been presented for five different sound reproduction systems: mono, two channel stereo, 3/2 stereo (5.0 surround), WFS and HOA. The results show that the WFS and HOA systems perform better and are much more consistent across a range of conditions than the other three sound reproduction systems.

However, the model presented in this paper can currently only handle monochromatic, time-invariant signals. Real sources contain more than a single frequency and vary with time. In particular, the initial transients and changes in the signal content at the onset of a sound provide important localisation cues (the precedence effect [1, 5, 14]). Cues from steady-state components of a signal (ITDs and ILDs) sometimes disagree with transient cues, which are used as the most reliable localisation cues by the listener. The two-channel and five-channel microphone techniques used in the model rely strongly on the precedence effect to create convincing stereo images. As the model presented in this paper only deals with monochromatic, time-invariant signals, the precedence effect does not affect the model's results. Hence, the perceptual effect of the two-channel stereo and 3/2 stereo systems will probably be improved with real world signals (i.e. more than one frequency and varying over time), relative to the WFS and HOA systems.

Another limitation of the model is that it assumes both the original and reproduced soundfields occur in anechoic conditions. With simulated room reflections, the model will become more realistic and allow the calculation of other spatial perceptual attributes, such as source distance and source width.

5. REFERENCES

- [1] J. Blauert. *Spatial Hearing: The psychophysics of Human Sound Localisation*. MIT Press, Cambridge, Massachusetts, 1983.

- [2] J. Daniel, R. Nicol, and S. Moreau. Further Investigations of Higher Order Ambisonics and Wavefield Synthesis of Holophonic Sound Imaging. 114th Convention of the Audio Eng. Soc., Preprint 5788, Amsterdam, The Netherlands, March 2003.
- [3] B. Gardner and K. Martin. HRTF measurements of a KEMAR dummy-head microphone. <http://sound.media.mit.edu/KEMAR.html>, May 18, 1994 (last revised July 18, 2000).
- [4] M. A. Gerzon. Surround-sound psychoacoustics, *Wireless World*, 80:483-486, December 1974.
- [5] H. Haas. The influence of a single echo on the audibility of speech. *J. Audio Eng. Soc.*, 20:146-159, 1972.
- [6] G. Kuhn. Model for the interaural time difference in the azimuthal plane. *J. Acoustical Soc. of America*, 62:157-167, July 1977.
- [7] E. M. Macpherson. A computer model of binaural localization for stereo imaging measurement. *J. Audio Eng. Soc.*, 39(9):604-622, September 1991.
- [8] R. Nicol and M. Emerit. 3D Sound Reproduction Over an Extensive Listening Area: A Hybrid Method Derived From Holophony and Ambisonic. In *Proceedings of the AES 16th International Conference: Spatial Sound Reproduction*, Rovaniemi, Finland, April 1999.
- [9] M. Pocock. A Computer Model of Binaural Localization. 72nd Convention of the Audio Eng. Soc., Preprint 1951, Anaheim, California, USA, September 1982.
- [10] V. Pulkki, M. Karjalainen, and J. Huopaniemi. Analysing virtual sound source attributes using a binaural auditory model. *J. Audio Eng. Soc.*, 47(4):203-217, April 1999.
- [11] F. Rumsey. Spatial Quality Evaluation for Reproduced Sound Terminology, Meaning, and a Scene-Based Paradigm. *J. Audio Eng. Soc.*, 50(9):651-666, September 2002.
- [12] B. Supper, T. Brookes, and F. Rumsey. A lateral angle tool for spatial auditory analysis. 116th Convention of the Audio Eng. Soc., Preprint 6068, Berlin, Germany, May 2004.
- [13] E. Verheijen. *Sound Reproduction by Wave Field Synthesis*. PhD thesis, Faculty of Applied Physics, Delft, The Netherlands, 1996.
- [14] H. Wallach, E. B. Newman, and M. R. Rosenzweig. The Precedence Effect in Sound Localisation. *Am. J. Psychol.* 52:315-336, 1949. Reprinted in *J. Audio Eng. Soc.*, 21:817-826, 1973.
- [15] M. Williams and G. Le Dù. The Quick Reference Guide to Multichannel Microphone Arrays, Part 1: using Cardioid Microphones. 110th Convention of the Audio Eng. Soc., Preprint 5336, Amsterdam, The Netherlands, May 2001.
- [16] R. S. Woodworth and G. Schlosberg, *Experimental Psychology*, pp. 349-361. Holt, Rinehard and Winston, New York, 1962.