isvr Institute of Sound and Vibration Research

Robustness of the Performance of the "Stereo Dipole" to Head Misalignment

T.Takeuchi and P.A. Nelson

ISVR Technical Report No. 285

October 1999

University of Southampton

SCIENTIFIC PUBLICATIONS BY THE ISVR

Technical Reports are published to promote timely dissemination of research results by ISVR personnel. This medium permits more detailed presentation than is usually acceptable for scientific journals. Responsibility for both the content and any opinions expressed rests entirely with the author(s).

Technical Memoranda are produced to enable the early or preliminary release of information by ISVR personnel where such release is deemed to the appropriate. Information contained in these memoranda may be incomplete, or form part of a continuing programme; this should be borne in mind when using or quoting from these documents.

Contract Reports are produced to record the results of scientific work carried out for sponsors, under contract. The ISVR treats these reports as confidential to sponsors and does not make them available for general circulation. Individual sponsors may, however, authorize subsequent release of the material.

UNIVERSITY OF SOUTHAMPTON

INSTITUTE OF SOUND AND VIBRATION RESEARCH

FLUID DYNAMICS AND ACOUSTICS GROUP

Robustness of the Performance of the "Stereo Dipole" to Head Misalignment

by

T Takeuchi and P A Nelson

ISVR Technical Report No. 285

October 1999

Authorized for issue by
Professor P A Nelson
Group Chairman

ACKNOWLEDGEMENTS

We would like to thank Dr Ole Kirkeby and Professor Hareo Hamada for helpful discussions. This research was supported by Yamaha Corporation, Alpine Electronics and Hitachi Ltd. T. Takeuchi is supported by Kajima Corporation.

Contents

Abstract

When binaural sound signals are presented with two loudspeakers, the listener's ears are required to be in the relatively small region which is under control of the system. Misalignment of the head results in inaccurate synthesis of the binaural signals. Consequently, directional information associated with the acoustic signals is inaccurately reproduced. When the two loudspeakers are placed close together, the spatial rate of change of the generated sound field is much smaller than that generated by two loudspeakers spaced apart. Therefore, the performance of such a system is expected to be more robust to misalignment of the listener's head. Robustness of performance is investigated here with respect to head displacement in three translational and three rotational directions. A comparison is given between systems consisting of two loudspeakers either placed close together or spaced apart. The extent of effective control with head displacement and the resulting deterioration in directional information is investigated by analysing temporal and spectral localisation cues estimated from synthesised binaural signals. Subjective localisation experiments are performed for cases in which notable differences in performance are expected from the previous analysis. It is shown that the system comprising two loudspeakers that are close together is very robust to misalignment of the listener's head.

1 INTRODUCTION

Binaural technology [1]-[3] is often used to present a virtual acoustic environment to a listener. The principle of this technology is to control the sound pressure at the listener's ears so that the reproduced sound pressure coincides with that which would be produced when he is in the desired real sound field. Producing the correct ear sound pressures should lead to almost the same sensation as the listener would experience in the real sound field for most realistic sound signals. The superiority of this binaural technique lies in its capability of providing very accurate spatial impression to a listener. Appropriate control of directional information of direct and reflected sounds, as well as information regarding reflecting surfaces, distance which the sound has travelled and information from the sound source itself, is essential to create a convincing virtual auditory space.

Unlike other types [4]-[6] of attempt to give virtual directional information to a listener, binaural technology requires the control of sound at each of two ears independently. One way of achieving this is to use a pair of headphones or similar types of transducers. An alternative to this is to use two loudspeakers at different positions in a listening space with the help of signal processing to ensure that appropriate binaural signals are obtained at the listener's ears [7]-[10].

One disadvantage of binaural sound reproduction over loudspeakers is that the listener's ears must be in the relatively small region in space at which the control is effective. Misalignment of the head position and orientation results in the inaccurate synthesis of the binaural signals at the ears. This results from the change in the transfer functions between the transducers and the listener's ears. Consequently, the performance of the system deteriorates, i.e., directional information associated with the sound is smeared as is other information.

It can also be shown that it is possible to achieve independent control of the sound signal at two ears with a monopole transducer and a dipole transducer at the same position [10],[11]. When two closely spaced monopole transducers are used, the sound field produced is a good approximation to that produced by a point monopole and a point dipole transducer up to a given frequency. We refer to such system as a "Stereo Dipole" [12]. The sound field generated by such a system has a distinct character in that its rate of change over space is much smaller than that generated by two monopole transducers spaced apart [13]. As a conse-

quence, it is expected to be more robust to misalignment of the position and orientation of the listener's head [14].

The objective of this study is to investigate the robustness of the performance of such a system when the listener's head is misaligned. Comparison between two different transducer arrangements is made; two transducers placed close together and two transducers spaced apart. The consequence of three translational and three rotational displacements of the head is examined. Much emphasis is put upon the preservation of directional information which depends mostly upon the head related transfer functions (HRTFs). First, the effectiveness of control is investigated by synthesis of a unit impulse at both ears in both the time and frequency domains. Presentation of an incident sound from various directions are then investigated as the very basic components of a virtual sound environment. Prediction of subjective response is attempted by analysing the synthesised HRTFs. As a temporal localisation cue, the interaural time difference (ITD) is investigated here. The monaural spectral shape cue is also investigated as a spectral localisation cue. In addition, binaural spectral cues, i.e., interaural level difference (ILD) used to localise along the interaural direction and interaural difference of spectral shape to localise around the interaural axis, are also considered. Cues related to the dynamics of head movement are outside the scope of this study. Subjective localisation experiments are performed for displacements for which notable differences in performance are expected from the previous analysis.

2 FACTORS IN THE SYNTHESIS OF A VIRTUAL ACOUSTIC ENVIRONMENT

The principle of the system under investigation is illustrated in Fig. 1. The following is described with a frequency domain representation of the acoustic paths (transfer functions) and sound signals. All the spatial information is in the transfer functions between sound source and both of the listener's ears. As the very basic components of a virtual sound environment, generation of a single incident sound wave is taken as an example here. A pair of binaural signals $\mathbf{d}(z)$ corresponding to a single incident sound wave are generated by filtering a sound source signal $S(z)$ through a vector of filters $\mathbf{a}(z)$ which contains a pair of HRTFs for both ears corresponding to the desired direction of the incident sound. Thus

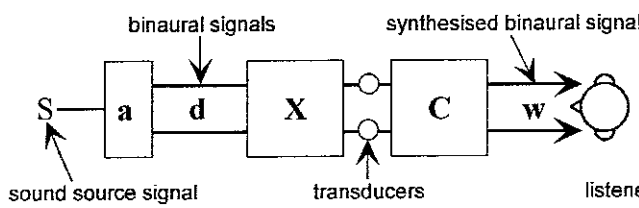

Fig. 1. The principle of binaural synthesis over loudspeakers.

$$\mathbf{d}(z) = \mathbf{a}(z)S(z) \tag{1}$$

When the signals at both ears of the listener are controlled by two transducers in the listening space, the 2x2 matrix $\mathbf{C}(z)$ of transfer functions can be defined between the transducers and the ears. In order to present the binaural signals at each ear, the signals $\mathbf{d}(z)$ are filtered through a 2x2 matrix $\mathbf{X}(z)$ of control filters which contains the pseudo-inverse of the transfer function matrix $\mathbf{C}(z)$. Then the synthesised binaural signals $\mathbf{w}(z)$ and the sound source signal $S(z)$ are related by

$$\mathbf{w}(z) = \mathbf{C}(z)\mathbf{X}(z)\mathbf{a}(z)S(z) \tag{2}$$

For convenience, the control performance matrix $\mathbf{R}(z)$ and vector of synthesised HRTFs $\mathbf{q}(z)$ are defined as follows

$$\mathbf{R}(z) = \mathbf{C}(z)\mathbf{X}(z) \tag{3}$$

$$\mathbf{q}(z) = \mathbf{C}(z)\mathbf{X}(z)\mathbf{a}(z) = \mathbf{R}(z)\mathbf{a}(z) \tag{4}$$

A number of filter design methods have been presented [15],[16]. In short, with the use of a modelling delay Δ and a regularisation parameter for causal stable inversion, $\mathbf{X}(z)$ is designed so that

$$\mathbf{R}(z) = \mathbf{C}(z)\mathbf{X}(z) \approx z^{-\Delta}\mathbf{I} \tag{5}$$

is satisfied where $\mathbf{I}$ is the identity matrix. This ensures the synthesised HRTFs $\mathbf{q}(z)$ are a good approximation to the original HRTFs $\mathbf{a}(z)$. Thus, from Eq. (4) and (5),

$$\mathbf{q}(z) \approx z^{-\Delta}\mathbf{a}(z) \tag{6}$$

As the listener's head is displaced away from the exact position for which control filters $\mathbf{X}(z)$ are calcu-

lated, the transfer functions $\mathbf{C}(z)$ change gradually. Thus the pseudo-identity matrix $\mathbf{R}(z)$ and, as a consequence, the synthesised binaural HRTFs $\mathbf{q}(z)$ are degraded and may result in the wrong subjective perception.

3 ESTIMATING THE SUBJECTIVE RESPONSE

3.1 Model

The physical acoustic paths **a** and **C** are modelled with free field head related impulse responses (HRIRs: the time domain representation of HRTFs). A database comprising directionally discrete HRIRs on a virtual spherical surface 1.4m from a KEMAR dummy head is obtained from MIT Media Lab [17]. Those between sampled directions are obtained by bilinear interpolation on the virtual spherical surface of magnitude and phase spectra in the frequency domain. Those at a different distance from a head are obtained by extrapolation with an appropriately chosen delay and spherical attenuation (Appendix). The loudspeaker response is deconvolved from the data and thus each control transducer of the system is modelled as an ideal monopole source. The control filter matrix **X** is determined by the frequency domain deconvolution method [16].

The listener's head is displaced with respect to six orthogonal axes (three translational and three rotational) as in Table 1 and Fig. 2. Since the robustness to relatively small displacement of the head position and orientation is of interest here, the robustness of the virtual sound image is evaluated relative to the listener's head, not relative to the listening

Description	Terminology
Translation along x-axis	lateral
Translation along y-axis	fore-and-aft
Translation along z-axis	vertical
Rotation about x-axis	pitch
Rotation about y-axis	roll
Rotation about z-axis	yaw

Table 1. Terminology used to describe head displacement.

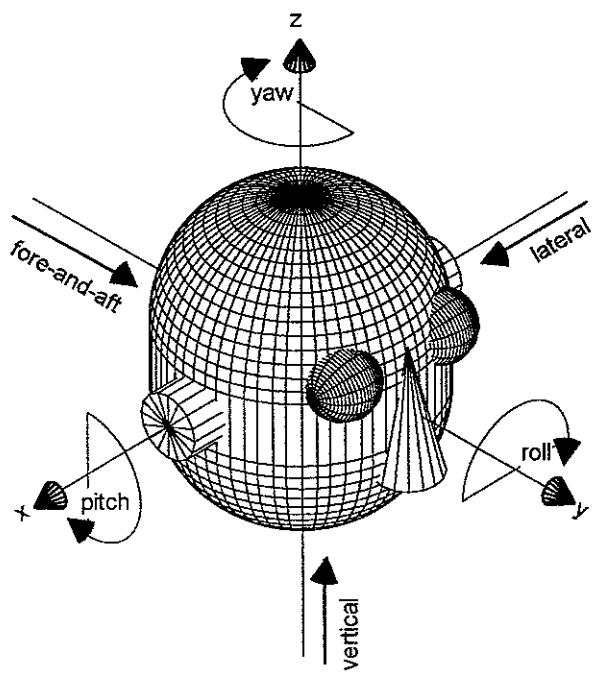

Fig. 2. The Cartesian co-ordinate system used to define head displacement relative to the optimal head position and orientation.

space. In other words, when the listener's head i displaced, he should ideally perceive the same vi tual sound image as in the optimal position and ori entation, unlike those applications where the listene may want to move around in a virtual sound envi ronment.

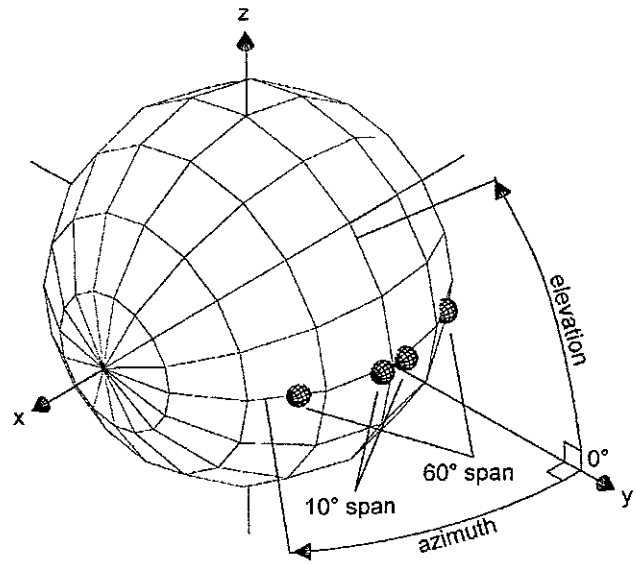

Fig. 3. The spherical co-ordinate system used to define the direction of sound sources relative to the listener's head position and orientation. The two different transducer arrangements investigated are also shown (relative to the optimal head position and orientation).

The spherical co-ordinate system used to defin direction of sound and of transducers is shown i Fig. 3. The origin is at the intersection of th interaural axis and the median plane. The polar axis coincides with the interaural axis. The azimuth angle ranges from -90° to 90° as the direction changes from the pole at the left to the other pole at the right. A cone of constant azimuth is approximately the same as the cone of confusion where there are no ITDs. The elevation angle ranges from -180° on the horizontal plane behind the head to -90° below, 0° on the horizontal plane in front, 90° above the head to 180° again on the horizontal plane behind. Two different transducer arrangements are investigated for comparison. In both cases, two transducers are placed in front of the listener on the horizontal plane (0° elevation) and aligned symmetrically with respect to the median plane. The transducers positioned spanning 60° as seen by the listener (±30° azimuth) are representative of a popular arrangement. The span of 10° (±5° azimuth) represents close spacing.

3.2 Evaluation of localisation cues

In the temporal domain, the interaural cross-correlation function $\Psi_a(t)$ of HRIRs $\mathbf{a}(t)$ corresponding to the real source direction are examined and the time lag which gives the peak values of $\Psi_a(t)$ is used as an estimate of ITD. The interaural cross-correlation function $\Psi_a(t)$ is expressed as follows in terms of the elements of $\mathbf{a}(t)$.

$$\Psi_a(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} a_1(t) a_2(t+\tau) dt \tag{7}$$

There are other possible methods to estimate ITD, for example, by detecting the leading-edge in the HRIRs, or by computing the phase spectrum or group delay of the binaural signals. However, the leading-edge method may misjudge ITD by detecting the less potent onset ITD rather than the ongoing ITD to which neurones are sensitive.[18]-[20] There is no indication that the nervous system could detect the high-frequency phase spectrum nor group delay. Anatomical and physiological studies strongly suggest that ITD information is extracted with the interaural cross-correlation of the auditory-nerve responses to the stimuli in the superior olivary complex then further processed at a higher level of auditory pathway [21],[22]. The envelope delay of high frequency signals as an ITD cue [23],[24] can be extracted by the cross-correlation method as well as the phase delay of low frequency signals. Here we are interpreting the results from a large number of psychophysiological studies that stimuli with ITD related to phase delay or envelope delay produce the response in the cross-correlation mechanism in the auditory nervous system that leads to the perception of directional information.

In the spectral domain, an analysis is performed over a logarithmic scale both in frequency and magnitude to account for the basic property of auditory filters. The monaural spectral shape is regarded as an important cue to identify one direction out of directions with no interaural differences. This cue utilises the change of the spectral shape of the sound source signal due to the HRTF for each ear. The monaural spectral cues also have supplemental role in localisation along the interaural direction [25]. Interaural difference of spectra could have two roles. The major role is to localise along the interaural direction (azimuth discrimination) with interaural level difference (ILD). It could also be another cue to resolve confusion among directions with no interaural time difference (elevation discrimination) by utilising the pattern of frequency dependent interaural spectral difference [26]. The advantage of this cue over the monaural spectral shape cue in practice would be that it does not depend on the spectrum of sound source signal.

3.3 Robustness of temporal cues

First, the effectiveness of control as a function of head displacement is evaluated by analysing the matrix of electro-acoustic paths $\mathbf{R}(t)$ which is independent of the direction of the virtual source. Following this, the

synthesised HRIRs $\mathbf{q}(t)$ with head displacement are analysed in order to demonstrate what happens to temporal cues as a function of the relative direction of the virtual sound source.

3.3.1 Control performance (Temporal)

When the inputs to $\mathbf{R}(t)$ is a pair of simultaneous delta functions $\mathbf{d}(t)$ rather than binaural signals, the interaural cross-correlation function, $\Psi_p(t)$, of the synthesised signals is expressed as

$$\Psi_p(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} p_1(t) p_2(t+\tau) dt \tag{8}$$

where

$$\mathbf{p}(t) = \begin{bmatrix} p_1(t) \\ p_2(t) \end{bmatrix} = \begin{bmatrix} R_{11}(t) + R_{12}(t) \\ R_{21}(t) + R_{22}(t) \end{bmatrix} \tag{9}$$

When the listener's head is at the optimal position and orientation, the synthesised signals $\mathbf{p}(t)$ are approximately delta functions with an identical delay. Thus $\Psi_p(t)$ is a delta function with ITD = 0 (ms). In this way, the directional dependence in $\mathbf{a}(t)$ can be excluded from the analysis of the interaural cross-correlation functions. As the head is displaced away from the optimal position and orientation, the synthesised signals $\mathbf{p}(t)$ are no longer delta functions. Thus $\Psi_p(t)$ is also no longer a delta function. A degraded $\Psi_p(t)$ indirectly suggests the degradation of the ITD cue of the synthesised HRIRs for all directions. A shift of the peak in $\Psi_p(t)$ suggests a shift in the ITD of the synthesised HRIRs and multiple peaks in $\Psi_p(t)$ may cause ambiguity or result in the wrong perception among multiple directions of sound.

Fig. 4 shows the degradation of $\Psi_p(t)$ (the interaural cross-correlation functions for the synthesised simultaneous unit impulses) with lateral displacement over the range of ±250mm. The maximum value of $\Psi_p(t)$ '1' at 0 lag can be observed at 0mm displacement (the optimal position) for both transducer arrangements. When the listener's head is displaced laterally, an ITD shift for the 60° transducer arrangement increases significantly as displacement increases, which is at the rate of approximately 2.5ms/mm.

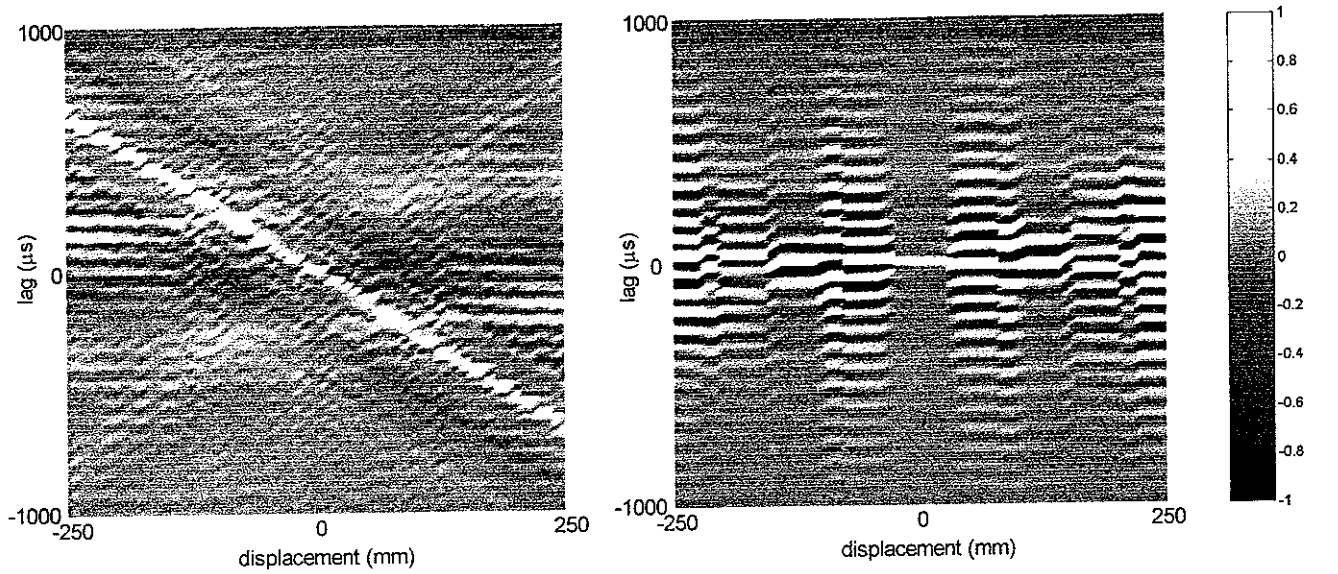

Fig. 4. The effect of lateral displacement on the interaural cross-correlation functions for the synthesis of simultaneous unit impulses. Left panel: 60° transducer span. Right panel: 10° transducer span.

For example, 25mm displacement results in about 65ms ITD shift which corresponds to about 8° shift in azimuth direction. The threshold for ITD discrimination is considered to be approximately 10ms [27] and corresponds to about 4mm displacement with the 60° arrangement. On the other hand, the rate of shift is much less for the 10° transducer arrangement (0.1ms/mm) and so 100mm displacement would just enough produce the threshold value for ITD discrimination. When the listener's head is rolled, the ITD shift is again greater for the 60° arrangement though the difference between two arrangements is much smaller (about 1.6ms/° and 1ms/°) than the lateral displacement (Fig. 5). $\Psi_p(t)$ for fore-and-aft displacement (Fig. 6) shows no shift of the original peak for both transducer arrangements, as expected from the symmetry, but slightly

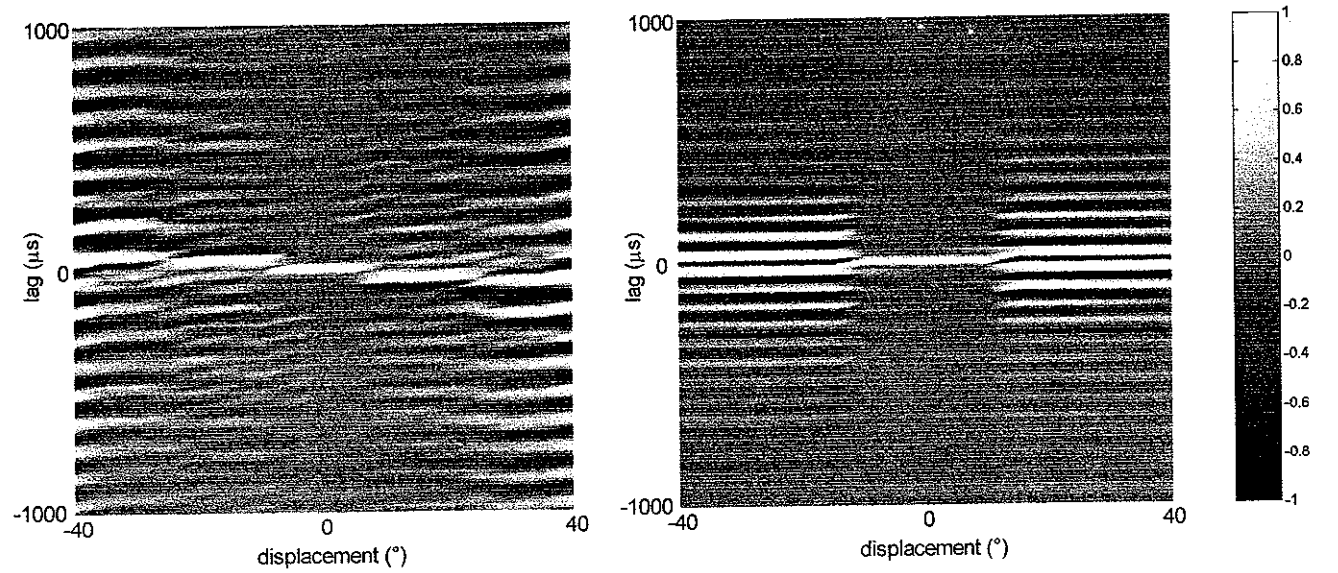

Fig. 5. The effect of roll displacement on the interaural cross-correlation functions for the synthesis of simultaneous unit impulses. Left panel: 60° transducer span. Right panel: 10° transducer span.

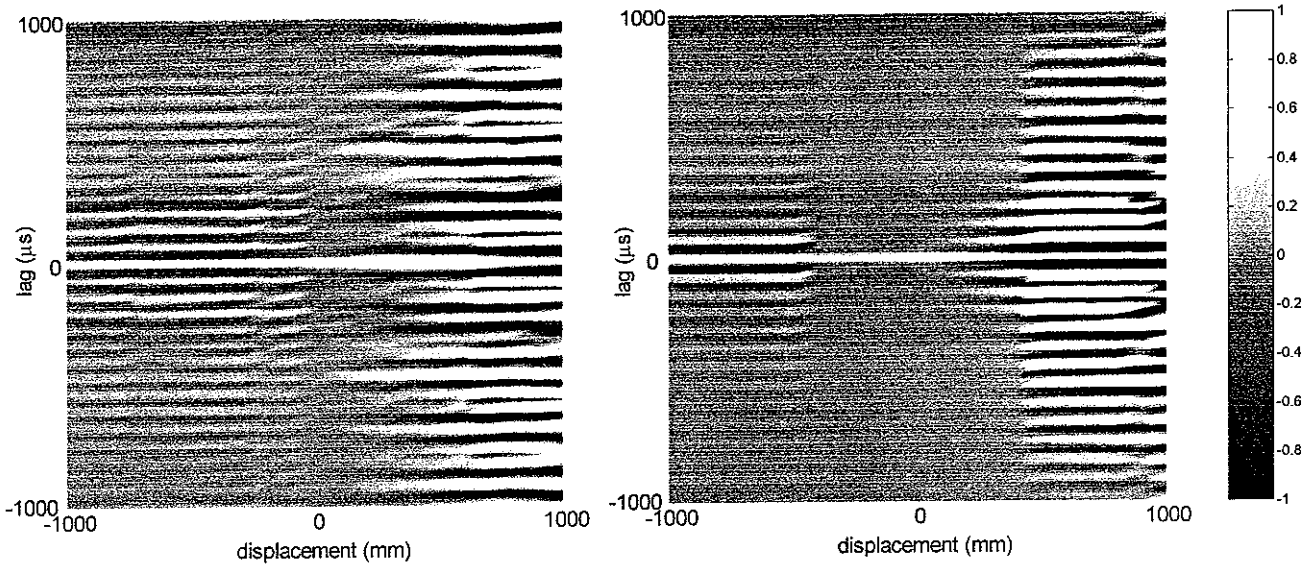

Fig. 6. The effect of fore-and-aft displacement on the interaural cross-correlation functions for the synthesis of simultaneous unit impulses. Left panel: 60° transducer span. Right panel: 10° transducer span.

better preservation (smaller amplitude of additional maxima) of the interaural cross correlation function can be observed for the 10° arrangement. Yaw displacement showed the same ITD shift (about 8ms/°) which corresponds exactly to the yaw displacement angle for both of the two transducer arrangements (Fig. 7). Vertical (Fig. 8) and pitch (Fig. 9) displacement did not show any ITD shift for both arrangements. The results for the six types of displacement are summarised in Table 2.

Comparisons can be made between the six types of displacement by normalising the results by the amount of displacement of the ears produced by each of the six types of head displacement. The ITD cue is

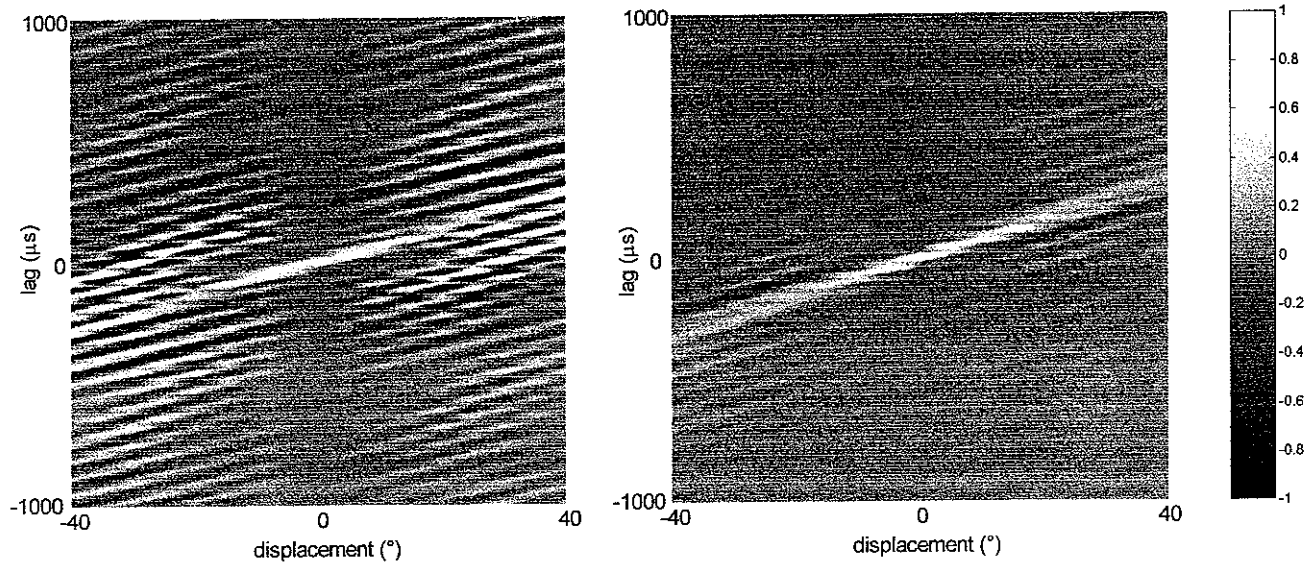

Fig. 7. The effect of yaw displacement on the interaural cross-correlation functions for the synthesis of simultaneous unit impulses. Left panel: 60° transducer span. Right panel: 10° transducer span.

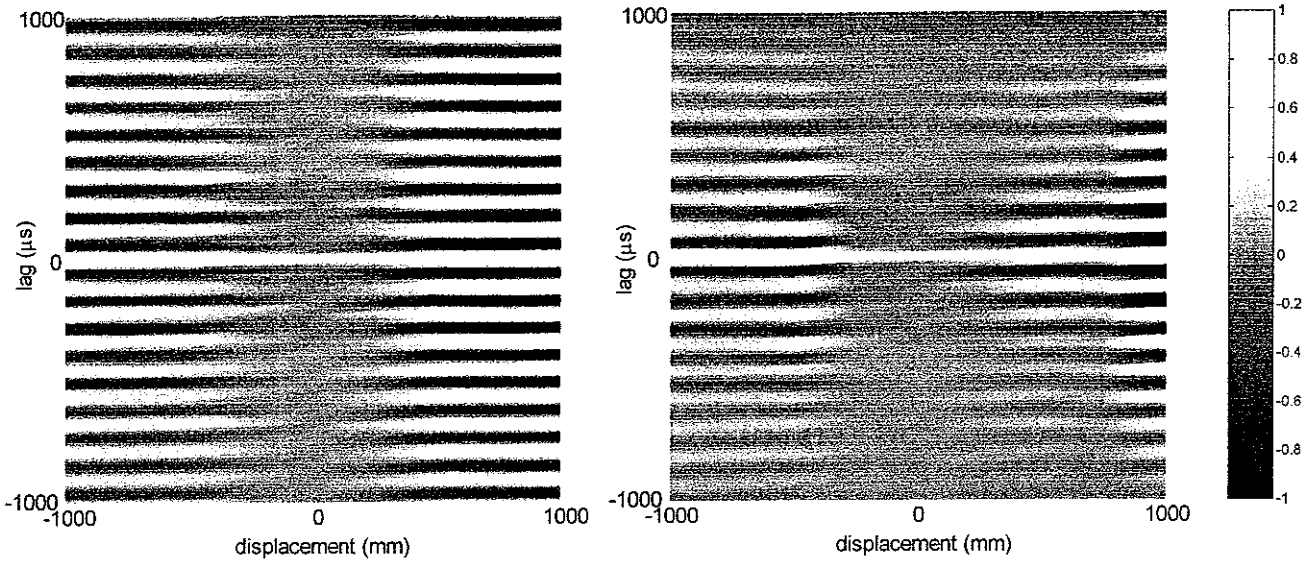

Fig. 8. The effect of vertical displacement on the interaural cross-correlation functions for the synthesis of simultaneous unit impulses. Left panel: 60° transducer span. Right panel: 10° transducer span.

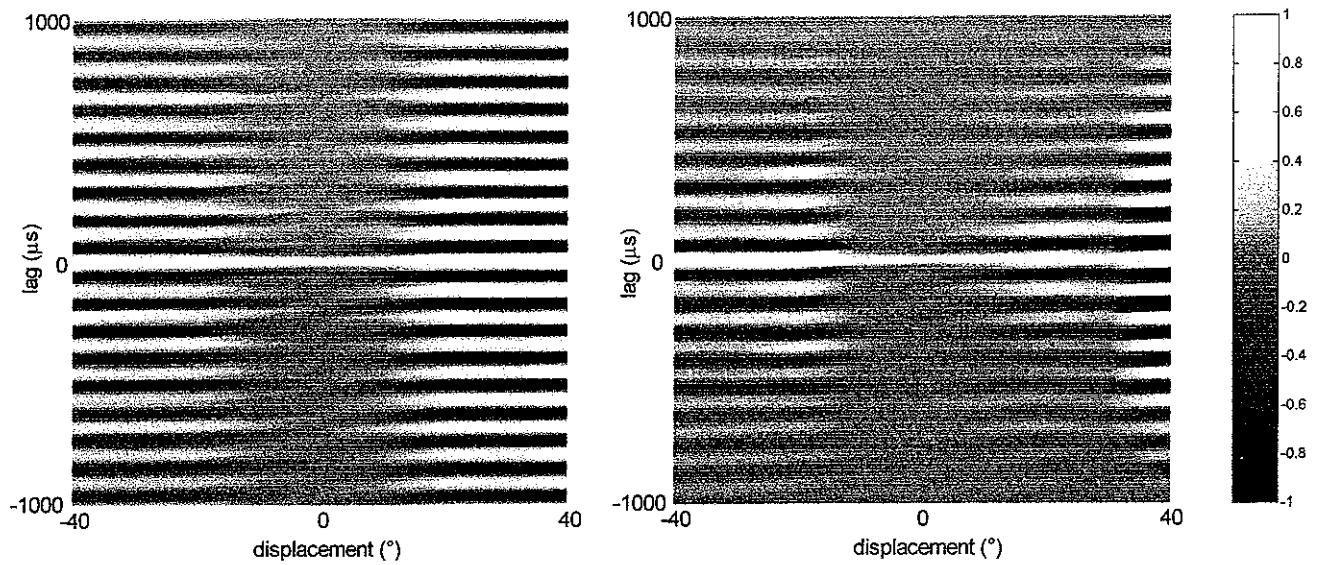

Fig. 9. The effect of pitch displacement on the interaural cross-correlation functions for the synthesis of simultaneous unit impulses. Left panel: 60° transducer span. Right panel: 10° transducer span.

the most sensitive to yaw displacement followed by lateral and roll displacements. It is very robust to fore-and-aft, pitch and vertical displacement. However, the difference in the robustness of the ITD cue between two different transducer arrangements is most significant for lateral displacement followed by roll and fore-and-aft displacements. There are no obvious differences between two transducer arrangements for the other three displacements (yaw, vertical, pitch).

3.3.2 Accuracy of synthesis (Temporal)

By analogy with $\Psi_a(t)$, the interaural cross-correlation functions of synthesised HRIRs, $\Psi_q(t)$, is expressed as follows in terms of the elements of $\mathbf{q}(t)$.

$$\Psi_q(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} q_1(t) q_2(t+\tau) dt \qquad (10)$$

As the ITD cue is regarded as the most salient cue that is used to determine the azimuth direction [28], directions on the horizontal plane which contain two sets of all the azimuth directions are taken as examples to show the interaural cross-correlation functions of HRIRs (Fig. 10). That of the original HRIRs, $\Psi_a(t)$, is shown in Fig. 10a and that of synthesised HRIRs, $\Psi_q(t)$, when the listener's head is displaced 25 mm laterally are shown in Fig. 10b and Fig. 10c. In Fig. 10a, it can be observed that ITD is increasing almost linearly with respect to azimuth angle over most of the range. (Note that the variation is not sinusoidal which would be the case if there were no head in the sound field.) $\Psi_q(t)$ is severely degraded with the 60° transducer arrangement (Fig. 10b); a few large additional local maxima (especially around ±250ms, corresponding to ±30° azimuth which are the control trans-

type of displacement	rate of ITD shift		displacement at 10 ms ITD shift	
	60° span	10° span	60° span	10° span
lateral	2.5 ms/mm	0.1 ms/mm	4 mm	100 mm
fore-and-aft	0 ms/mm	0 ms/mm	-	-
vertical	0 ms/mm	0 ms/mm	-	-
pitch	0 ms/°	0 ms/°	-	-
roll	1.6 ms/°	1.0 ms/°	6°	10°
yaw	8 ms/°	8 ms/°	1.3°	1.3°

Table 2. Estimated rate of ITD shift and displacement which gives the threshold value of ITD discrimination (10ms) for six types of displacement and two different transducer arrangements.

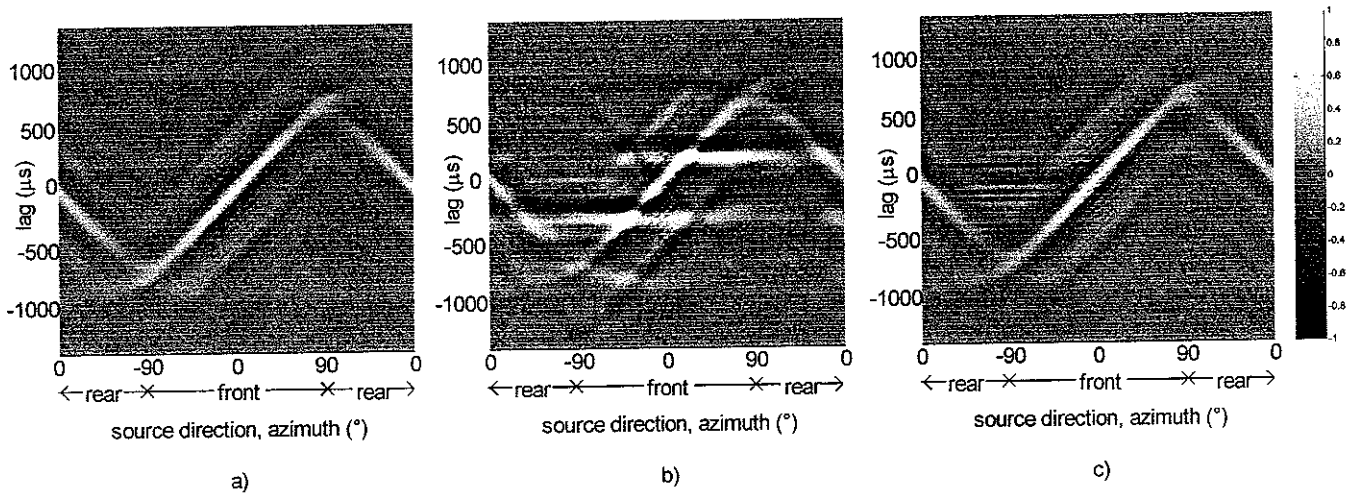

Fig. 10. Interaural cross-correlation functions of the original and synthesised HRIRs corresponding to source directions on the horizontal plane. a) Calculated from the original HRIRs. b) Calculated from the synthesised HRIRs with 60° transducer span when the listener's head is displaced 25 mm laterally. c) Calculated from the synthesised HRIRs with 10° transducer span when the listener's head is displaced 25 mm laterally.

ducer directions) can be observed over wide range of virtual source directions as well as a shift (about 65ms, 8° azimuth) of the original peak. However, $\Psi_q(t)$ is better preserved with the 10° arrangement (Fig. 10c) except for minor local maxima (the largest around -60ms which again corresponds to the control transducer directions) at virtual source directions around -90° azimuth.

ITD estimated from the synthesised HRIRs without displacement for both transducer arrangements are identical to the estimate from the original HRIRs for all the directions around the head. The estimated ITD from the synthesised HRIRs when the listener's head is displaced 25 mm laterally for most of the directions around the head is plotted in Fig. 11. There are no data points on bottom part of the spherical plot. In general, it is observed that cones of constant ITD are shifted from the original value (Fig. 11a) for the 60° arrangement (Fig. 11b) but little shift is observed for the 10° arrangement (Fig. 11c), as observed in Fig. 4. The system with the 10° transducer arrangement preserved the synthesised ITD value for larger azimuth directions better than that of the 60° arrangement. A slightly worse performance is expected on the left side of the head than the other side (right) for the 10° arrangement. Whereas the right side shows worse performance than the left side for the 60° arrangement. The loss of a large ITD value around large azimuth directions (e.g. $|\text{azimuth}| > \pm 30°$ in Fig. 11b, around -90° azimuth in Fig. 11c) is primarily because the additional peaks in the interaural cross-correlation function became larger than the original peak. When the head is displaced, large additional peaks which give ITD values corresponding to the direction of the control transducers appear. In cases when these additional peaks are larger than the original peaks, if the largest peak is taken to estimate ITD, the virtual sound source would vanish and the listener would localise the sound source in the direction of the control transducers. However, with existence of the other types of cue such as monaural spectral shape cues, the smaller magnitude of the original peak could be more plausible in estimating ITD. If it is taken to estimate ITD, it would result in much better preserved ITD value and thus better preserve the direction of virtual sound. This is down to psychological function at higher levels of the nervous system. It is likely, inferring from the results from subjective experiment presented in a later section, that a smaller but more plausible original peak would result in the estimated ITD for head displacements below a certain value.

3.4 Robustness of spectral cues

As in the analysis of temporal cues, the effectiveness of control as a function of head displacement is

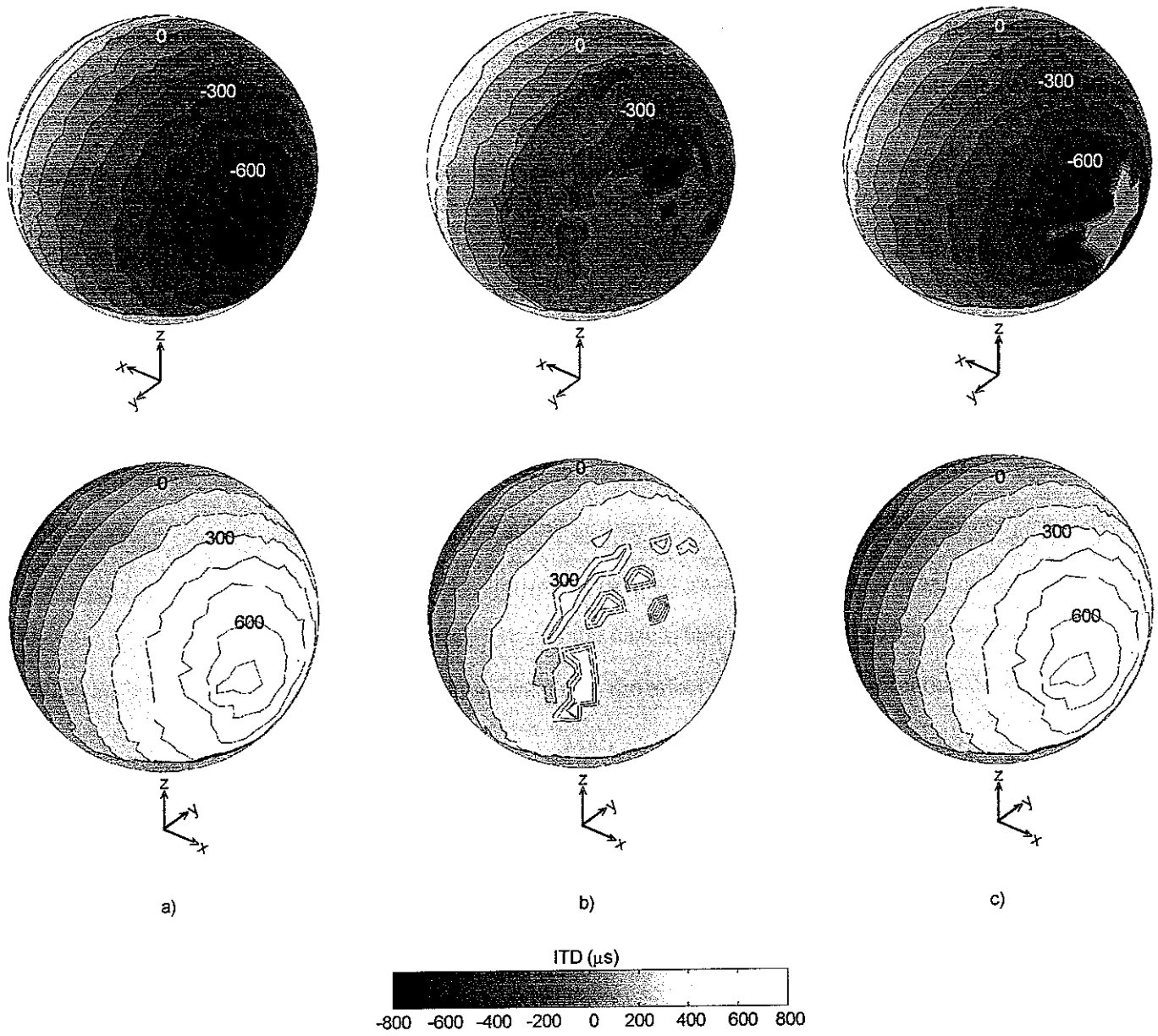

Fig. 11. Estimated ITD, plotted as a function of the intended direction of the virtual sound source. a) Left column: estimated from the original HRIRs. b) Middle column: estimated from synthesised HRIRs with 25mm lateral displacement for the 60° transducer span. c) Right column: estimated from synthesised HRIRs with 25mm lateral displacement for the 10° transducer span. Upper row: view from the upper-front-left (azimuth=-45°, elevation=30°). Lower row: view from the upper-rear-right (azimuth=45°, elevation=150°).

evaluated first by analysing the matrix of transfer functions $\mathbf{R}(z)$ which is independent of the virtual source direction. Then, synthesised HRTFs $\mathbf{q}(z)$ are analysed in order to demonstrate what happens to spectral cues depending on the direction of the virtual sound source.

3.4.1 Control performance (Spectral)

When the control system is required to synthesise particular spectra at two ears, head displacement results in leakage of some of the signal intended for one of the ears to the other ear. This is the so called

"cross-talk" component of the signals, i.e. the component of the signal for right ear fed to the left ear and vice versa. This can be regarded as noise component in the intended signal. The components of the synthesised HRTFs $\mathbf{q}(z)$ are given by

$$\mathbf{q}(z) = \begin{bmatrix} Q_1(z) \\ Q_2(z) \end{bmatrix} = \begin{bmatrix} R_{11}(z)\,A_1(z) + R_{12}(z)A_2(z) \\ R_{21}(z)\,A_1(z) + R_{22}(z)\,A_2(z) \end{bmatrix} \tag{11}$$

where $R_{11}(z)$ and $R_{22}(z)$ are the elements which contribute towards the correct synthesis of the HRTFs but $R_{12}(z)$ and $R_{21}(z)$ are noise elements which smear the synthesis. For the left ear, the signal (signal intended for the left ear) to noise (signal intended for the right ear) ratio of the control system is estimated from $|R_{11}(z)|/|R_{12}(z)|$. This is the case when the time histories of the inputs to $\mathbf{R}(z)$ are a pair of identical delta functions. This again excludes the effect of $\mathbf{a}(z)$, i.e. the direction dependence. Fig. 12 shows the degradation of the S/N for the HRTF synthesis at the left ear with lateral displacement over the range of ±250mm. The signal to noise ratio (S/N) at the right ear, $|R_{22}(z)|/|R_{21}(z)|$, can be obtained by flipping over the left and right of the figure. Much larger displacements which maintain good S/N over wide frequency range (>500Hz) are allowed for the 10° transducer arrangement (roughly ±40mm for 10dB S/N) compared to the 60° transducer arrangement (roughly ±8mm for 10dB S/N). The dip in S/N around 9kHz and 13kHz even when the head is at the optimal position is due to low signal to noise ratio of the measurement of the HRTFs. Good S/

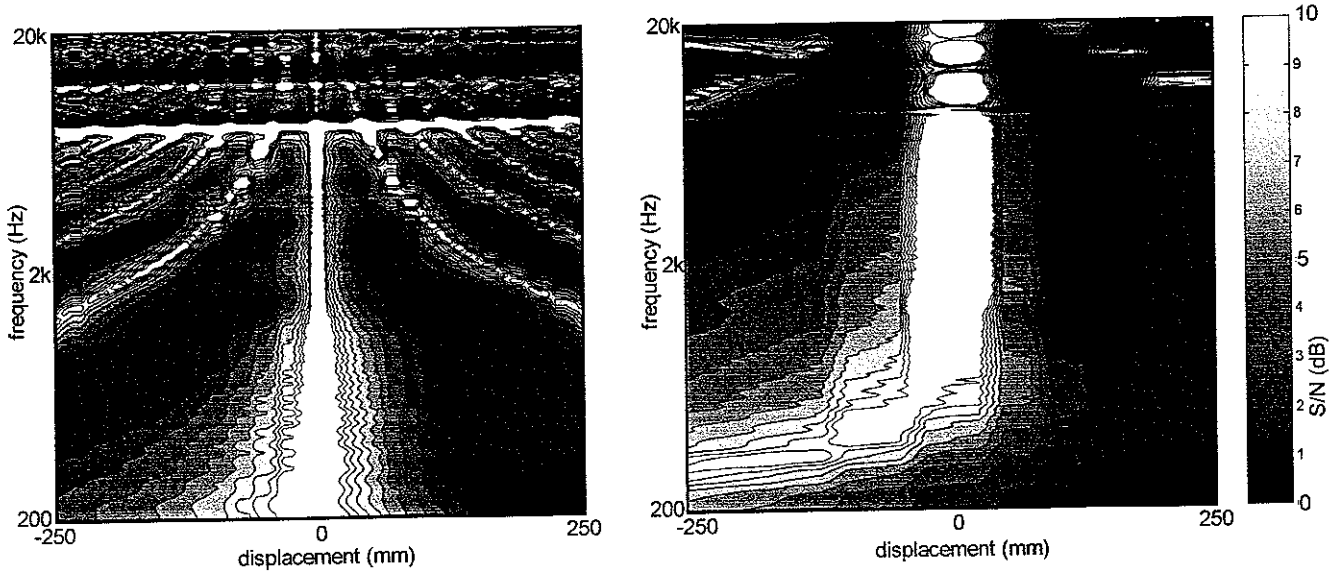

Fig. 12. Signal to noise ratio for the HRTF synthesis at the left ear as a function of lateral displacement. Left panel: 60° transducer span. Right panel: 10° transducer span.

N with larger displacement for the 10° arrangement can also be observed for fore-and-aft (roughly ±410mm compared to ±120mm for 10dB S/N) and yaw (roughly ±12° compared to ±6° for 10dB S/N) displacement as shown in Fig. 13 and Fig. 14. The 60° transducer arrangement has the advantage at frequencies below 500Hz, however. This is where ILD cues are less potent than ITD cues. A slightly better S/N is preserved with the 60° arrangement for pitch (Fig. 15) and vertical (Fig. 16) displacement. With rotation about the interaural axis, transducers being at large azimuth angle means less change of transducer direction than transducers being around the median plane. There are not large differences between two arrangements for roll displacements (Fig. 17). The results for six types of displacements are summarised in Table 3.

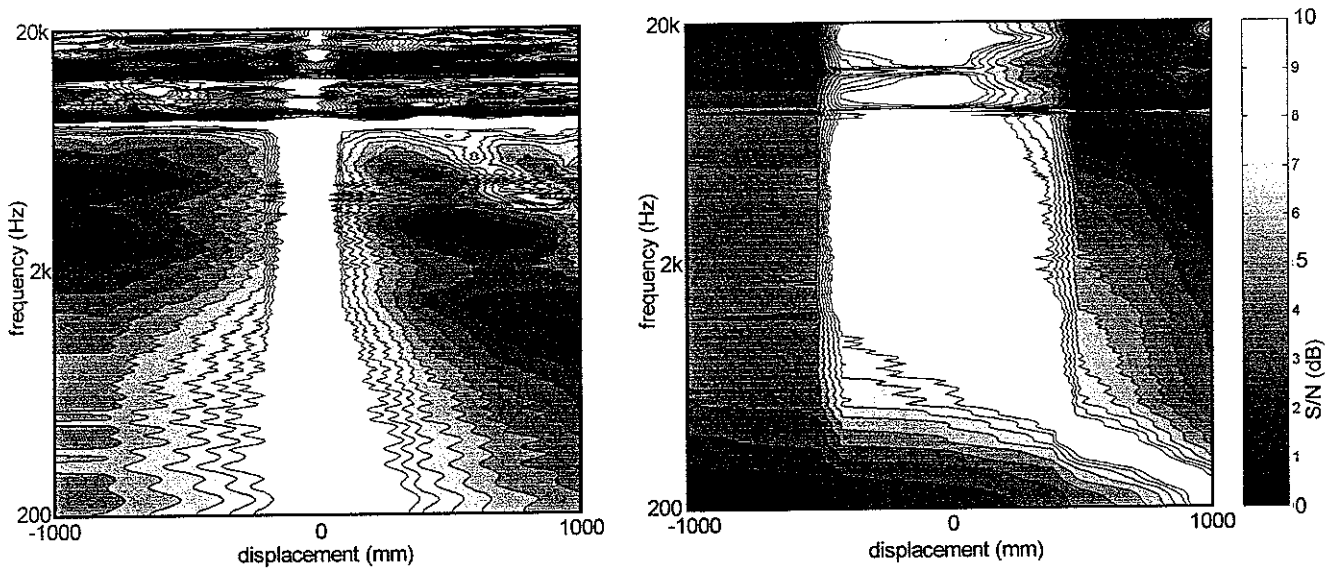

Fig. 13. Signal to noise ratio for the HRTF synthesis at the left ear as a function of fore-and-aft displacement. Left panel: 60° transducer span. Right panel: 10° transducer span.

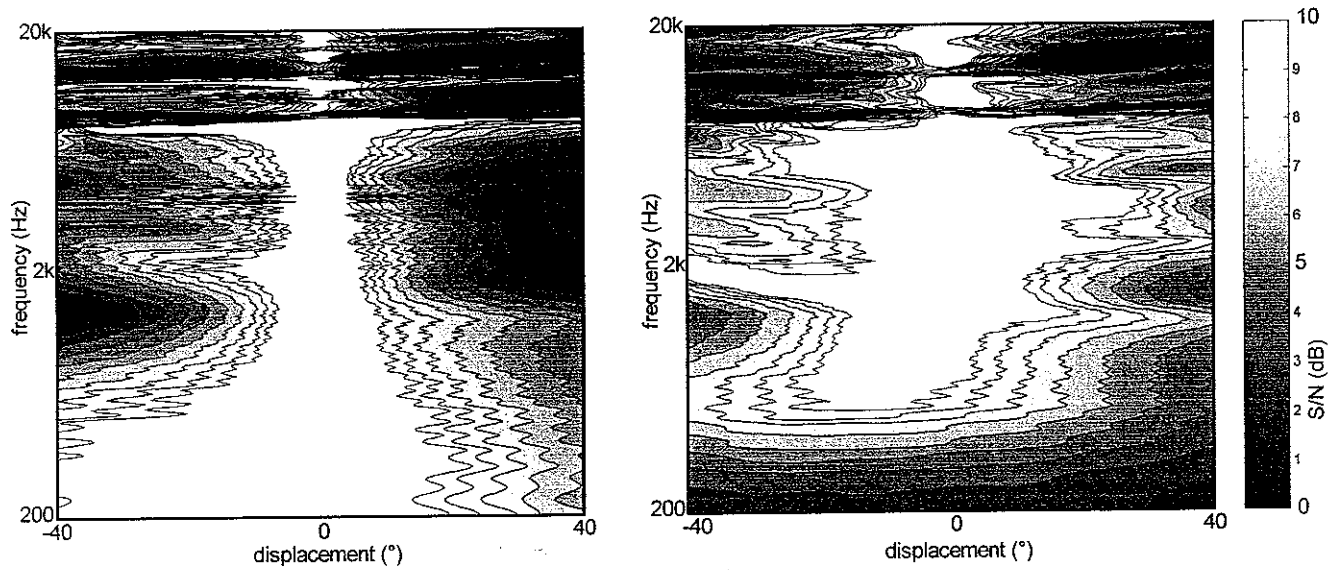

Fig. 14. Signal to noise ratio for the HRTF synthesis at the left ear as a function of yaw displacement. Left panel: 60° transducer span. Right panel: 10° transducer span.

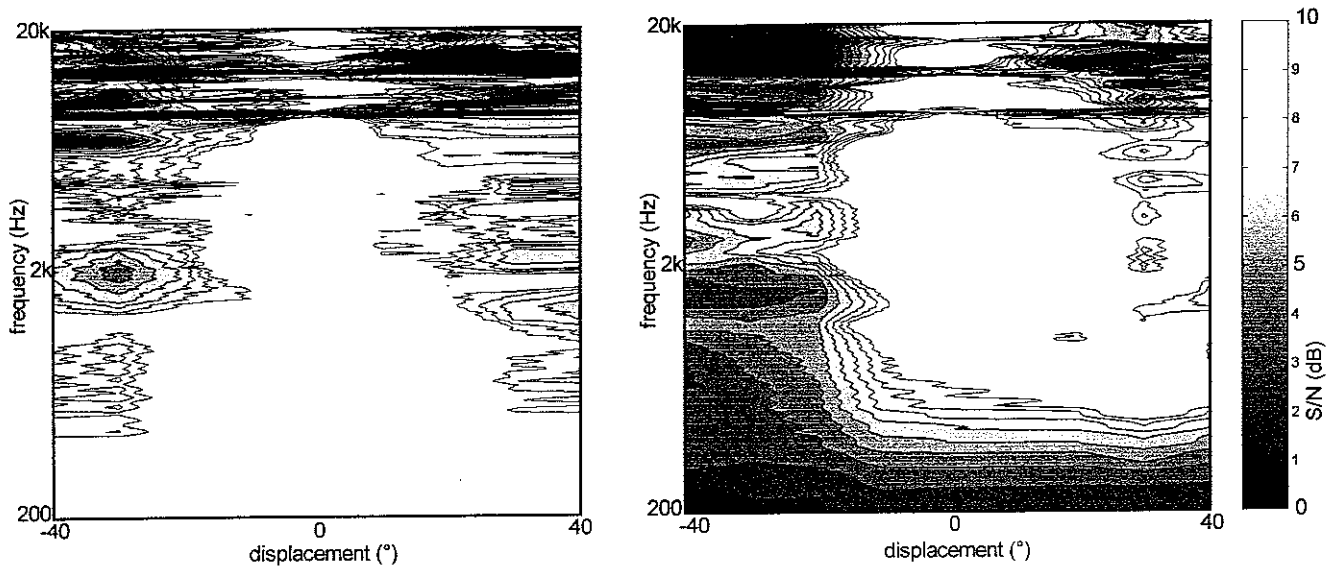

Fig. 15. Signal to noise ratio for the HRTF synthesis at the left ear as a function of pitch displacement. Left panel: 60° transducer span. Right panel: 10° transducer span.

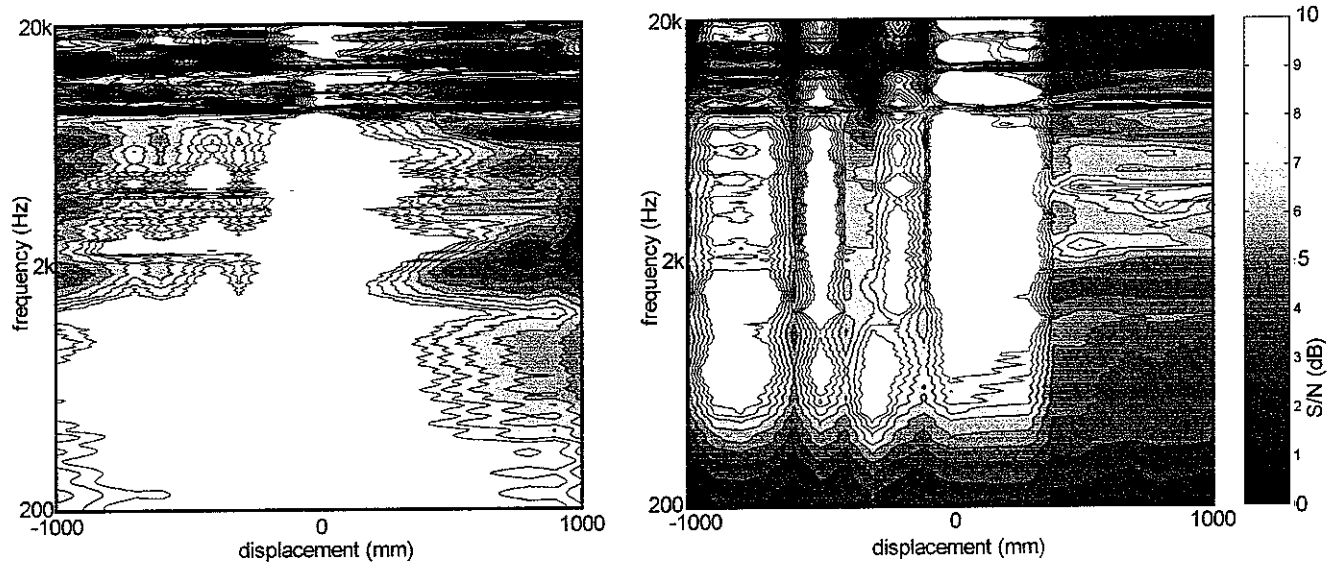

Fig. 16. Signal to noise ratio for the HRTF synthesis at the left ear as a function of vertical displacement. Left panel: 60° transducer span. Right panel: 10° transducer span.

When compared in the same way as used in the temporal cue analysis, spectral cues are most sensitive to lateral and roll displacement followed by yaw, pitch, vertical and fore-and-aft displacements. However, the difference in robustness of spectral cues between two different transducer arrangements is most significant for lateral displacement followed by fore-and-aft and yaw displacements. Note that 10dB S/N is roughly sufficient to synthesise the monaural spectra for the ipsi-lateral ear but much better S/N is required for the contra-lateral ear. This is because, if the level of two desired ear signals $\mathbf{d}(z)$ is compared, the level of the signal for the ipsi-lateral ear is smaller than that for the other ear over most of the frequency range and for most directions. As a result, at the contra-lateral ear, binaural synthesis is affected by a smaller signal input

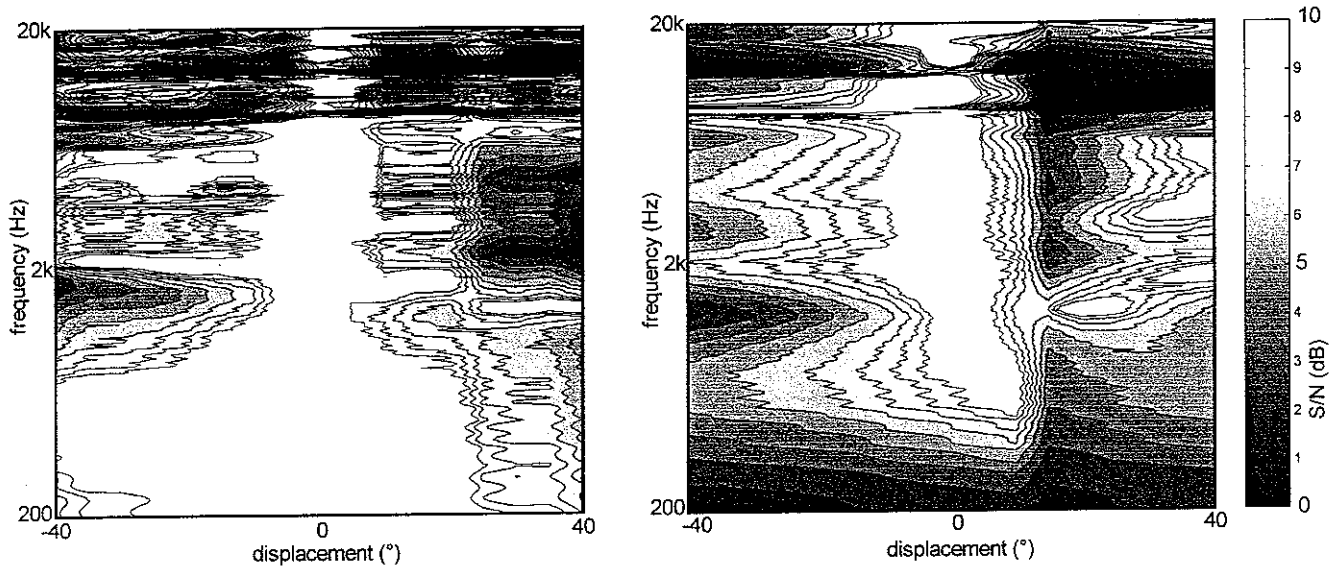

Fig. 17. Signal to noise ratio for the HRTF synthesis at the left ear as a function of roll displacement. Left panel: 60° transducer span. Right panel: 10° transducer span.

with a much larger noise input in addition to the response of the control performance of the system.

type of displacement	displacement at 10dB S/N	
	60° span	10° span
lateral	±8 mm	±40 mm
fore-and-aft	±120 mm	±410 mm
vertical	±220 mm	±190 mm
pitch	±18°	±14°
roll	±9°	±9°
yaw	±6°	±12°

Table 3. Estimated displacement which gives 10dB signal to noise ratio of the control system for six types of displacement and two different transducer arrangements.

3.4.2 Accuracy of synthesis (Spectral)

As the role of the monaural spectral shape cue is primarily to determine the elevation direction of sources located on the cone of confusion, directions along the cone of constant azimuth (50°) are taken as examples to illustrate the monaural spectral shape cue in the HRTFs. Fig. 18 shows examples of monaural spectral shape in HRTFs for the ipsi-lateral (right) ear at directions along the cone of constant azimuth (50°). Significant differences in spectrum pattern between sources below (at negative elevation) and above (positive elevation) the horizontal plane can be observed easily in Fig. 18a for real sound sources (estimated from $|A_2(z)|$). There are less significant differences between sources in front (0~±90°) and in the rear (±90°~±180°) except on the horizontal plane where a significant dip in spectra around ±180° compared to those around ±0° can be seen in the mid frequency range. The synthesised monaural spectral shape (estimated from $|Q_2(z)|$) when the listener's head is displaced 40 mm laterally are shown in Fig. 18b and Fig. 18c. The elevation dependency is less clear for that of the 60° arrangement (Fig. 18b). However, the synthesised monaural spectral shape for the 10° transducer arrangement (Fig. 18c) shows similar elevation dependent monaural spectra to the original spec-

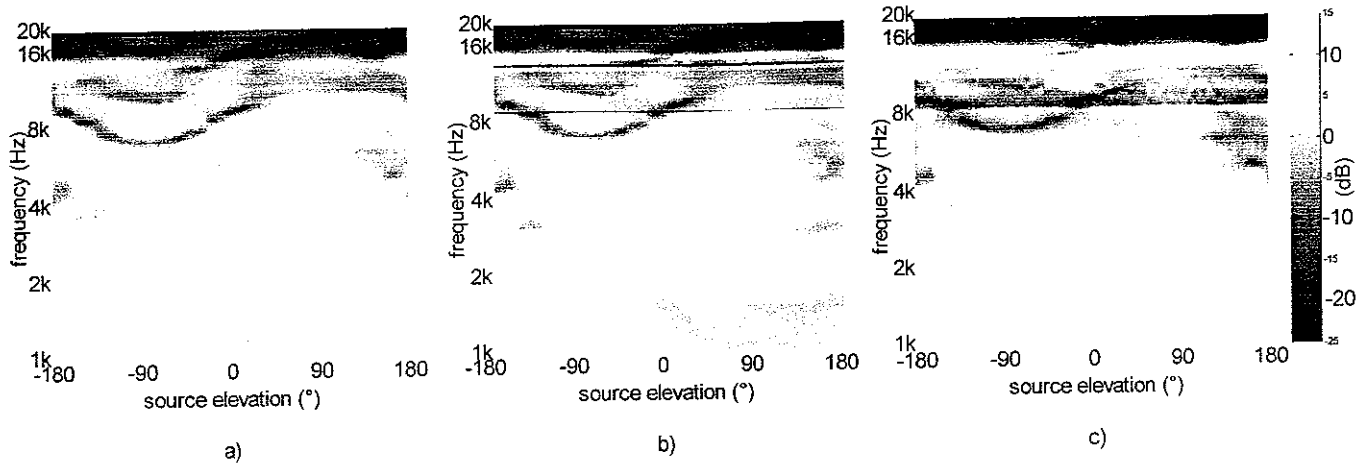

Fig. 18. Monaural spectral shape in HRTFs for the ipsi-lateral (right) ear. Sound source directions are along the cone of constant azimuth (50°). a) Left panel: monaural spectral shape by real sound sources. b) Middle panel: monaural spectral shape synthesised by the 60° transducer arrangement. The listener's head is displaced 40mm laterally. c) Right panel: monaural spectral shape synthesised by the 10° transducer arrangement. The listener's head is displaced 40mm laterally.

tra (Fig. 18a). The consequence of degraded monaural spectral shape would be an increased number of confusions among the directions on the constant azimuth cone. The degradation of this cue may also affect the azimuth localisation since the monaural spectral cue has a supplemental role for azimuth discrimination, especially when the interaural cross-correlation function $\Psi_q(t)$ is degraded to present ambiguity in estimating ITD due to a multiple choice of peaks.

Fig. 19 shows examples of monaural spectral shape in HRTFs for the contra-lateral (left) ear (estimated from $|A_1(z)|$ and $|Q_1(z)|$) at directions along the cone of constant azimuth (50°). As for the ipsi-lateral ear, differences in spectrum pattern for real sound sources is more significant between sources below and above

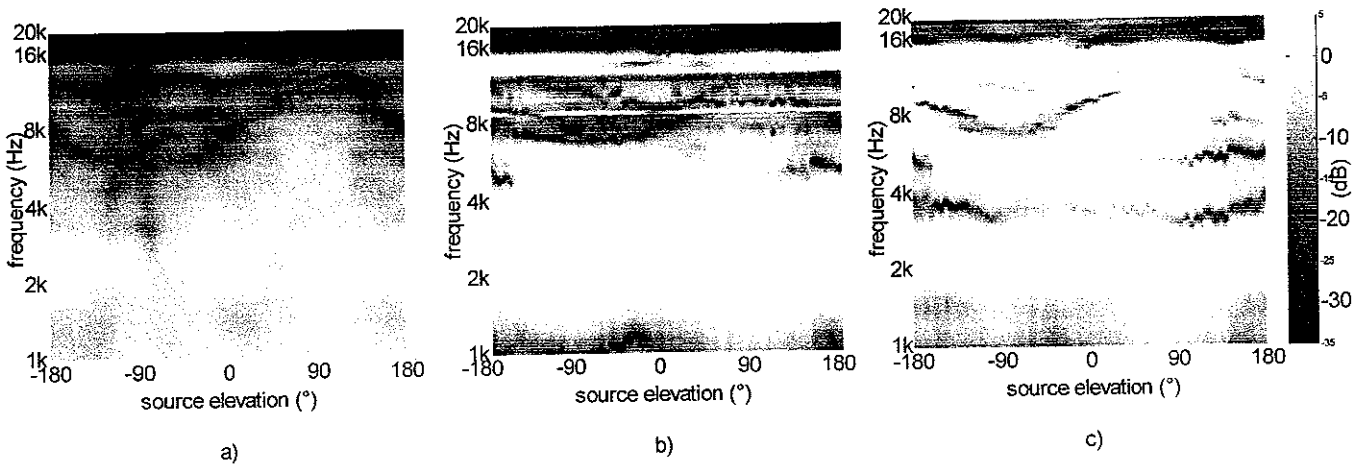

Fig. 19. Monaural spectral shape in HRTFs for the contra-lateral (left) ear. Sound source directions are along the cone of constant azimuth (50°). a) Left panel: monaural spectral shape by real sound sources. b) Middle panel: monaural spectral shape synthesised by the 60° transducer arrangement. The listener's head is displaced 40mm laterally. c) Right panel: monaural spectral shape synthesised by the 10° transducer arrangement. The listener's head is displaced 40mm laterally.

than between front and rear (Fig. 19a). When the listener's head is displaced 40 mm laterally, the monaural spectral shape cue for the synthesised contra-lateral (left ear) HRTF is dominated by the noise, i.e., the cross-talk component, even for the 10° transducer arrangement due to the low S/N (Fig. 19b and Fig. 19c). The requirement for the preservation of monaural spectra for the contra-lateral ear is much more severe than that of the ipsi-lateral ear as pointed out in the previous section. For example, a lateral displacement of not more than 25 mm even for the 10° transducer arrangement and less than 5 mm for the 60° arrangement is required for the 50° azimuth directions. Obviously, the requirement varies as the direction of virtual sound source varies. The variation of azimuth direction (along the interaural axis) has more influence on it than the variation of the elevation direction (around the interaural axis).

Naturally, the same requirement as the contra-lateral monaural spectral shape cue, which is more severe than ipsi-lateral ear, applies for the both of the binaural spectral cues. In terms of analysis, these binaural spectral cues are essentially identical to the difference between the two monaural spectral shapes and estimated from $|Q_2(z)|/|Q_1(z)|$. Examples of the interaural spectral shape difference at directions along the cone of constant azimuth (50°) is shown in Fig. 20. As a matter of course, it has features of monaural spectral shape for both ears (Fig. 20a). For preservation of this cue (as well as monaural spectral shape for the contra-lateral ear), Fig. 20b and Fig. 20c also shows that 25mm lateral displacement is just within the limit with the 10° transducer arrangement for the cone of 50° azimuth directions but not with the 60° transducer arrangement. Above all, these monaural and binaural spectral shape cues are well preserved by the 10° transducer arrangement, so less confusion along the cone of confusion is expected with this arrange-

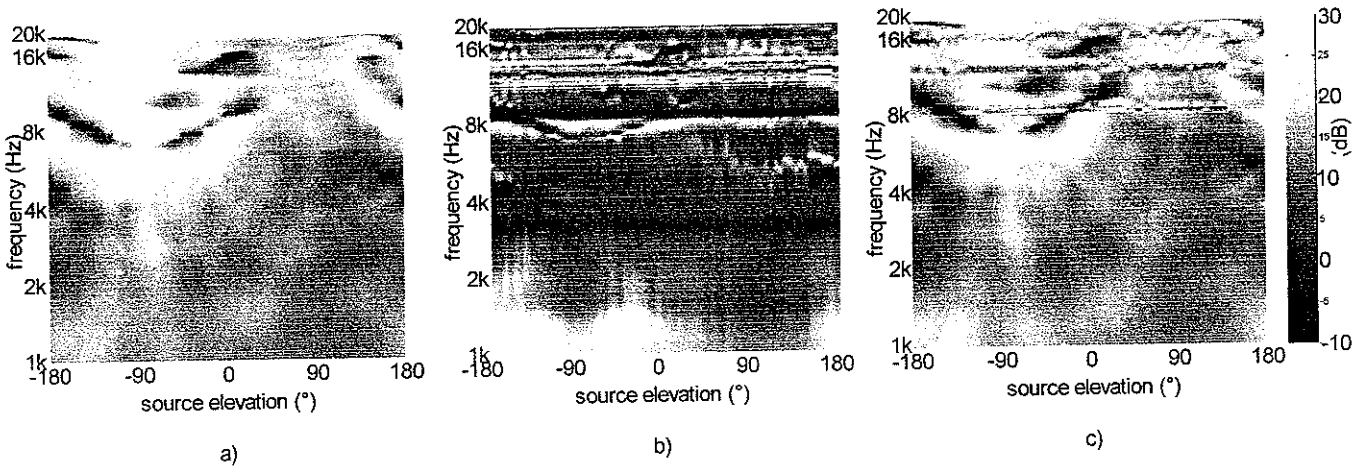

Fig. 20. Interaural spectral shape difference in HRTFs. Sound source directions are along the cone of constant azimuth (50°). a) Left panel: interaural spectral shape difference by real sound sources. b) Middle panel: interaural spectral shape difference synthesised by the 60° transducer arrangement. The listener's head is displaced 25mm laterally. c) Right panel: interaural spectral shape difference synthesised by the 10° transducer arrangement. The listener's head is displaced 25mm laterally.

ment.

Examples of another type of binaural spectral cue, the interaural level difference (ILD), are shown in Fig. 21 for sound source directions on the horizontal plane. As can be seen in Fig. 21a, which shows the ILD with real sound sources, it is not a simple task to allocate one ILD value to a particular azimuth angle. Since complex interference at higher frequencies yields multiple (often more than 4) azimuth angles for one ILD value at each frequency. In addition, the ILD value for a particular azimuth direction varies depending on frequency. The ILD with synthesised HRTFs when the listener's head is displaced 25mm laterally are shown in Fig. 21b and Fig. 21c. The ILD with the 60° transducer arrangement is degraded severely but those with the 10° span preserved well. Generally speaking, the ILD value for larger azimuth angles cannot be achieved without a very good preservation of monaural spectra for the contra-lateral ear. For example, with the 60° transducer arrangement with 50mm lateral head displacement, the ILD value (averaged over the mid-frequency range) for azimuth directions larger than ±30° cannot be achieved [14].

4 SUBJECTIVE EXPERIMENT

The virtual directional information synthesised with two different arrangements of monopole transducers were investigated by using subjective localisation experiments. Experiments with real sound sources were also performed to establish the accuracy of the experimental procedure itself. Source directions on the

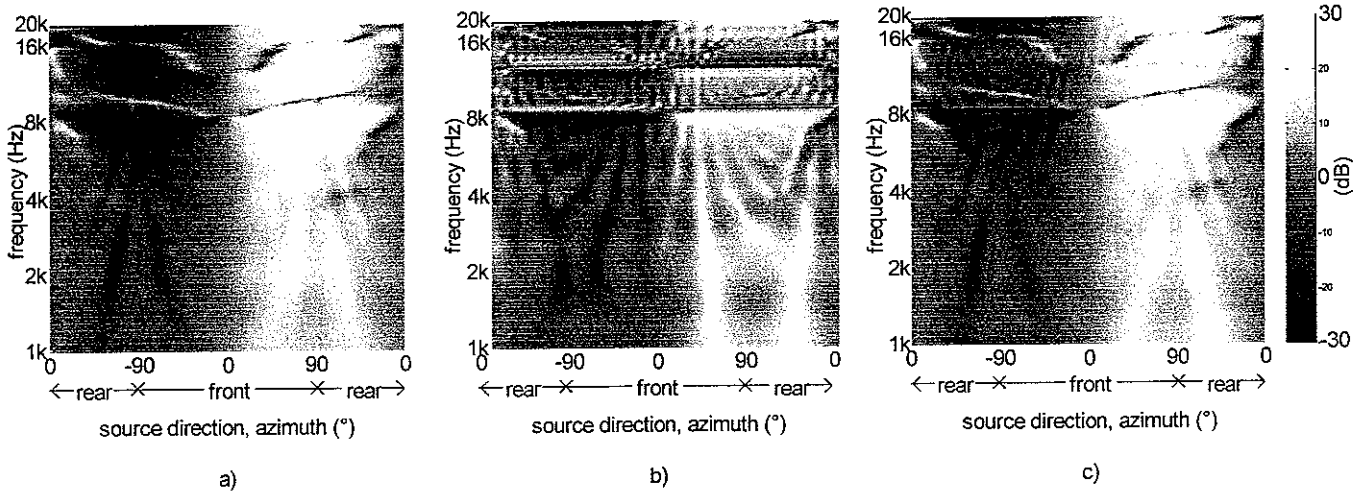

Fig. 21. Interaural level difference (ILD) for sound source directions on the horizontal plane. a) Left panel: ILD produced by real sound sources. b) Middle panel: ILD synthesised by the 60° transducer arrangement. The listener's head is displaced 25mm laterally. c) Right panel: ILD synthesised by the 10° transducer arrangement. The listener's head is displaced 25mm laterally.

horizontal plane were chosen to be examined since this covers the whole range of azimuth directions and two alternative elevation directions, i.e. 0° (front) and 180° (rear), in each cone of constant azimuth.

4.1 Procedure

A weighted noise signal (EAIJ RC-7603) was used as source signal to minimise the consequence of the large high frequency discrepancy between the HRTFs of the subjects and the KEMAR HRTFs used in the filter design procedure. The signal has a flat spectrum between 200Hz and 2kHz and gradually rolls off towards lower and higher frequencies (Fig. 22). The relative level is about -2dB at 5kHz, -5dB at 10kHz, -13dB at 20Hz and 20kHz with respect to the level between 200Hz and 2kHz. Each stimulus consisted of a reference signal and a test signal. A reference signal was presented at 0° azimuth and 0° elevation, i.e., directly in front of the listener before each test signal. Both signals had the same sound source signal with a duration of 3 seconds for the reference signal and 5 seconds for the test signal with a gap of 3 seconds in between. In order to avoid the effect of presentation order, the order of presentation from different directions was randomised. The reference stimulus not only cancelled the order effect, but also gave subjects prior knowledge of the sound source signal spectrum which is important for the monaural spectral cue. Stimuli, a set of reference and test signals, were repeated when subjects had difficulty in making a judgement.

Subjects were required to choose the closest marker to the perceived direction of sound. The markers were placed all around the head in the horizontal plane 1m from the origin of the co-ordinate with 10° intervals (Fig. 23). The subjects were allowed

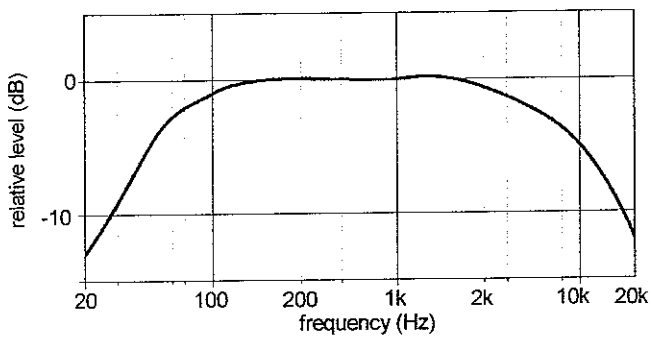

Fig. 22. The spectrum of weighted noise signal (EAIJ RC-7603) used as source signal.

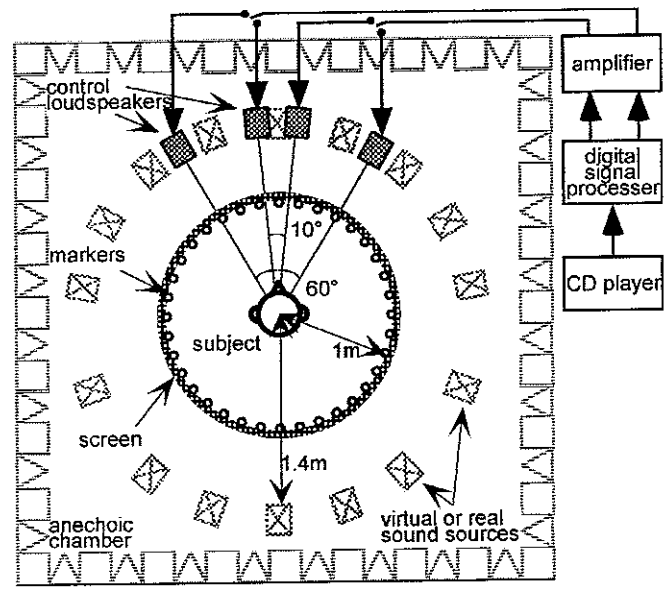

Fig. 23. The arrangement for the subjective experiment.

to choose more than one marker when they perceived two or more separate directions of sound. In order to avoid introducing dynamic cues which relate to head movement, the subject was instructed not to move the head nor body while the stimuli were presented. However, the subject was allowed to turn his head to see markers after each test stimulus had stopped. The subject's head was not physically fixed but supported by a small head rest. The subject was surrounded by a thin black curtain placed between markers and loudspeakers in order to minimise the effect of visual information (Fig. 23). Subjects were all European males with normal hearing.

The loudspeakers used had a fairly flat response between about 250Hz and 5kHz which gradually rolls off towards lower and higher frequencies (Fig. 24). The relative level at 20kHz is about 10dB smaller with respect to the frequency which gives maximum response. The characteristics of the loudspeakers were well-matched (0.5dB difference in amplitude and a few degrees difference in phase response). Difference in responses between two loudspeakers degrades the HRTF synthesis. When their responses are identical, their effects become independent of virtual source directions and can be regarded as degrading the sound source signal rather than synthesised HRTFs. The responses of the loudspeakers of course affect monaural cues and they also affect those associated with the real sound sources, but they do not affect the binaural cues. Therefore, for binaural synthesis, it is important to use a well-matched pair of loudspeakers. The loudspeaker pairs for different transducer arrangements were swapped for half of the subjects with the aim of minimising bias errors which are induced by different responses between the loudspeakers.

In order to minimise other factors than head misalignment which affect synthesis, the experiments were carried out in an anechoic chamber. The same data for the acoustic paths **a** and the control filter matrix **X** as those used in the analysis were implemented by digital filters using an MTT Lory Accel digital signal processing system. The output of the digital filters were fed via an amplifier to two pairs of loudspeakers with the same geometrical arrangements as used in the analysis. The loudspeakers as control transducers and as real sound sources were placed 1.4m from the origin of the spherical co-ordinate system (Fig. 23). It is very important to bear in mind that there is considerable amount of variability of the HRTFs among individuals. Inevitably, the matrix **C** containing each subject's HRTFs in this experiment is different from that as-

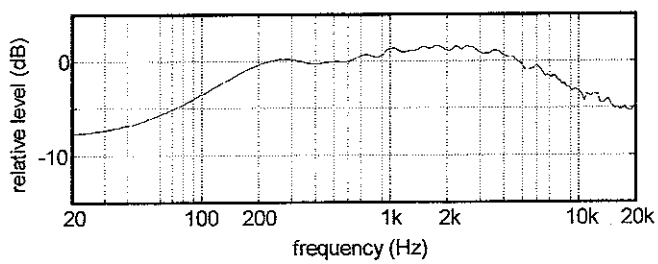

Fig. 24. The frequency response of the loudspeaker.

sumed when the matrix **X** is designed. This is the largest source of error when comparing the results with the analysis. The loudspeakers, rather than the listener's head, were displaced in both the lateral direction and in the fore-and-aft direction in order to achieve the displacement of the listeners head from the optimal position. The precision of the arrangement of the loudspeakers and listener's head was of the order of ±10mm.

The results from the subjective experiments are presented in the following format. The area of each circle in the figures indicates the number of subjects who perceived the source to be in the given direction. In cases where the subjects perceived sound sources in more than two directions, the area of the circle is distributed into those positions in accordance with the number of the directions. The dash-dot line shows the position of the circles when the perceived direction is the same as the presented direction. The dotted line is in a symmetric position to the dash-dot line with respect to the interaural axis. Therefore, the subjective responses due to front-back confusion fall around these lines.

4.2 Real sound sources

Nine real sound sources were placed at 10° increments at different azimuthal angles except ±20° and ±90°, and two sources were placed at azimuth 0° with different elevations of 0° and 180° (front and rear). Five of them were positioned in front (elevation 0°) and four of them were positioned in the rear (elevation 180°). Four of them were positioned to the left (negative azimuth) and five of them were positioned to the right (positive azimuth). The performance with real sound sources (Fig. 25) shows the localisation performance of the subjects and the accuracy of the experimental procedure itself. This therefore implies the maximum precision achievable with the following experiments with synthesised virtual sound sources. More than 60% of the responses resulted on the correct marker and more than 90% of the responses resulted within the smallest (±10°) measurable error with the method. The repeatability of the response is exceptionally good in that the responses associated with a particular direction for a particu-

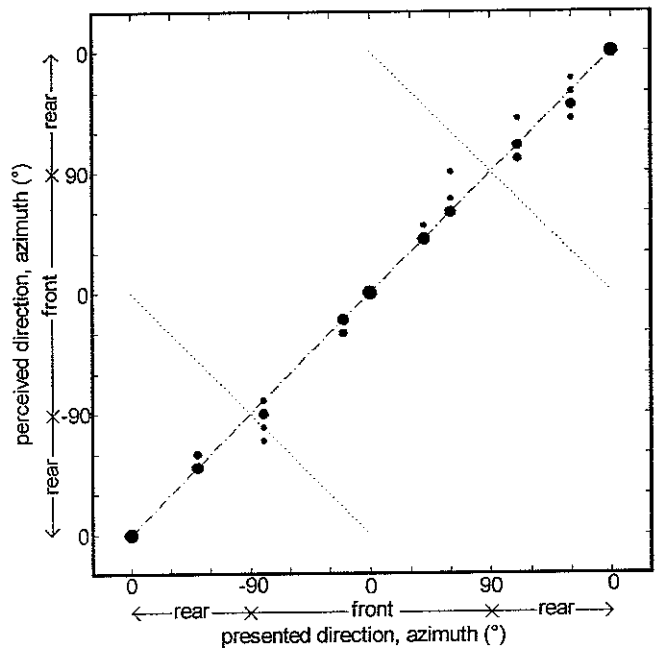

Fig. 25. Results of the subjective experiment for localising real sound sources. 6 subjects were tested.

lar subject almost always (more than 95%) resulted at the same marker (even for the wrong marker). The accuracy can be observed best at small azimuth directions (closer to the median plane) and deteriorates towards large azimuth directions (the side of the listener). There are no obvious signs of confusion along the cone of constant azimuth, i.e., front-and-back confusion. The subjects reported after the experiments that the task was very easy and did not have any ambiguity in deciding which marker to choose.

4.3 Virtual sound sources

Localisation experiments with binaural synthesis over loudspeakers were first carried out with the listener's head at the optimal position. Sixteen virtual sound sources were placed at 0°, ±20°, ±40°, ±60° and ±80° azimuth with 0° elevation (front) and 0°, ±20°, ±40° and ±70° azimuth with 180° elevation (rear). It was revealed that there was a population of subjects for whom the synthesis of virtual sound sources works reasonably well ("good" subjects) whereas it does not work so effectively for the rest of subjects ("poor" subjects). Fig. 26 shows the localisation performance for 11 subjects when the head is at the optimal position. Only a few front-back confusions can be observed with the 7 "good" subjects. The 4 "poor" subjects did not localise the virtual sound sources in the rear half of the horizontal plane correctly, and instead, localised them around symmetric positions in the front half plane. Moreover, virtual sound sources at large azimuth directions (around ±90° azimuth) were perceived at the offset position towards the centre (smaller azimuth angle). Clearly, the grouping of subjects has no relation to the different transducer span. It also has no relation to the ability of the subjects to localise real sound sources. Further investigation confirmed that a large disparity between each individual and KEMAR HRTFs resulted in inaccurate synthesis of binaural signals and resulted in systematic bias error for the "poor" subjects [29].

In principle, different transducer arrangements should not produce much difference in performance when the listener's head is at the optimal position and orientation. Nevertheless, the 10° transducer span showed slightly better performance, especially for "good" subjects around 0° azimuth where it showed no front-back confusion, contrary to the considerable amount of confusions with the 60° span. The 60° span transducer arrangement has a slight advantage around its transducer positions (around ±30° azimuth) for the "poor" subjects for obvious reasons, but poorer performance (more front-back confusion) for the "good" subjects. Although the listener's head was supposed to be at the optimal position and orientation in this experiment, some misalignment of the head is inevitable in practice. This may have caused the increase in front-back

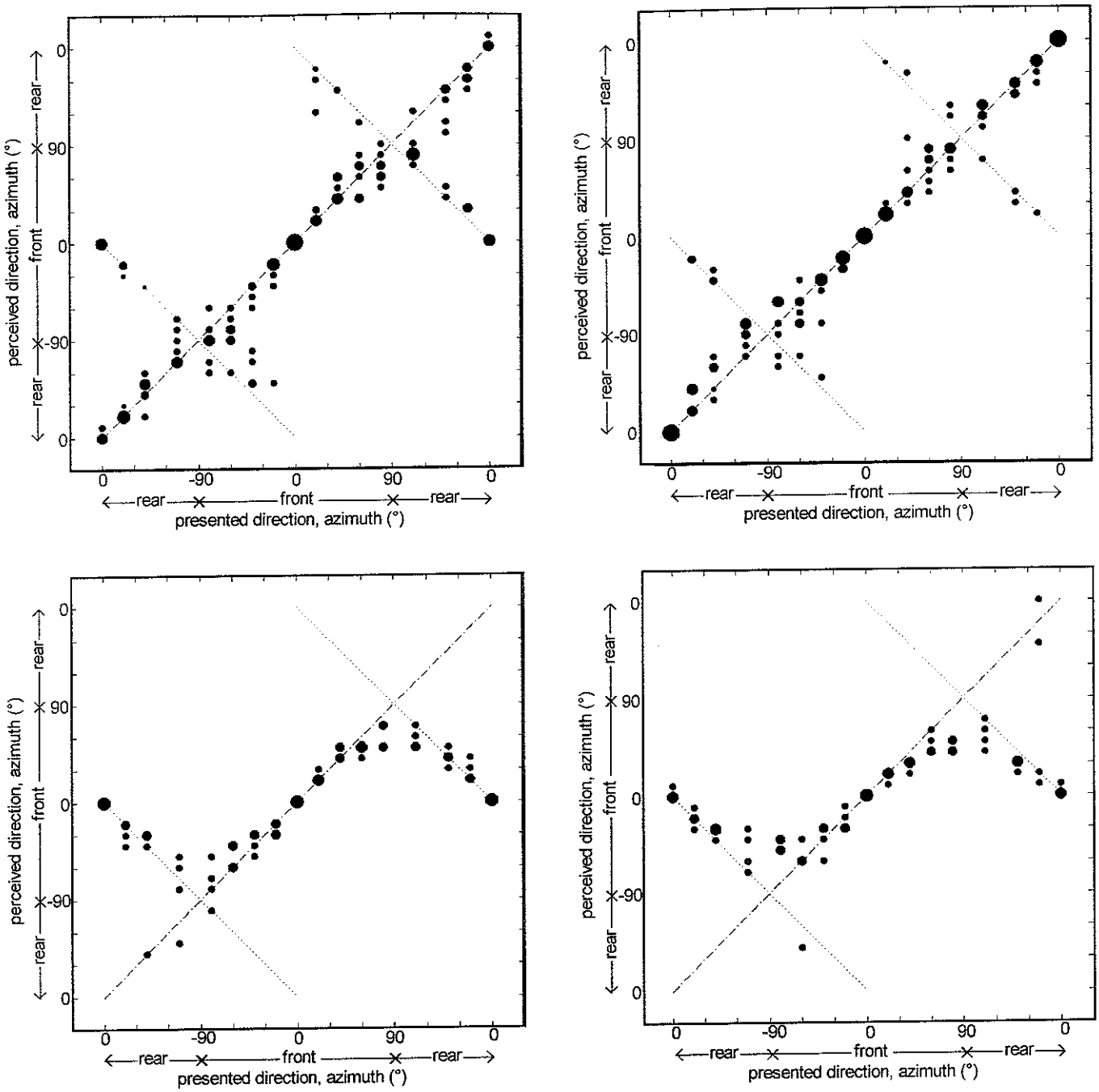

Fig. 26. Results of the localisation experiment with binaural synthesis over loudspeakers. The listener's head is at the optimal position and orientation. 11 subjects were tested. Upper row: Responses by the subjects for whom the systems work well. Lower row: Responses by the subjects for whom the systems do not work well. Left column: 60° transducer span. Right column: 10° transducer span.

confusion with the 60° transducer span. The slight unintended displacement of the head would have probably exceeded the severe limit for the good synthesis of spectral cues for the 60° transducer arrangement, even though the same displacement may have been within the required limit for the 10° transducer arrangement.

4.4 Head displacement

Further experiments with head displacement were carried out only with the 7 "good" subjects. The results when the listener's head is displaced 50mm to the right are shown in Fig. 27. The subjects reported after this experiment that the task was very difficult since sometimes they did not perceive a clear direction and sometimes they perceived source to be at multiple directional locations. The multiple perception may be the consequence of multiple maxima in the interaural cross-correlation function. Discrepancy in different cues (e.g. ITD and ILD) could also be the cause. Virtual sound sources presented by the 60° transducer arrangement intended at 0° azimuth angle (both in front and rear) are often perceived at 10°~20° offset direction, whereas the virtual sources were mostly perceived in the intended direction by the 10° arrangement. These results agree with predicted direction by the ITD analysis where a 16° offset is expected from the 60° arrangement but a 0° offset is expected from the 10° arrangement. Considerable offset around ±40°~±60° azimuth is also noticeable for the 60° arrangement. More front-back confusions for the 60° arrangement than the 10° arrangement can still be observed. Degradation of spectral shape cues does not seem to affect the performance very much since little increase of front-back confusion can be observed, although some effect may have already been in the results at the optimal head position as discussed earlier. Another possibility is that the head displacement may not have degraded the spectral shape very much more than the disparity between each individual HRTFs and the KEMAR HRTFs. A slightly better performance is

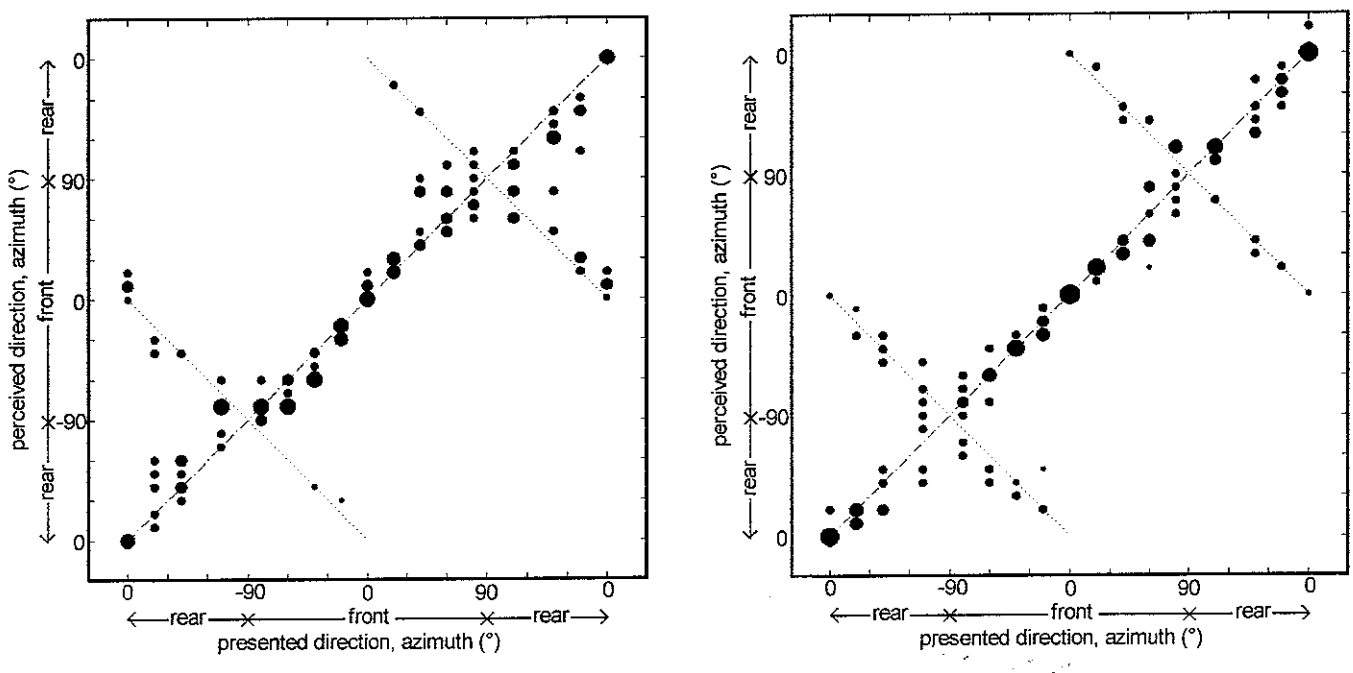

Fig. 27. Results of the localisation experiment with binaural synthesis over loudspeakers when the listener's head is displaced 50mm laterally (to the right). 7 subjects were tested. Left panel: 60° transducer span. Right panel: 10° transducer span.

observed on the side which the head is displaced to (right) for the 10° arrangement, whereas the other side (left) shows better performance for the 60° arrangement as predicted by ITD analysis. Contrary to the poor ILD values obtained, azimuth localisation seems surprisingly accurate. Considering that the additional local maxima of the cross-correlation function start to become larger than the original maximum around 25 mm displacement for the 10° arrangement and much smaller displacement for the 60° arrangement, the performance of azimuth estimation is more likely to be determined by a more plausible local maximum than by the absolute maximum of the interaural cross-correlation function, as discussed in the analysis of temporal cues.

When the listener's head is displaced 200mm to the rear, the 10° span transducer arrangement showed slightly better performance than the 60° arrangement for both azimuth localisation and front-back discrimination (Fig. 28). However, the difference in performance between the two transducer arrangements are much less significant compared to lateral displacement.

5 CONCLUSIONS

1. In binaural synthesis over two loudspeakers, yaw, lateral and roll displacement results in a shift of ITD as well as the generation of additional local maxima in the interaural cross-correlation function. Fore-and-aft, vertical and pitch displacement results only in the generation of additional local maxima. There are less

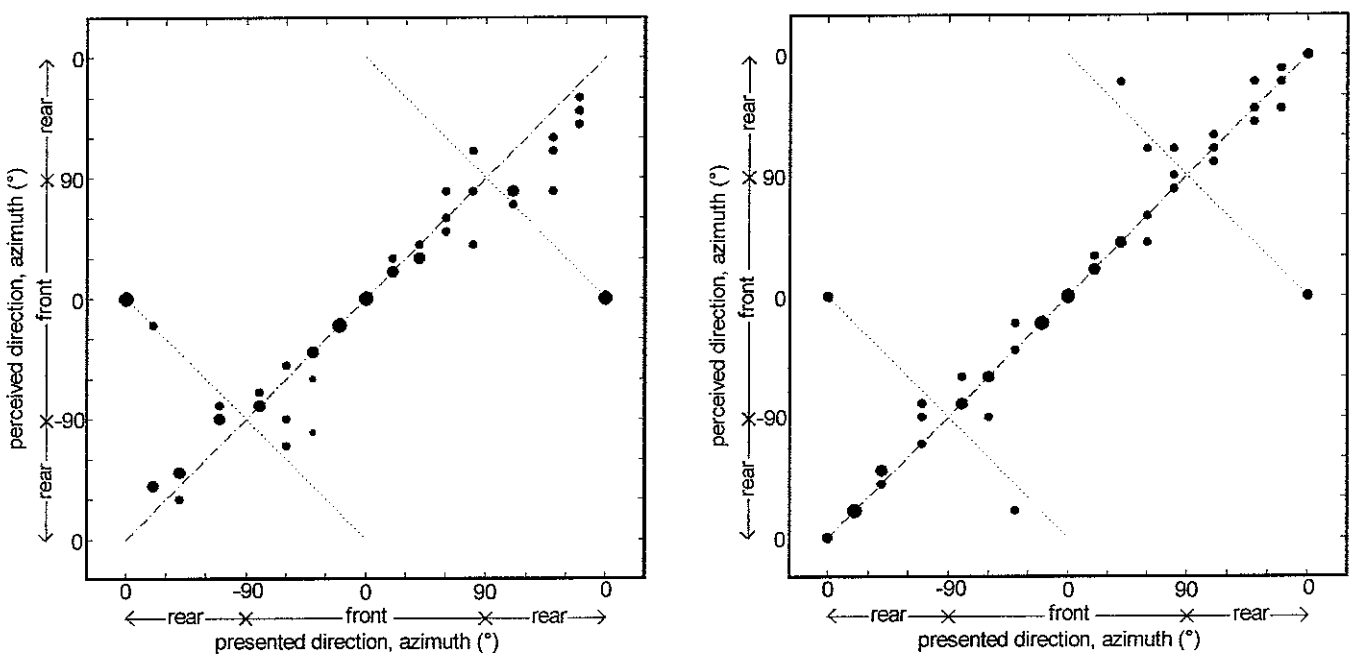

Fig. 28 Results of the localisation experiment with binaural synthesis over loudspeakers when the listener's head is displaced 200mm to the rear. 3 subjects were tested. Left panel: 60° transducer span. Right panel: 10° transducer span.

degradation of temporal cues for lateral, roll and fore-and-aft displacements when two loudspeakers are placed close together.

2. Any displacement induces more "cross-talk" components in synthesised spectra. There is less degradation of the spectral cue for lateral, fore-and-aft and yaw displacement when two loudspeakers are placed close together.

3. The ITD cue is the most robust to head misalignment followed by the monaural spectral cue for the ipsi-lateral ear. The monaural spectral cue for the contra-lateral ear is the least robust together with binaural spectral cues (including ILD cues).

4. Subjective experiments confirmed that two closely spaced loudspeakers have an advantage in performance with regard to the misalignment of the listener's head. The localisation performance with subjective experiments were better than those predicted with any one individual localisation cue. This suggests the importance of the combination of different localisation cues.

References

[1] J. Blauert, Spatial Hearing; The Psychophysics of Human Sound Localization (MIT Press, Cambridge, MA, 1997).

[2] H. Møller, "Fundamentals of Binaural Technology," Appl. Acoust. **36**, 171-218 (1992).

[3] D. R. Begault, 3-D Sound for Virtual Reality and Multimedia (AP Professional, Cambridge, MA, 1994).

[4] A. D. Blumlein, British patent specification 394.324 (1958)

[5] M. A. Gerzon, "Ambisonics in Multichannel Broadcasting and Video", J. Audio Eng. Soc., **33** (11), 859-871 (1985)

[6] A. J. Berkhout, D. de Vries, and P. Vogel, "Acoustic Control by Wave Field Synthesis," J. Acoust. Soc. Am. **93**, 2764-2778 (1993).

[7] M. R. Schroeder, B. S. Atal, "Computer Simulation of Sound Transmission in Rooms," IEEE Intercon. Rec. Pt7, 150-155 (1963).

[8] P. Damaske, "Head-related Two-channel Stereophony with Reproduction," J. Acoust. Soc. Am. **50**, 1109-1115 (1971).

[9] H. Hamada, N. Ikeshoji, Y. Ogura And T. Miura, "Relation between Physical Characteristics of Orthostereophonic System and Horizontal Plane Localisation," Journal of the Acoustical Society of Japan, (E) **6**, 143-154, (1985).

[10] J. L. Bauck and D. H. Cooper, "Generalized Transaural Stereo and Applications," J. Acoust. Soc. Am. **44** (9), 683-705 (1996).

[11] P. A. Nelson, O. Kirkeby, T. Takeuchi, and H. Hamada, "Sound fields for the production of virtual acoustic images," J. Sound. Vib. **204** (2), 386-396 (1997).

[12] O. Kirkeby, P. A. Nelson, and H. Hamada, "Stereo Dipole," UK Patent Application, 9603236.2, 1996.

[13] O. Kirkeby, P. A. Nelson, and H. Hamada, "Local Sound Field Reproduction Using Two Closely Spaced Loudspeakers," J. Acoust. Soc. Am. **104** (4), 1973-1981 (1998).

[14] T. Takeuchi, P. A. Nelson, O. Kirkeby, and H. Hamada, "Robustness of the Performance of the "Stereo Dipole" to Misalignment of Head Position," 102nd AES Convention Preprint 4464(17), (1997).

[15] P.A. Nelson, F. Orduna-Bustamante, and H. Hamada, "Inverse Filter Design and Equalisation Zones in Multi-Channel Sound Reproduction," IEEE Trans. Speech Audio Process. **3**(3), 185-192 (1995).

[16] O. Kirkeby, P. A. Nelson, F. Orduna-Bustamante, and H. Hamada, "Local Sound Field Reproduction Using Digital Signal Processing," J. Acoust. Soc. Am. **100**, 1584-1593 (1996).

[17] B. Gardner, and K. Martin, "HRTF Measurements of a KEMAR Dummy-Head Microphone," MIT Media Lab Perceptual Computing - Technical Report No. 280 (1994).

[18] T. C. T. Yin, and J. C. Chan, "Interaural Time Sensitivity in Medial Superior Olive of Cat," J. Neurophysiol. **64**, 465-488 (1990).

[19] T. R. Stanford, S. Kuwada, and R. Batra, "A Comparison of the Interaural Time Sensitivity of Neurones in the Inferior Colliculus and Thalamus of the Unanesthetized Rabbit," J. Neurophysiol. **12**, 3200-3216 (1992).

[20] T. N. Buell, C. Trahiotis, and L. R. Bernstein, "Lateralization of Low-Frequency Tones: Relative Potency of Gating and Ongoing Interaural Delays," J. Acoust. Soc. Am. **90**, 3077-3085 (1991).

[21] L. A. Jeffress, "A Place Theory of Sound Localization," J. Comp. Physiol. Psychol. **41**, 35-39 (1948).

[22] H. S. Colburn, "Theory of Binaural Interaction Based on Auditory-Nerve Data. I. General Strategy and Preliminary Results on Interaural Discrimination," J. Acoust. Soc. Am. **54**, 1458-1470 (1973).

[23] G. B. Hanning, "Detectability of Interaural Delay in High-frequency complex waveforms," J. Acoust. Soc. Am. **55**, 84-90 (1974).

[24] J. C. Middlebrooks, and D. M. Green, "Directional Dependence of Interaural Envelope Delays," J. Acoust. Soc. Am. **87**, 2149-2162 (1990).

[25] R. A. Butler, and R. Flannery, "The Spatial Attributes of Stimulus Frequency and Their Role in Monaural Localisation of Sound in the Horizontal Plane," Percept. psychophys. **28**, 449-457 (1980).

[26] C. Lim, and R. O. Duda, "Estimating the Azimuth and Elevation of a Sound Source from the Output of a Cochlea Model," Proc. Twenty-eighth Annual Asilomer Conference on Signals, Systems and Computers (IEEE, Asilomar, CA), 399-403 (1994).

[27] R. G. Klump, and H. R. Eady, "Some Measurements of Interaural Time Difference Thresholds," J. Acoust. Soc. Am. **28**, 859-860 (1956).

[28] F. L. Wightman, and D. J. Kistler, "The Dominant Role of Low-frequency Interaural Time Differences in Sound Localization," J. Acoust. Soc. Am. **91**, 1648-1661 (1992).

[29] T. Takeuchi, P.A. Nelson, O. Kirkeby and H. Hamada, "Influence of Individual Head Related Transfer Function on the Performance of Virtual Acoustic Imaging Systems", 104th AES Convention Preprint 4700 (P4-3).