

**A Study of Virtual Acoustic Imaging Systems for Asymmetric  
Listener Locations**

**J. Rose, P.A. Nelson, B. Rafaely and T. Takeuchi**

ISVR Technical Report No 295

November 2001



## 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 be 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.

### COPYRIGHT NOTICE

(c) ISVR University of Southampton All rights reserved.

ISVR authorises you to view and download the Materials at this Web site ("Site") only for your personal, non-commercial use. This authorization is not a transfer of title in the Materials and copies of the Materials and is subject to the following restrictions: 1) you must retain, on all copies of the Materials downloaded, all copyright and other proprietary notices contained in the Materials; 2) you may not modify the Materials in any way or reproduce or publicly display, perform, or distribute or otherwise use them for any public or commercial purpose; and 3) you must not transfer the Materials to any other person unless you give them notice of, and they agree to accept, the obligations arising under these terms and conditions of use. You agree to abide by all additional restrictions displayed on the Site as it may be updated from time to time. This Site, including all Materials, is protected by worldwide copyright laws and treaty provisions. You agree to comply with all copyright laws worldwide in your use of this Site and to prevent any unauthorised copying of the Materials.

UNIVERSITY OF SOUTHAMPTON  
INSTITUTE OF SOUND AND VIBRATION RESEARCH  
FLUID DYNAMICS AND ACOUSTICS GROUP

**A Study of Virtual Acoustic Imaging Systems for  
Asymmetric Listener Locations**

by

**J Rose, P A Nelson, B Rafaely and T Takeuchi**

ISVR Technical Report No. 295

November 2001

Authorized for issue by  
Professor C L Morfey, Group Chairman

© Institute of Sound & Vibration Research



## CONTENTS

CONTENTS .....	ii
LIST OF FIGURES .....	iii
ABSTRACT .....	v
I. INTRODUCTION .....	1
II. FILTER DESIGN .....	2
III. COMPUTER SIMULATIONS .....	2
A. HRTF APPROXIMATIONS .....	3
1. <i>Free field approximation</i> .....	3
2. <i>Spherical head approximation</i> .....	4
3. <i>Dummy head</i> .....	4
B. RINGING FREQUENCY .....	5
C. CONDITION NUMBER .....	6
D. CROSS-TALK CANCELLATION EFFECTIVENESS .....	6
E. CALCULATING THE SWEET SPOT BOUNDARY FROM JUST NOTICEABLE INTERAURAL TIME DIFFERENCE .....	9
IV. SUBJECTIVE EXPERIMENT .....	10
A. PROCEDURE .....	10
B. RESULTS .....	11
C. LIMITATIONS OF RESULTS .....	12
V. CONCLUSION.....	13
REFERENCES.....	14
FIGURES .....	16



## LIST OF FIGURES

Fig. 1. Plan view depicting lateral head translations and the “stereo dipole” geometrical arrangement. The elements of the electroacoustic path plant matrix  $C$  and the virtual paths  $a$  are shown as well as the corresponding distances used in equations (5,6).

Fig. 2. Block diagram of the implemented virtual acoustic imaging system.

Fig. 3. Variables used when calculating the sound field scattered from a rigid sphere.

Fig. 4. Magnitudes of KEMAR’s left ear HRTFs at angles used for the simulations. For angles in between the curves shown, linear interpolation was performed. This appears to be a reasonable approximation with a possible exception around the dip at about 8kHz.

Fig. 5. Calculated results for the  $10^\circ$  loudspeaker arrangement at head positions 0-2m off-axis a) free field ringing frequency, b) free field condition number (maximum values of 280,000 at 1Hz and 1,100 at 22kHz), c) rigid sphere model condition number (maximum values of 13,000 at 1Hz and 74 at 22kHz), and d) KEMAR dummy condition number (maximum values of 50,000 at 1Hz, 2,900 at 8kHz, and 1,200 at 16kHz).

Fig. 6. Elements  $|R_{11}|$  and  $|R_{12}|$  of control performance matrix  $R$  for filters designed at the on-axis head location. The left column shows results when the head is displaced 0.1mm from optimum. The right column shows results when the head is displaced 5cm from optimum. The calculations are based on a-b) the free field approximation, c-d) the spherical head model, and e-f) KEMAR dummy HRTFs.

Fig. 7. Cross-talk cancellation effectiveness for the left ear (left column) and right ear (right column) at designed virtual acoustic imaging filter listener locations a-b) on-axis, c-d) 50cm off-axis, e-f) 1m off-axis, and g-h) 2m off-axis. In these figures, the abscissa represents the head position relative to the assumed designed optimum listener location, the ordinate represents frequency and the grey scale represents the ratio of the cross-talk path to the direct path in decibels. Good cross-talk cancellation performance is achieved in the dark areas. All of the calculations are based on the free field model.

Fig. 8. Cross-talk cancellation effectiveness for the left ear (left column) and right ear (right column) at designed virtual acoustic imaging filter listener locations a-b) on-axis, c-d) 50cm off-axis, e-f) 1m off-axis, and g-h) 2m off-axis. In these figures, the abscissa represents the head position relative to assumed designed optimum listener location, the ordinate represents frequency and the grey scale represents the ratio of the cross-talk path to the direct path in decibels. Good cross-talk cancellation performance is achieved in the dark areas. All of the calculations are based on the spherical head model.

Fig. 9. Cross-talk cancellation effectiveness for the left ear (left column) and right ear (right column) at designed virtual acoustic imaging filter listener locations a-b) on-axis, c-d) 50cm off-axis, e-f) 1m off-axis, and g-h) 2m off-axis. In these figures, the abscissa represents the head position relative to assumed designed optimum listener location, the ordinate represents frequency and the grey scale represents the ratio of the cross-talk path to the direct path in



decibels. Good cross-talk cancellation performance is achieved in the dark areas. All of the calculations are based on measured KEMAR dummy HRTFs.

Fig. 10. Boundaries of the sweet spot for a range of head locations  $\pm 2\text{m}$  off-axis as calculated from a 10dB cross-talk cancellation performance criterion in the frequency range 0.3-3kHz when using a,b) the free field model, c,d) spherical head model, and e,f) the KEMAR dummy. The left and right columns are for the listener's left and right ears respectively. Dashed lines show the positions examined subjectively in section IV.

Fig. 11. Boundaries of the sweet spot for a range of head locations  $\pm 2\text{m}$  off-axis as calculated from a  $10\mu\text{s}$  shift in ITD of the single virtual source represented by the small circle in a) when using b) the free field model, c) spherical head model, and d) the KEMAR dummy. Dashed lines show the positions examined subjectively in section IV.

Fig. 12. Boundaries of the sweet spot for a range of head locations  $\pm 2\text{m}$  off-axis as calculated from a  $10\mu\text{s}$  shift in ITD of either of the two virtual sources represented by the small circles in a) when using b) the free field model, c) spherical head model, and d) the KEMAR dummy.

Fig. 13. Boundaries of the sweet spot for a range of head locations  $\pm 2\text{m}$  off-axis as calculated from a  $10\mu\text{s}$  shift in ITD of any of the 18 virtual sources represented by the small circles in a) when using b) the free field model and c) the KEMAR dummy.

Fig. 14. Set-up for subjective experiment inside anechoic chamber.

Fig. 15. Plan view of the experimental arrangement with the subject inside the sphere, angles marked, and loudspeakers on a moveable slide.

Fig. 16. Plan view depicting the 6 designed virtual acoustic imaging filter positions that were examined subjectively. The small circles represent the point at the middle of the two loudspeakers for the 6 positions.

Fig. 17. Subjects' responses for the 5cm off-axis designed virtual acoustic imaging filter position. Different symbols denote different subjects. Responses are shown with a) loudspeakers moving away from designed filter location, to the left, b) loudspeakers moving away from designed filter location, to the right, c) loudspeakers moving toward the designed filter location, from the left, and d) loudspeakers moving toward the designed filter location, from the right.

Fig. 18. Average subject responses for the designed virtual acoustic imaging filter positions 0-25cm off-axis. Different lines represent different designed filter positions. Responses are shown with a) loudspeakers moving away from designed filter location, to the left, b) loudspeakers moving away from designed filter location, to the right, c) loudspeakers moving toward the designed filter location, from the left, and d) loudspeakers moving toward the designed filter location, from the right.



## ABSTRACT

Virtual acoustic imaging systems are effective when the listener's head location is close to the head location assumed when the system was designed. The sweet spot refers to the spatial bubble of head location in which the system is still effective. Previous work investigating the "stereo dipole" acoustic imaging system shows that for the traditional on-axis listener location the sweet spot is about  $\pm 5\text{cm}$  for lateral head translations. Larger head movements than this require an update of the virtual acoustic imaging filters. The interest here is the sweet spot size at off-axis asymmetric listener locations or an understanding of how often one needs to update the filters to ensure the listener perceives a stable virtual image as they move. The examination of the off-axis sweet spot size comprises a theoretical acoustic analysis, computer simulations, and a subjective study. The simulations and subjective evaluation both demonstrate that the width of tolerable lateral head translations is comparable for the symmetric on-axis listener location and asymmetric listener locations that are as far as 25cm off-axis.



## I. INTRODUCTION

The principal idea of binaural technology<sup>1</sup> is to present listeners with sound signals at their two ears that coincide with the signals of a real auditory environment. This gives listeners the perception of virtual acoustic images existing where there are no real sound sources. Virtual acoustic imaging systems can use headphones to present the binaural signals<sup>1,2</sup> or loudspeakers<sup>3-5</sup>. One problem with using loudspeakers is cross-talk. This occurs when the right ear can hear the signal meant for the left ear and vice versa. Proper filtering of the signals input into the loudspeakers overcomes this difficulty<sup>3-6</sup>. The design of these filters incorporates knowledge of the impulse responses from the loudspeakers to the listener's ears. These impulse responses depend on the listener's head location and so the design of the filters assumes a listener location. Without physically restraining the listener's head to a single location, one expects the listener's actual position to stray from the optimal listening location. Virtual acoustic imaging systems tolerate some head displacement before the listener begins to lose the desired perception of the virtual acoustic images<sup>7</sup>. The sweet spot or equalization zone is the name given to the spatial bubble of head location, in which the virtual acoustic imaging system is effective. The geometrical arrangement of the system affects the sweet spot size. This report concentrates on one particular virtual acoustic imaging system known as the "stereo dipole", which has been shown to be particularly robust to head movements<sup>5,7-10</sup>. The "stereo dipole" utilizes an arrangement of two loudspeakers subtending an angle of  $10^\circ$  with the listener's head, as shown in Fig. 1. If one is given a way to track the listener's movements, such as by employing a video camera and an image processing head tracking algorithm<sup>11</sup>, the knowledge of the varying head position can be used to continuously select appropriate virtual acoustic imaging filters so that the listener location is always close to the optimum location. An adaptive system needs only to update the virtual acoustic imaging filters before the desired subjective impression begins to degrade. Therefore, it is important to know the sweet spot size at off-axis locations before implementing such a system.

Takeuchi and Nelson<sup>7</sup> examine the sweet spot size for the traditional symmetric on-axis listener location. The attempt here is to understand where the boundary of the sweet spot lies for both the traditional symmetric on-axis listener location and asymmetric off-axis listener locations. The study is restricted to lateral head translations. Figure 1 illustrates this type of head movement. The work here should lead to an understanding of the required spatial resolution of virtual acoustic imaging filters in order to achieve a seamless performance from an adaptive virtual acoustic imaging system. The results suggest that asymmetric listener locations, up to 25cm off-axis, are as robust to head displacements as the traditional on-axis listener location.

The report's presentation is as follows. Section II provides a brief overview of the virtual acoustic imaging filter design procedure. Section III presents computer simulations, which offer insight into the physical nature of the sweet spot at different head positions. In addition, section III gives derivations of sweet spot boundaries as calculated from a cross-talk cancellation performance criterion and a strict just noticeable interaural time difference criterion. Section IV presents a subjective localization experiment with results that support the derived sweet spot size of section III. Section V is the conclusion.

## II. FILTER DESIGN

This section presents an overview of well-established material<sup>4,7-12</sup>. Figure 2 shows a block diagram of the implementation of the virtual acoustic imaging system in the  $z$ -domain. The virtual paths are the vector of filters  $\mathbf{a}(z)$  that represent the transfer functions from a virtual source to the listener's ears. The plant matrix  $\mathbf{C}(z)$  represents the transfer functions of the paths from the two real sources to the two ears.

$$\mathbf{a}(z) = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}, \quad \mathbf{C}(z) = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \quad (1a,b)$$

Figure 1 provides an illustration of the elements in equation (1). The goal of the system is to produce at the ears of the listener a vector of desired binaural signals  $\mathbf{d}(z)$ . One acquires the desired signals  $\mathbf{d}(z)$  by filtering the source signal  $S(z)$  with  $\mathbf{a}(z)$ . The method of presenting the listener with the sound through two loudspeakers ensures a modification of  $\mathbf{d}(z)$  by the plant matrix  $\mathbf{C}(z)$ . The problem is to design a matrix of filters  $\mathbf{X}(z)$  that will cancel the effect of  $\mathbf{C}(z)$  i.e.,

$$\mathbf{X}(z) = \mathbf{C}(z)^{-1}. \quad (2)$$

This criterion is relaxed by allowing for some error and by employing a modelling delay  $\Delta$  to ensure a realizable causal stable inverse so that equation (2) becomes,

$$\mathbf{C}(z)\mathbf{X}(z) \approx z^{-\Delta}\mathbf{I}, \quad (3)$$

where  $\mathbf{I}$  is the identity matrix. The filter matrix  $\mathbf{X}(z)$  is the cross-talk cancellation filter matrix, and its job is to cancel the effect of the plant matrix  $\mathbf{C}(z)$ . In the frequency domain, the vector of virtual acoustic imaging filters  $\mathbf{h}(z)$  is given by the matrix multiplication of the cross-talk cancellation filters  $\mathbf{X}(z)$  and the virtual path vector  $\mathbf{a}(z)$ , i.e.

$$\mathbf{h}(z) = \mathbf{X}(z)\mathbf{a}(z). \quad (4)$$

## III. COMPUTER SIMULATIONS

This section considers the basic physics of the problem to gain some understanding into how designing virtual acoustic imaging filters at off-axis locations affect the sweet spot throughout the audible frequency range. Subsection A presents the three head related transfer function (HRTF) models that the computer simulations employ. Subsection B introduces the concept of ringing frequency and considers its relation to geometry and robustness of the virtual acoustic imaging system. Subsection C considers the condition number of the plant  $\mathbf{C}$  as a function of head position and its relation to robustness. Subsection D considers the performance of cross-talk cancellation at different head positions and uses a performance criterion to derive a sweet spot size. Finally, subsection E uses an interaural time difference criterion to derive a sweet spot size.

## A. HRTF approximations

It is possible to measure directly the path responses, shown in Fig. 1, on each individual listener<sup>13</sup>. This approach is impractical if there are many different or unknown people that will use the virtual acoustic imaging system. This subsection describes three alternative approaches to modelling HRTFs. As the models progress in realism, their complexities also increase. Subsections B-E present simulations that employ the models. Comparing the results of the different models helps attribute certain aspects of the results to specific model characteristics. The free field model provides insight into geometrical effects. In addition to geometrical effects, the spherical head model well approximates the shadowing effects of the listener's head. The dummy head model includes these effects and the effects of the listener's pinnae and torso.

### 1. Free field approximation

The simplest approximation is to remove the physical head and replace it with two monopole receivers at the position of the ears. In addition, point monopole sources model the loudspeakers and the environment for this entire arrangement is anechoic. This conceptual situation has the advantage of having a simple analytical solution. The acoustic complex pressure  $p$  due to a point monopole source at a distance  $r$  from the source is<sup>14</sup>

$$p(r) = \frac{j\omega\rho_0 q e^{-jkr}}{4\pi r} \quad (5)$$

where an  $e^{j\omega t}$  time dependence is assumed and where  $k$ ,  $\rho_0$ ,  $c_0$ ,  $\omega$ , and  $q$  are the wave number  $\omega/c_0$ , density of the medium, sound speed, angular frequency, and effective complex source strength respectively. By designating complex acoustic pressure  $p$  as the output and complex source volume acceleration  $j\omega q$  as the input the transfer functions of equation (1a,b) become,

$$\mathbf{a}(j\omega) = \frac{\rho_0}{4\pi} \begin{bmatrix} \frac{e^{-jkr_{a1}}}{r_{a1}} \\ \frac{e^{-jkr_{a2}}}{r_{a2}} \end{bmatrix}, \quad \mathbf{C}(j\omega) = \frac{\rho_0}{4\pi} \begin{bmatrix} \frac{e^{-jkr_{11}}}{r_{11}} & \frac{e^{-jkr_{12}}}{r_{12}} \\ \frac{e^{-jkr_{21}}}{r_{21}} & \frac{e^{-jkr_{22}}}{r_{22}} \end{bmatrix} \quad (6a,b)$$

where Fig. 1 shows the distances  $r_{11}$ ,  $r_{12}$ ,  $r_{21}$ ,  $r_{22}$ ,  $r_{a1}$ , and  $r_{a2}$ . In this case, the inverse of the plant matrix  $\mathbf{C}^{-1}$  has a simple analytical solution and equation (2) becomes,

$$\mathbf{X}(j\omega) = \mathbf{C}(j\omega)^{-1} = \frac{4\pi}{\rho_0} \begin{pmatrix} 1 \\ \frac{e^{-jk(r_{11}+r_{22})}}{r_{11}r_{22}} - \frac{e^{-jk(r_{12}+r_{21})}}{r_{12}r_{21}} \end{pmatrix} \begin{bmatrix} \frac{e^{-jkr_{22}}}{r_{22}} & -\frac{e^{-jkr_{12}}}{r_{12}} \\ \frac{e^{-jkr_{21}}}{r_{21}} & \frac{e^{-jkr_{11}}}{r_{11}} \end{bmatrix}. \quad (7)$$

All of the free field computer simulations discussed below, employ equations (6,7). The monopole receivers (ears) are always set at 18cm apart. The head position changes by varying

the distances ( $r_{11}$ ,  $r_{12}$ , etc.) appropriately. The low computation required for this type of simulation provides quick results that provide insight into the basic effects of the geometry.

## 2. Spherical head approximation

A common approach to improve the approximation is to model the head as a perfectly rigid sphere where the two ears are on the surface at ends of a diameter. By again assuming an anechoic environment and modelling the sound sources as point monopole sources this model yields an analytical solution. Taking the complex source volume acceleration  $j\omega q$  of a point monopole source as the input and acoustic complex pressure  $p$  on the surface of the sphere as the output, the total frequency response transfer function  $C_t(j\omega)$  is<sup>5</sup>

$$C_t(j\omega) = C_{ff}(j\omega) + C_s(j\omega). \quad (8)$$

The two components of equation (8) correspond to the incident  $C_{ff}(j\omega)$  and scattered  $C_s(j\omega)$  sound fields. Equations (9-11) calculate these elements.

$$C_{ff}(j\omega) = -\frac{j\rho_0 k}{4\pi} \sum_{m=0}^{\infty} (2m+1) j_m(ka) \times [j_m(kr) - jn_m(kr)] P_m(\cos\phi) \quad (9)$$

$$C_s(j\omega) = \frac{\rho_0 k}{4\pi} \sum_{m=0}^{\infty} b_m [j_m(ka) - jn_m(ka)] P_m(\cos\phi) \quad (10)$$

$$b_m = j(2m+1) \frac{j_m(kr) - jn_m(kr)}{1 - \frac{jn'_m(ka)}{j'_m(ka)}} \quad (11)$$

In these equations  $r$  is the distance from the source to the centre of the sphere,  $a$  is the sphere's radius and the angle  $\phi$  is defined in Fig. 3. The functions  $j_m$  and  $n_m$  are  $m$ th-order spherical Bessel functions of the first and second kind respectively and  $P_m$  is the  $m$ th-order Legendre polynomial.

At frequencies near and below 1kHz, this model well approximates the shadowing effects of a human head. The model's predictions of the binaural localization cue, interaural time delay (ITD), agree closely with measurements made on man-shaped dummies<sup>4,15</sup> and on subjects<sup>16</sup>. The binaural localization cue, interaural level difference (ILD), is somewhat underestimated by this model due to the absence of a neck<sup>15</sup>. A practical disadvantage of this model is the slow convergence of the Bessel functions making the calculations time consuming. In the simulations that follow, the sphere was set to have a radius of 9cm.

## 3. Dummy head

The previous two models made no effort to model the effects of the listener's pinnae and torso. The last method employed makes use of direct measurements on a KEMAR (Knowles Electronics Manikin for Acoustic Research) dummy<sup>17</sup>. The KEMAR dummy has median human adult dimensions including its pinnae. The acoustic behaviour of KEMAR's ear canals and eardrum simulators matches that of real ears. The MIT Media Lab measured HRTFs of a

KEMAR dummy in an anechoic chamber at a sampling frequency of 44.1kHz<sup>18</sup>. This database is available for downloading<sup>19</sup>. The measurements at 0° elevation at a radial distance of 1.4m for the full 360° of azimuth sampled in 5° increments are employed below with the loudspeaker responses deconvolved.

To match the angle and distance of interest some manipulation of the data is necessary. The first manipulation is linear interpolation in the frequency domain for angle locations between those measured. The frequency responses shown in Fig. 4 are of thirteen, measured, left ear HRTFs at the angles that are of primary interest. In view of this figure, linear interpolation seems to be a reasonable approximation with the possible exception of frequencies close to the dip at about 8kHz. The next manipulation is to correct for distance by scaling the magnitude according to the  $1/r$  law of spherical spreading<sup>20</sup> and by applying a fractional delay FIR filter to account for the time difference. The fractional delay filter is a shifted truncated sinc function<sup>21</sup>. Error introduced by this filter is concentrated around the Nyquist frequency (22.05kHz) due to the Gibbs phenomenon. The actual time delay added in the simulations is determined by assuming a 344m/s speed of sound.

## B. Ringing frequency

Employing the relationship for a geometric series of the type

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n \quad (\text{for } |x| < 1) \quad (12)$$

to equation (7) yields,

$$\mathbf{X}(j\omega) = \frac{4\pi}{\rho_0} \begin{bmatrix} r_{11} e^{jk r_{11}} & -\frac{r_{11} r_{22}}{r_{12}} e^{-jk(r_{12}-r_{11}-r_{22})} \\ -\frac{r_{11} r_{22}}{r_{21}} e^{-jk(r_{21}-r_{11}-r_{22})} & r_{22} e^{jk r_{22}} \end{bmatrix} \sum_{n=0}^{\infty} \left( \frac{r_{11} r_{22}}{r_{12} r_{21}} \right)^n e^{-jk(r_{12}+r_{21}-r_{11}-r_{22})n} \quad (13)$$

Equation (13) reveals cross-talk cancellation as inherently a recursive process with an associated “ringing” frequency  $f_r$  related to the path length differences.

$$f_r = \frac{c_0}{r_{12} + r_{21} - r_{11} - r_{22}} \quad (14)$$

Employing equation (12) to the infinite summation in equation (13) yields,

$$\frac{1}{1 - \frac{r_{11} r_{22}}{r_{12} r_{21}} e^{-jk(r_{12}+r_{21}-r_{11}-r_{22})}} \quad (15)$$

which approaches infinity at frequencies  $f = 0, f_r, 2f_r, 3f_r, \dots$  for  $\frac{r_{11}r_{22}}{r_{21}r_{12}} \approx 1$ . The next subsection shows these frequencies to be ill-conditioned and therefore not very robust.

Figure 5a shows the ringing frequency for listener positions 0-2m off-axis for the  $10^\circ$  loudspeaker arrangement shown in Fig. 1. At the on-axis location,  $f_r$  is about 11kHz. As the listener moves off-axis, the ringing frequency increases monotonically. At about 110cm off-axis  $f_r$  passes beyond the audible frequency range.

### C. Condition number

The condition number with respect to matrix inversion is the ratio of the matrix's maximum to minimum singular value. This is a measure (for a system of linear equations) of the solution's sensitivity or "vulnerability" to small perturbations or errors. A well-conditioned matrix has a condition number equal or close to one. An ill-conditioned matrix has a large condition number.

In this case, the interest is in the sensitivity to error of system inversion of an assumed plant matrix  $C$ . A source of error in the plant matrix  $C$  is in differences of the HRTF model used and the actual listener's HRTF. These differences include head displacements from the assumed listener location. Therefore,  $C$ 's condition number reflects the robustness of the system to head displacements.

Figures 5b-5d show the condition number of  $C$  as a function of frequency for head locations 0-2m off-axis for the free field approximation (Fig. 5b), spherical head model (Fig. 5c), and for the KEMAR dummy (Fig. 5d). In these figures, the abscissa represents the head position relative to the inter-source axis and the ordinate represents frequency. The grey scale represents the condition number with black and white corresponding to low and high condition numbers respectively. The free field and sphere models both show ill conditioning at the ringing frequency and at very low frequencies. The dummy HRTF is ill conditioned at very low frequencies but also at other frequencies especially at farther off-axis positions. The dummy is also ill conditioned around the 8kHz dip (Fig. 3) for head positions greater than 1m off-axis.

The condition number simulations using the free field and spherical head approximations predict off-axis positions to be more robust over a greater frequency range than on-axis. When including pinnae and torso effects the condition number of the dummy predicts a more robust system on-axis.

### D. Cross-talk cancellation effectiveness

The effect of the listener's location on cross-talk cancellation performance of the system is now considered. A cross-talk cancellation performance criterion is a useful basis to estimate the sweet spot size. Introducing a control performance matrix  $R$  helps evaluate cross-talk cancellation performance.

$$\mathbf{R} = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix} = \mathbf{C}\mathbf{X}. \quad (16)$$

Perfect cross-talk cancellation occurs when the diagonal terms of  $\mathbf{R}$ , or the direct paths, are time shifted delta functions in the time domain and the off-diagonal terms of  $\mathbf{R}$ , or the cross-talk paths, are zero. Channel separation refers to a ratio of two of these responses to one ear (e.g.  $R_{12}/R_{11}$  or  $R_{21}/R_{22}$ )<sup>4</sup>. This is essentially a measure of the effectiveness of the cross-talk cancellation process.

Figure 6 shows the magnitudes of the frequency responses  $R_{11}$  and  $R_{12}$  for head displacements 0.1mm and 5cm from the optimal on-axis location with the free field approximation (Figs. 6a,b), the spherical head model (Figs. 6c,d), and the KEMAR dummy (Figs. 6e,f). Optimally,  $|R_{11}|$  should be 0dB over the whole frequency range and  $|R_{12}|$  should be minus infinity. For a 0.1mm head displacement  $|R_{11}|$  is very flat over the whole frequency range and  $|R_{12}|$  is predominately below  $-40$ dB. A peak in  $|R_{12}|$  occurs at the ringing frequency at about 11kHz. The dummy's results show poor cross-talk cancellation above 20kHz due to the error introduced by the fractional delay interpolation filter. A head displacement of 5cm degrades performance greatly. At this head displacement, Figs. 6b,d,f show about 10dB cross-talk cancellation below 5kHz but poor performance above this frequency.

Figures 7-9 show the cross-talk cancellation performance at both ears for filters designed for the on-axis head location and at 50cm, 1m, and 2m off-axis. In these figures, the abscissa represents the head position relative to the assumed design location, the ordinate represents frequency, and the grey scale represents the ratio of the cross-talk path to the direct path in decibels. The dark areas represent good cross-talk cancellation. The bases of the calculations for Figs. 7, 8, and 9 are the free field approximation, spherical head model, and the KEMAR dummy measurements respectively. In these figures, the right and left columns show the results for the left and right ear respectively.

The basis of the calculations for Fig. 9 is the free field approximation. The vertical white lines at head displacements of  $\pm 18$ cm correspond to situations where an ear is at the intended location for the other ear. Consideration of Figs. 7a,b reveals that at on-axis the system tolerates head displacements of about 4-5cm at low frequencies. Around 11kHz (the ringing frequency), the on-axis system tolerates practically no head displacement. Above the ringing frequency, the system tolerates about 2-3cm displacements from the optimum position. Figures 7c,d show the cross-talk cancellation performance for filters designed at 50cm off-axis. The performance is similar to that at the on-axis position except that the ringing frequency is at about 13kHz. At 1m off-axis (Figs. 7e,f), the ringing frequency is at about 20kHz. At 2m off-axis (Figs. 7g,h), the ringing frequency is beyond the audio frequency range. These results imply that an expansion of the frequency range where the system is the most robust occurs as the result of a higher ringing frequency for off-axis systems. The higher ringing frequency results from a decrease in the difference in the path lengths.

Figure 8 shows results using the spherical head model. This model reveals the same behaviour of the ringing frequency. However, when including the shadowing effects of the sphere, the cross-talk cancellation at the right ear (right column) appears to be much more robust than the

left ear (left column). The left ear appears to have the most robust cross-talk cancellation on-axis. As filters are designed farther off-axis to the listener's right, the sweet spot's frequency range increases but its width decreases for the left ear and increases for the right. The limiting case might be the smaller of the two widths and so the expectation might be that the sweet spot size decreases in width for off-axis listening.

Figure 9 shows results from the dummy HRTFs. This figure displays the same types of behaviour due to the ringing frequency and head shadowing. The results are obviously more complex in this case with the dip in the HRTFs at about 8kHz having an effect. The results above 20kHz reflect the error introduced by the fractional delay interpolation filter.

Designating at least 10dB cross-talk cancellation performance in the frequency range 0.3-3kHz as the criteria for the boundary of the sweet spot yields Fig. 10. This range of frequencies corresponds with the subjective experiment discussed in the next section. The authors feel the choice of 10dB cross-talk cancellation to be a reasonable estimate of the amount needed in order to ensure a desired subjective perception. In Fig. 10, the abscissas are the designed filter location relative to on-axis. The ordinates are the allowable head movements from the optimal position before there is less than 10dB cross-talk cancellation performance achieved. The solid lines give the sweet spot size for the range of head positions  $\pm 2$  m off-axis as defined by this criteria. The left and right columns are for the left and right ears respectively. Section IV discusses the subjective examination of the designed filter positions within the dashed lines of these figures.

The basis for Figs. 10a,b is the free field model. The sweet spot size for the ipsi-lateral ear (ear closer to the sources) is almost constant at about  $\pm 4$ cm -  $\pm 6$ cm from the optimal position. The sweet spot size for the contra-lateral ear (ear farthest from the sources) tolerates a little more movement away from the sources at farther off-axis positions. For example, at positions greater than 1m off-axis the contra-lateral ear has a sweet spot size of 10cm away from the sources and 5cm toward the sources.

The basis for Figs. 10c,d is the spherical head model. The effect of head shadowing greatly increases the robustness of cross-talk cancellation at the contra-lateral ear, especially at farther off-axis positions. The ipsi-lateral ear becomes less robust at farther off-axis positions, until it reaches about  $\pm 3$ cm at positions farther than 50cm off-axis.

The basis for Figs. 10e,f is the KEMAR dummy HRTFs. The sweet spot has the same general shape for this model as for the spherical head model. The measurements of the dummy with the inherent complexities of its shape introduce some randomness to the sweet spot and generally reduce its size by 1-2cm throughout.

The sweet spot's frequency range increases at off-axis listening positions because of a smaller path length difference at off-axis positions. Head shadowing counters this benefit by decreasing the sweet spot size for the ipsi-lateral ear (ear closer to the sources). However, head shadowing increases the sweet spot size for the contra-lateral ear (ear farthest from the sources) at off-axis listener locations. In practice, the expansion of the sweet spot's frequency range to include higher frequencies might not matter if the sound includes low frequencies (below 1.5kHz) due to the dominance of the low frequency interaural time difference cue for

sound localization. Within the dashed lines in Fig. 10, the sweet spot size is fairly constant and allows for head movements up to about 5cm to one side. This range of positions corresponds to the listener locations examined subjectively in section IV.

### E. Calculating the sweet spot boundary from just noticeable interaural time difference

The predominant horizontal localization cue at low frequencies (below 1.5kHz) is the interaural time difference (ITD) of sound arriving at the two ears<sup>1,22-24</sup>. There is convincing evidence that the brain extracts ITD by performing a running cross-correlation of the two-ear signals<sup>25-28</sup>. Computer models of the human auditory system generally first implement a band-pass filter bank to simulate the frequency selectivity of the cochlea and then perform an interaural cross-correlation (IACC) in each of the frequency bands<sup>27,29,30</sup>. The time lag at which the peaks in the IACC functions occur is then the estimate of ITD. Therefore, ITD is generally a function of critical frequency bandwidth. For the following simulations, a simple low-pass filtering of the two ear signals replaces the filter bank. A single cross-correlation calculation then follows. The reasoning behind this scheme is because humans utilize ITD at only low frequencies and ITD is fairly constant as a function of frequency. The cut-off frequency of the low pass filter is 4kHz. This has an effect of averaging the results for the critical frequency bandwidths below 4kHz.

A vector of synthesized HRTFs  $\mathbf{q}(z)$  in the  $z$  domain is defined as

$$\begin{bmatrix} Q_1(z) \\ Q_2(z) \end{bmatrix} = \mathbf{q}(z) = \mathbf{C}(z)\mathbf{X}(z)\mathbf{a}(z) = \mathbf{C}(z)\mathbf{h}(z) \quad (17)$$

The computation of these synthesized HRTFs  $\mathbf{q}(z)$  are for one set of designed virtual acoustic imaging filters  $\mathbf{h}(z)$  and initially with the plant matrix  $\mathbf{C}(z)$  corresponding to the optimal head location. Inverse fast Fourier transforms are performed on the two HRTFs yielding the discrete time domain synthesized head related impulse responses (HRIRs)  $q_1(n)$  and  $q_2(n)$ . These signals are passed through a low pass filter with a cut off frequency of 4kHz yielding  $q_{1L}(n)$  and  $q_{2L}(n)$ . These two signals are cross-correlated yielding an interaural cross-correlation function  $\Psi_{q_L}(m)$ ,

$$\Psi_{q_L}(m) = \sum_{n=0}^{N-m-1} q_{1L}(n)q_{2L}(n+m) \quad 0 \leq m < N \quad (18a)$$

$$\Psi_{q_L}(-m) = \sum_{n=0}^{N-m-1} q_{1L}(n+m)q_{2L}(n) \quad 0 \leq m < N \quad (18b)$$

The estimate of ITD is the amount of lag where the maximum value of this function occurs i.e.,

$$ITD = \left( m \Big|_{\max(\Psi_{q_L})} \right) \frac{1}{f_s}, \quad (19)$$

where  $f_s$  is the sampling frequency.

The initial calculation finds ITD with the listener at the optimal listener location. Simulations of incremental head displacements from the optimal listener location then follow with observations of the resulting change in ITD. The head position is changed by changing the plant matrix  $C(z)$  in equation (17).

Klumpp and Eady have suggested that the just noticeable ITD for humans is about  $10\mu\text{s}$ <sup>31</sup>. The following designations of sweet spot boundary correspond to head displacements that introduce this amount of ITD as found by the above method. The listener should hear no difference in the location of the virtual source with head displacements less than the size of the sweet spot as defined by this method.

Figures 11-13 show the virtual source locations and the sweet spot size as predicted by the three HRTF models. These figures include results for head positions through the range  $\pm 2\text{m}$  off-axis. Figure 11 shows the sweet spot size when only considering the ITD shift of a single virtual source  $45^\circ$  to the listener's right. Two striking peaks in the sweet spot size occur at 190cm and 100cm to the left of the loudspeakers. The peaks correspond to the listener moving to a position where the real sources are at the location of the virtual source ( $45^\circ$  to the listener's right). When considering more than one virtual source these peaks disappear as shown in Figs. 12,13. Adding more virtual sources into the calculations has the effect of decreasing the sweet spot size.

Generally, the free field model predicts a larger sweet spot for farther off-axis locations while the sphere and dummy models predict a constant sweet spot size. The free field model predicts a size of about  $\pm 4\text{cm}$  for locations close to on-axis and about  $\pm 10\text{cm}$  farther off-axis. The dummy predicts the smallest size sweet spot of about  $\pm 2\text{cm}$ . The sphere model predicts a size of about  $\pm 4\text{cm}$ . Section IV discusses the subjective examination of the designed filter positions within the dashed lines of Fig. 11. The free field model predicts a constant sweet spot size of about  $\pm 3.5\text{cm}$  at these listener locations. The shadowing effect of the sphere causes the size to vary within the dashed lines but the dummy again predicts a fairly constant size of about  $\pm 3\text{cm}$ .

#### IV. SUBJECTIVE EXPERIMENT

This section describes a subjective experiment undertaken in an anechoic chamber to investigate the sweet spot size at a range of head positions in a static case. The subjective results compare to the results of the calculated sweet spot size of the previous section.

##### A. Procedure

Band passed white noise with a pass-band of 300Hz to 3kHz was used as the source signal. This band of frequencies was chosen to include frequencies which humans localize predominately by utilizing ITD (below about  $1.5\text{kHz}$ <sup>1,22,24</sup>) and also to include the first ear canal resonance at about  $2\text{kHz}$ <sup>1</sup> but not include pinnae effects.

The subjects sat inside the 2m-diameter sphere shown in Fig. 14. A small headrest supported the subject's head and the subject was asked to limit their movement as much as possible. Sometimes a subject may have needed to move their head to read what the angle was from which they perceived the noise. A black curtain covered the sphere so the subject could not see out. The two loudspeakers were mounted on a moveable slide outside of the sphere, 1.4m away from the subject's head. The two loudspeakers subtend an angle of  $10^\circ$  when the listener is in the symmetric head position. The motion of the slide and thereby the loudspeakers was controlled via a computer in the anechoic chamber. The motion of the loudspeakers was limited to the horizontal direction parallel to a line joining the receiver's ears as shown in Fig. 15. When asked, the subjects voiced their perception of the direction of the noise. Responses were limited to the horizontal plane at angle locations marked inside of the sphere as illustrated in Fig. 15.

The virtual acoustic imaging filters were designed with the path responses approximated by monopole sources impinging sound on a perfectly rigid sphere. The experiment began by choosing one set of virtual acoustic imaging filters. This set of filters was then left unchanged while the loudspeakers were displaced incrementally to the left of the position for which the filters were designed. The farthest point examined was about 15cm away from the intended design listener location of the filters. The loudspeakers were then displaced incrementally to the right going back to the designed listener location. The loudspeakers were then displaced incrementally to the right of the designed listener location and then finally moved incrementally left going back to the designed listener location. After every increment, the subject was asked the azimuthal angle location where they perceive the location of the noise source. The loudspeaker increment distance was 1cm.

This procedure was repeated for several different sets of filters. The designed listener locations examined were when the inter-source axis exactly coincides with the inter-receiver axis and when it is offset 5cm, 10cm, 15cm, 20cm, and 25cm to the right. Figure 16 depicts these 6 positions with small circles denoting the centre of the two loudspeakers. There were 14 different subjects taking part in the experiment in all. Each subject listened to at least 2 different designed filter positions, although most of the subjects listened to 3 designed filter positions. On average then, each set of virtual acoustic imaging filters were examined by 7 different subjects. The virtual source was designed to be at  $45^\circ$  to the front right of the subject in all cases.

## **B. Results**

Figure 17 shows the subjects' responses for the 5cm off-axis designed virtual acoustic imaging filters. The results for this listener location give an example of the types of results obtained at all of the 6 positions examined. These figures show the difference in the angle perceived from the angle perceived when the listener is at the optimal location, plotted against the displacement of the loudspeakers from the designed, 5cm off-axis, optimal, listener location. The numbers 1-14 designate the different subjects and different symbols represent their individual responses in the plots. Solid lines in these graphs show the average responses. In addition, error bars show the standard deviation of the data. The standard deviation seems considerable at points but the means show clear trends. Generally, the standard deviation is greater for the larger loudspeaker displacements from optimal listener location. Front-back

confusions are resolved in the Fig. 17. That is, the plots show the listeners' responses of perceptions of the noise coming from behind them at the corresponding mirrored angles in front.

Figure 18 shows the means for all 6 of the examined designed filter positions. This plot shows the difference in the angle perceived, from the angle perceived when the listener is at the optimal location, against the displacement of the loudspeakers from the designed optimal listener location. At the optimal listener location, the subjects generally tended to perceive the direction to be slightly in front of the target location of  $45^\circ$  (e.g.  $30^\circ$ ). Other evaluations of virtual sound systems have also found this to occur<sup>28</sup>. Again, front-back confusions are resolved in the Fig. 18.

The top two plots, of Figs. 17,18, show the results with the loudspeakers moving incrementally away from the designed listener location. That is, the loudspeakers started out at the designed filter listener location and then were progressively displaced further and further away from the design listener location. The bottom two figures show the results with the loudspeakers moving incrementally toward the designed filter listener location. Comparisons of the bottom and top figures reveal discrepancies between responses given with the loudspeakers at the exact same locations. This exhibition of hysteresis suggests that the direction of the subject's last perception affects the subject's next direction perception. The subject was more likely to localize the sound from the same direction as their last response.

The left two plots, of Figs. 17,18, show the results with the loudspeakers left of the designed filter listener location while the right two graphs show the responses with the loudspeakers to the right of the designed listener location. Comparisons of the right and left figures suggest that the virtual sound image is more stable with the loudspeakers to the right. In fact, displacements in this direction correspond to the loudspeakers closer to the virtual source direction. This implies that the system is more robust for head movements away from the virtual source or loudspeaker movements toward the virtual source.

One might interpret these figures to imply a sweet spot size of about 3cm to one side of the designed filter listener location (6cm overall). The virtual image collapses somewhat gradually to the spot directly in front of the listener as the listener moves away from the optimum listening position. As a result, one must decide on how much error is tolerable in order to state the size of the sweet spot. A 3cm sweet spot boundary to one side of the optimal listener location corresponds with the results of the previous section. Both the ITD and cross-talk cancellation simulations predict close to the same size sweet spot making it difficult to define the limiting factor of the sweet spot size.

### **C. Limitations of results**

While the results of the experiment are useful and provide insight into the problem, this subsection discusses some observations made of the experimental procedure and possible limitations.

(1) The virtual acoustic imaging filters were designed with the path responses approximated by monopole sources impinging sound on a perfectly rigid sphere. Therefore, the

subjects' pinnae were not taken into consideration in the filter design. During the experiment, subjects often reported elevation movements. Many subjects localized the source behind them at times. These types of responses may have been avoided if the design of the filters had incorporated the pinna responses. Figs. 17,18 show the listeners' perceptions from behind them at their corresponding angles in the front.

(2) As was noted above, the subjects' responses show some bias toward the last direction that they perceived. An improvement to the experiment might be to present the different designed filters and loudspeaker displacements randomly. Although, this would make the experiment more time consuming.

(3) Toward the end of the experiment, some of the subjects expressed a feeling of fatigue in their right ear. This is because the intended virtual source location was invariably  $45^\circ$  to the subject's right. It would make for a more comfortable and interesting experiment by including some variation in the virtual source angle. This would complicate matters due to a dependence of sweet spot size on the number and locations of virtual sources as seen in the ITD simulations (section III).

## V. CONCLUSION

Virtual acoustic imaging systems with the geometrical arrangement shown in Fig. 1, are as robust to lateral head translations at asymmetric listener locations, that are offset up to 25cm from the inter-source axis, as the traditional symmetric on-axis listener location. A static subjective experiment suggests that virtual acoustic imaging filter adaptation should occur with lateral head translations of about 3cm from the designed optimum listener location. This value is in agreement with calculations made using a  $10\mu\text{s}$  just noticeable interaural time difference threshold and a 10dB cross-talk cancellation performance threshold. Farther off-axis listening has an advantage of a higher ringing frequency due to smaller path length differences. Head-shadowing effects decreasing the sweet spot size at low frequencies offset this advantage.

## REFERENCES

- <sup>1</sup>J. Blauert (1997), *Spatial Hearing: The Psychophysics of Human Sound Localization* (The MIT Press, Cambridge, MA).
- <sup>2</sup>F. L. Wightman and D. J. Kistler, "Headphone simulation of free-field listening. II: Psychophysical validation," *J. Acoust. Soc. Am.* **85**, 868-878 (1989).
- <sup>3</sup>J. Bauck and D. H. Cooper, "Generalized transaural stereo and applications," *J. Audio Eng. Soc.* **44**, 683-705 (1996).
- <sup>4</sup>W. G. Gardner, "3-D Audio Using Loudspeakers" (Ph.D. thesis, MIT Media Laboratory, Cambridge, MA, 1997).
- <sup>5</sup>O. Kirkeby, P. A. Nelson, and H. Hamada, "Local sound field reproduction using two closely spaced loudspeakers," *J. Acoust. Soc. Am.* **104**, 1973-1981 (1998).
- <sup>6</sup>B. S. Atal and M. R. Schroeder, "Apparent sound source translator," U.S. Patent 3,236,949 (1962).
- <sup>7</sup>T. Takeuchi and P. A. Nelson, "Robustness to head misalignment of virtual sound imaging systems," *J. Acoust. Soc. Am.* **109**, 958-971 (2001).
- <sup>8</sup>O. Kirkeby, P. A. Nelson, and H. Hamada, "Stereo dipole," Patent Application, PCT/GB97/00415, 1997.
- <sup>9</sup>P. A. Nelson, O. Kirkeby, T. Takeuchi, and H. Hamada, "Sound fields for the production of virtual acoustic images," *J. Sound. Vib. (Lett.)* **204**, 386-396 (1997).
- <sup>10</sup>O. Kirkeby, P. A. Nelson, and H. Hamada, "The "stereo dipole" a virtual source imaging system using two closely spaced loudspeakers," *J. Audio Eng. Soc.* **46**, 387-395 (1998).
- <sup>11</sup>J. F. W. Rose, "A Visually Adaptive Virtual Sound Imaging System" (MSc. thesis, Institute of Sound and Vibration Research, Southampton, UK, 1999).
- <sup>12</sup>P. A. Nelson, F. Orduna-Bustamante, and D. Engler, "Experiments on a system for the synthesis of virtual acoustic sources," *J. Audio Eng. Soc.* **44**, 990-1007 (1996).
- <sup>13</sup>F. L. Wightman and D. J. Kistler, "Headphone simulation of free-field listening. I: Stimulus synthesis," *J. Acoust. Soc. Am.* **85**, 858-867 (1989).
- <sup>14</sup>P. A. Nelson and S. J. Elliott (1992), *Active Control of Sound* (Academic Press, London).
- <sup>15</sup>F. A. Firestone, "The phase difference and amplitude ratio at the ears due to a source of pure tone," *J. Acoust. Soc. Am.* **2**, 260-270 (1930).
- <sup>16</sup>W. E. Feddersen, T. T. Sandel, D. C. Teas, and L. A. Jeffress, "Localization of high-frequency tones," *J. Acoust. Soc. Am.* **29**, 988-991 (1957).

- <sup>17</sup>M. D. Burkhard and R. M. Sachs, "Anthropometric manikin for acoustic research," *J. Acoust. Soc. Am.* **58**, 214-222 (1975).
- <sup>18</sup>W. G. Gardner and K. D. Martin, "HRTF measurements of a KEMAR," *J. Acoust. Soc. Am.* **97**, 3907-3908 (1995).
- <sup>19</sup>W. G. Gardner and K. D. Martin, *KEMAR HRTF Measurements*, (MIT's Media Lab through <http://sound.media.mit.edu/KEMAR.html>, 1994).
- <sup>20</sup>A. D. Pierce, (1989), *Acoustics: An Introduction to Its Physical Principles and Applications* (Acoustical Society of America, NY).
- <sup>21</sup>T. I. Laakso, V. Valimaki, M. Karjalainen, and U. K. Laine, "Splitting the unit delay: Tools for fractional delay filter design," *IEEE Signal Process. Mag.* **13**(1), 30-60 (1996).
- <sup>22</sup>J. C. Middlebrooks and D. M. Green, "Sound localization by human listeners," *Ann. Rev. Psychol.* **42**, 135-159 (1991).
- <sup>23</sup>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).
- <sup>24</sup>W. M. Hartmann, "How we localize sound," *Phys. Today* **52**(11), 24-29 (1999).
- <sup>25</sup>J. C. R. Licklider, "Auditory frequency analysis," in *Information Theory: Third London Symposium* edited by C. Cherry (Butterworths Scientific Publications, London, 1956), pp. 253-268
- <sup>26</sup>B. McA. Sayers and E. C. Cherry, "Mechanism of binaural fusion in the hearing of speech," *J. Acoust. Soc. Am.* **29**, 973-987 (1957).
- <sup>27</sup>J. C. R. Licklider, "Three Auditory Theories," in *Psychology: A Study of a Science, Study I. Conceptual and Systematic, Volume 1. Sensory, Perceptual, and Physiological Formulations* edited by S. Koch (McGraw-Hill Book Company, Inc., New York, 1959), pp. 41-144
- <sup>28</sup>B. McA. Sayers, "Acoustic-image lateralization judgments with binaural tones," *J. Acoust. Soc. Am.* **36**, 923-926 (1964).
- <sup>29</sup>W. Lindemann, "Extension of a binaural cross-correlation model by contralateral inhibition. I. Simulation of lateralization for stationary signals," *J. Acoust. Soc. Am.* **80**, 1608-1622 (1986).
- <sup>30</sup>W. Gaik, "Combined evaluation of interaural time and intensity differences: Psychoacoustic results and computer modeling," *J. Acoust. Soc. Am.* **94**, 98-110 (1993).
- <sup>31</sup>R. G. Klumpp and H. R. Eady, "Some measurements of interaural time difference thresholds," *J. Acoust. Soc. Am.* **28**, 859-860 (1956).

FIGURES

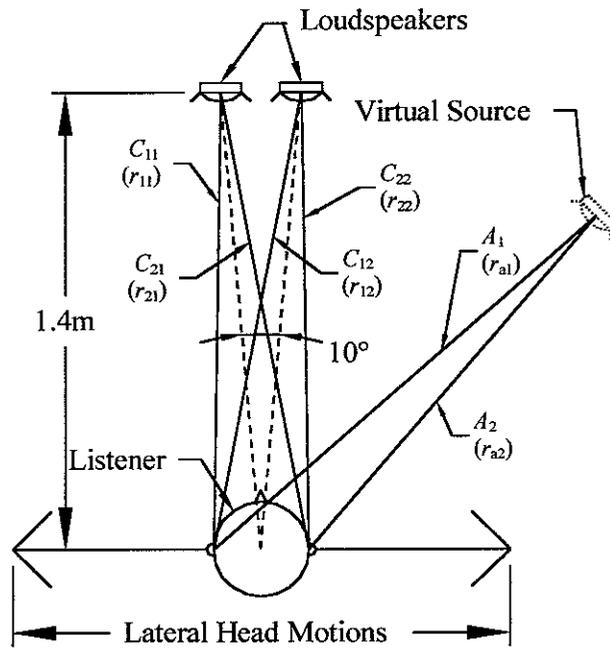


Figure 1. Plan view depicting lateral head translations and the “stereo dipole” geometrical arrangement. The elements of the electroacoustic path plant matrix  $C$  and the virtual paths  $a$  are shown as well as the corresponding distances used in equations (5,6).

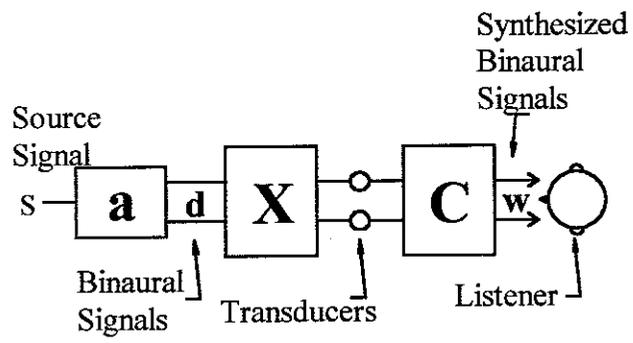


Figure 2. Block diagram of the implemented virtual acoustic imaging system.

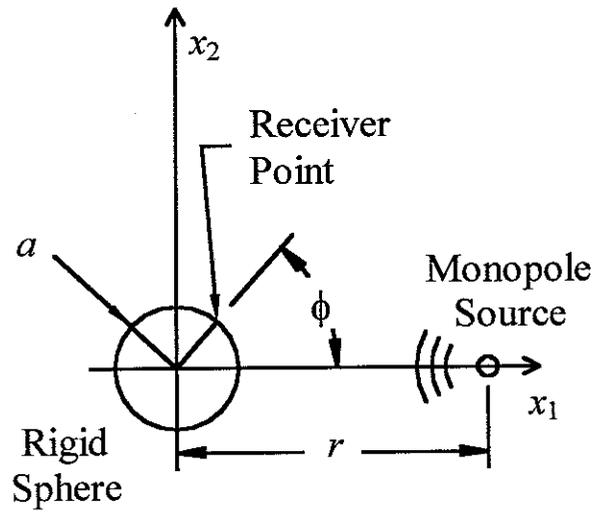


Figure 3. Variables used when calculating the sound field scattered from a rigid sphere.

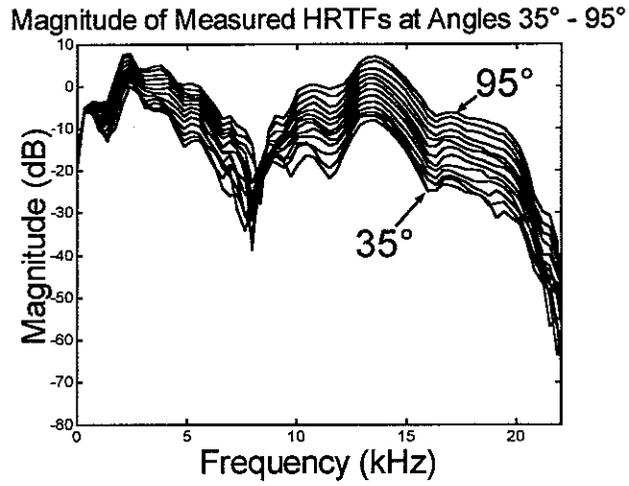


Figure 4. Magnitudes of KEMAR's left ear HRTFs at angles used for the simulations. For angles in between the curves shown, linear interpolation was performed. This appears to be a reasonable approximation with a possible exception around the dip at about 8kHz.

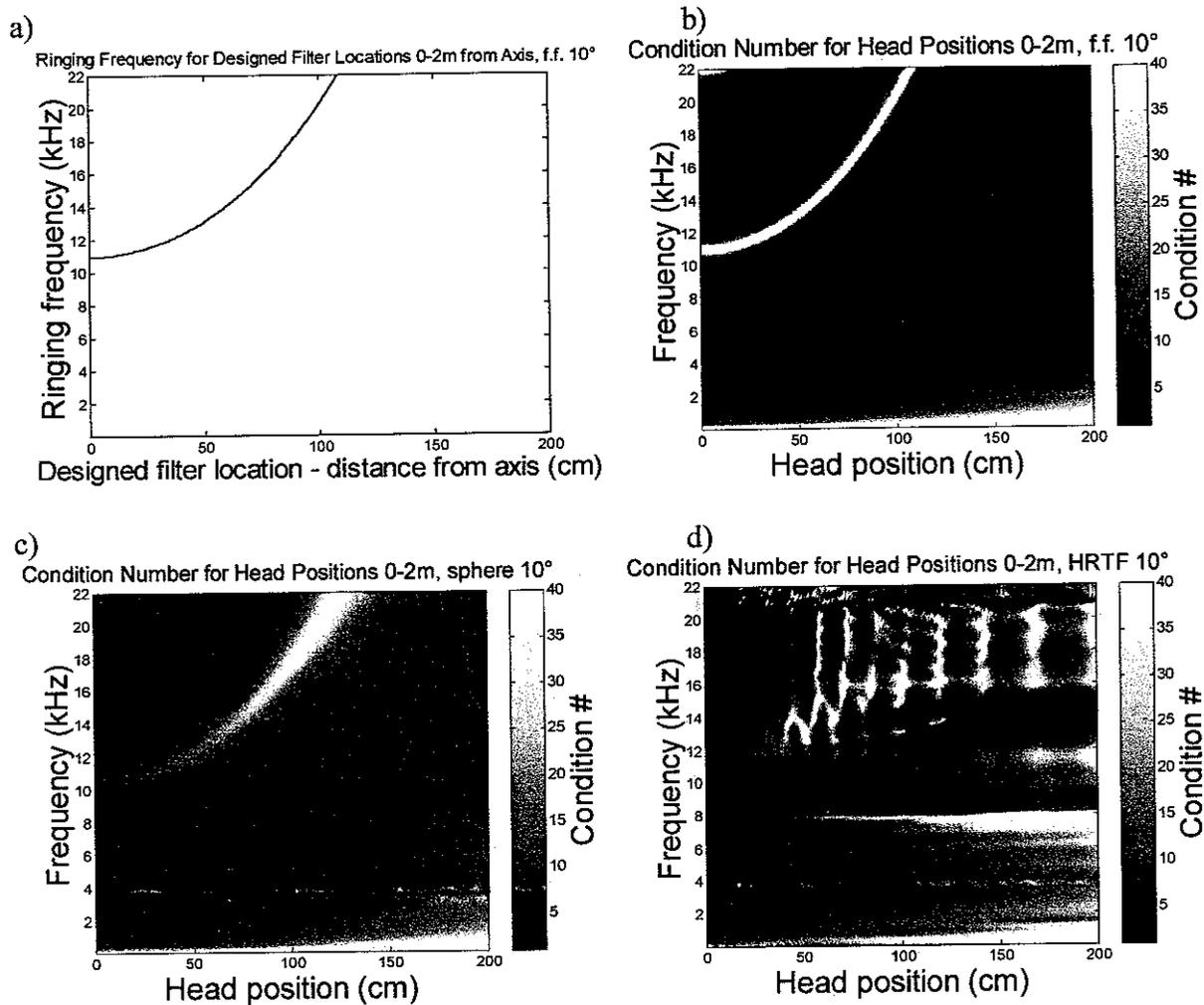


Figure 5. Calculated results for the 10° loudspeaker arrangement at head positions 0-2m off-axis a) free field ringing frequency, b) free field condition number (maximum values of 280,000 at 1Hz and 1,100 at 22kHz), c) rigid sphere model condition number (maximum values of 13,000 at 1Hz and 74 at 22kHz), and d) KEMAR dummy condition number (maximum values of 50,000 at 1Hz, 2,900 at 8kHz, and 1,200 at 16kHz).

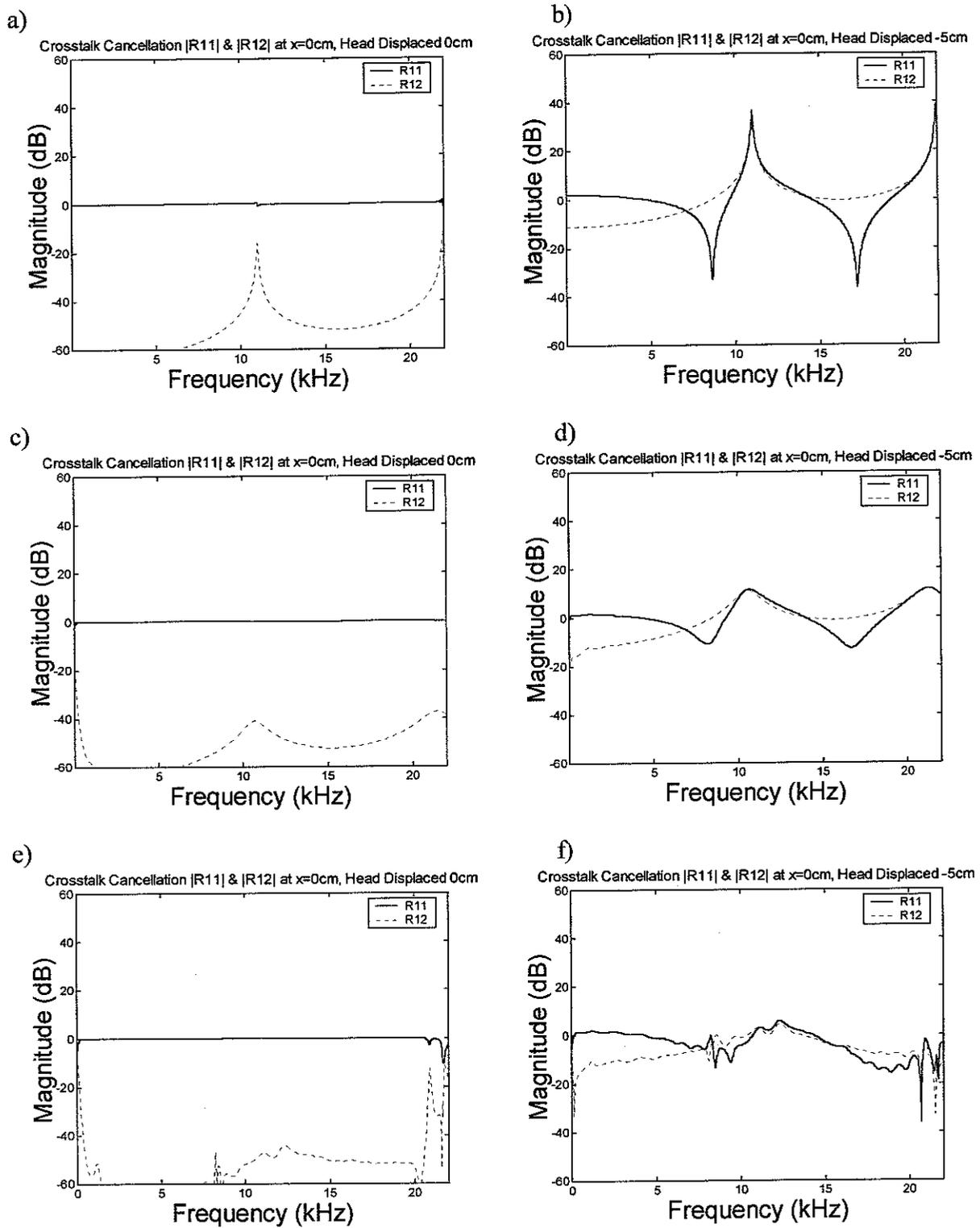


Figure 6. Elements  $|R_{11}|$  and  $|R_{12}|$  of control performance matrix  $\mathbf{R}$  for filters designed at the on-axis head location. The left column shows results when the head is displaced 0.1mm from optimum. The right column shows results when the head is displaced 5cm from optimum. The calculations are based on a-b) the free field approximation, c-d) the spherical head model, and e-f) KEMAR dummy HRTFs.

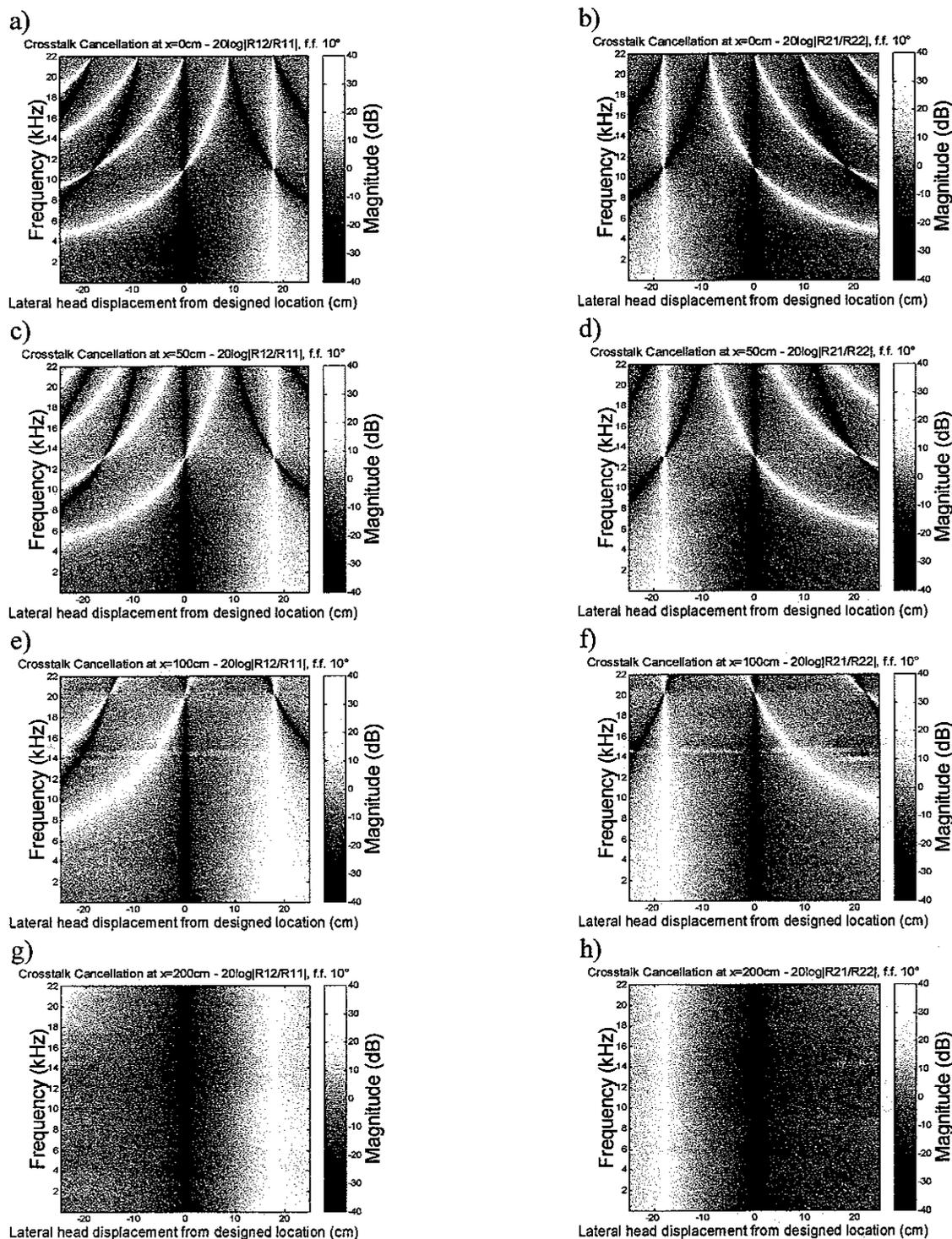


Figure 7. Cross-talk cancellation effectiveness for the left ear (left column) and right ear (right column) at designed virtual acoustic imaging filter listener locations a-b) on-axis, c-d) 50cm off-axis, e-f) 1m off-axis, and g-h) 2m off-axis. In these figures, the abscissa represents the head position relative to the assumed designed optimum listener location, the ordinate represents frequency and the grey scale represents the ratio of the cross-talk path to the direct path in decibels. Good cross-talk cancellation performance is achieved in the dark areas. All of the calculations are based on the free field model.

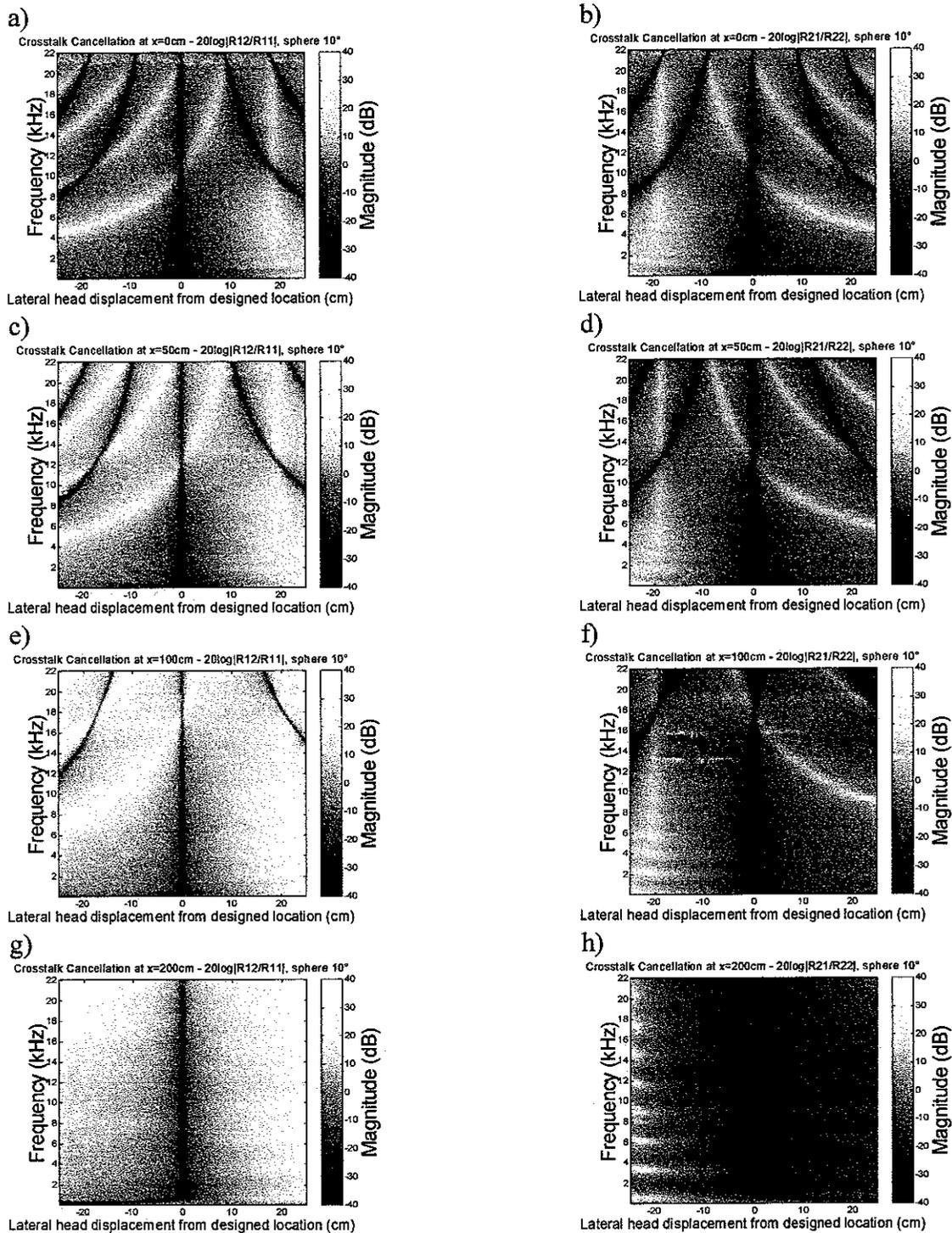


Figure 8. Cross-talk cancellation effectiveness for the left ear (left column) and right ear (right column) at designed virtual acoustic imaging filter listener locations a-b) on-axis, c-d) 50cm off-axis, e-f) 1m off-axis, and g-h) 2m off-axis. In these figures, the abscissa represents the head position relative to assumed designed optimum listener location, the ordinate represents frequency and the grey scale represents the ratio of the cross-talk path to the direct path in decibels. Good cross-talk cancellation performance is achieved in the dark areas. All of the calculations are based on the spherical head model.

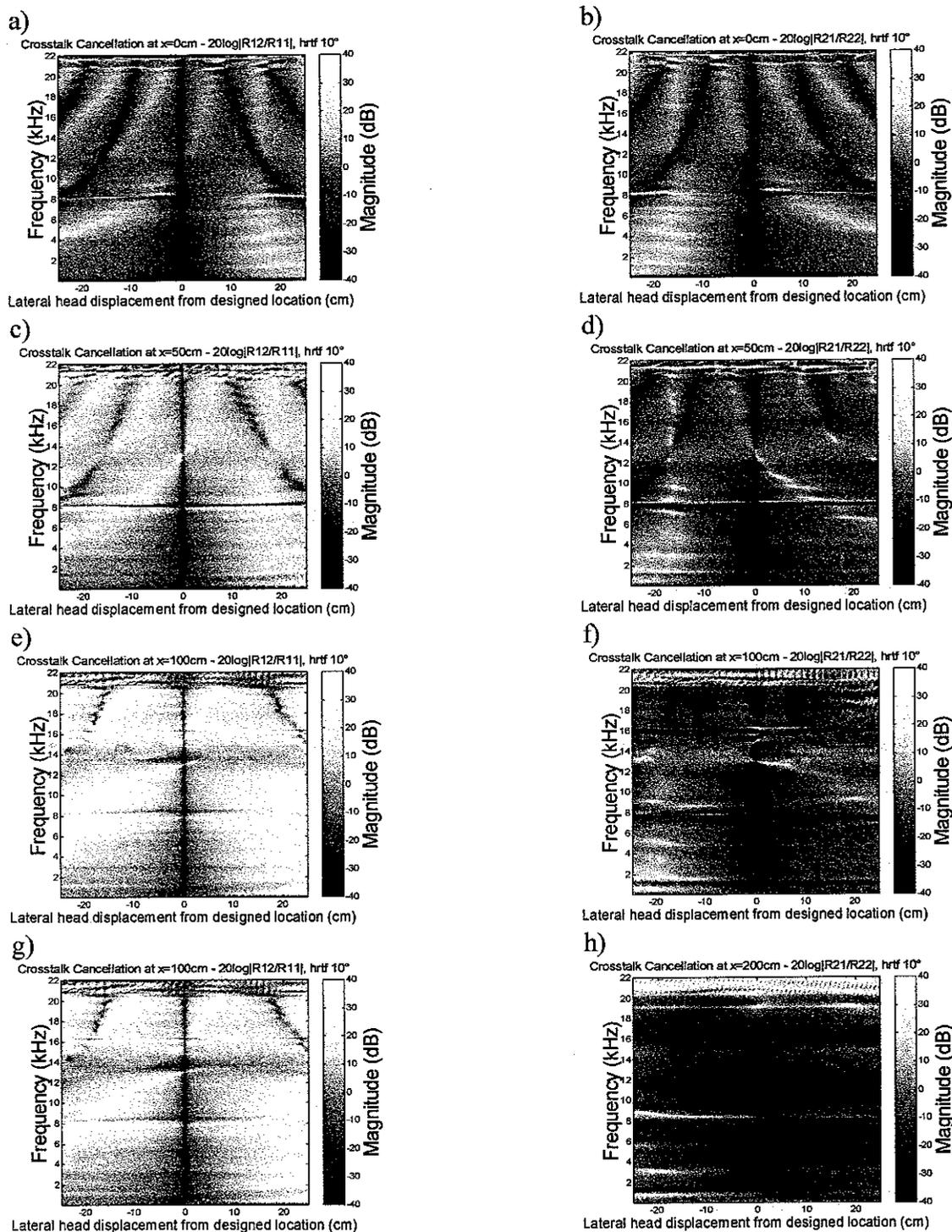


Figure 9. Cross-talk cancellation effectiveness for the left ear (left column) and right ear (right column) at designed virtual acoustic imaging filter listener locations a-b) on-axis, c-d) 50cm off-axis, e-f) 1m off-axis, and g-h) 2m off-axis. In these figures, the abscissa represents the head position relative to assumed designed optimum listener location, the ordinate represents frequency and the grey scale represents the ratio of the cross-talk path to the direct path in decibels. Good cross-talk cancellation performance is achieved in the dark areas. All of the calculations are based on measured KEMAR dummy HRTFs.

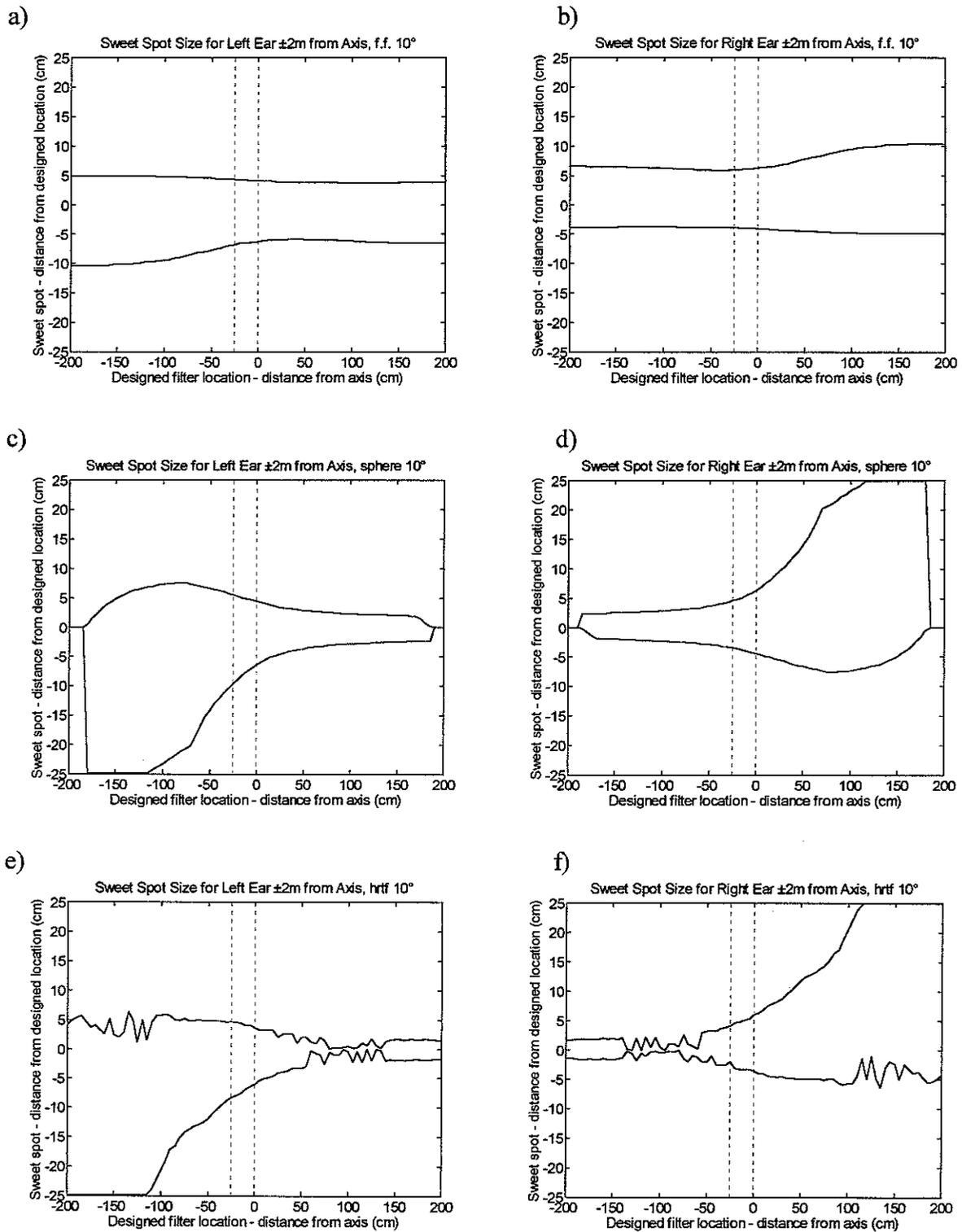


Figure 10. Boundaries of the sweet spot for a range of head locations  $\pm 2\text{m}$  off-axis as calculated from a 10dB cross-talk cancellation performance criterion in the frequency range 0.3-3kHz when using a) b) the free field model, c) d) spherical head model, and e) f) the KEMAR dummy. The left and right columns are for the listener's left and right ears respectively. Dashed lines show the positions examined subjectively in section IV.

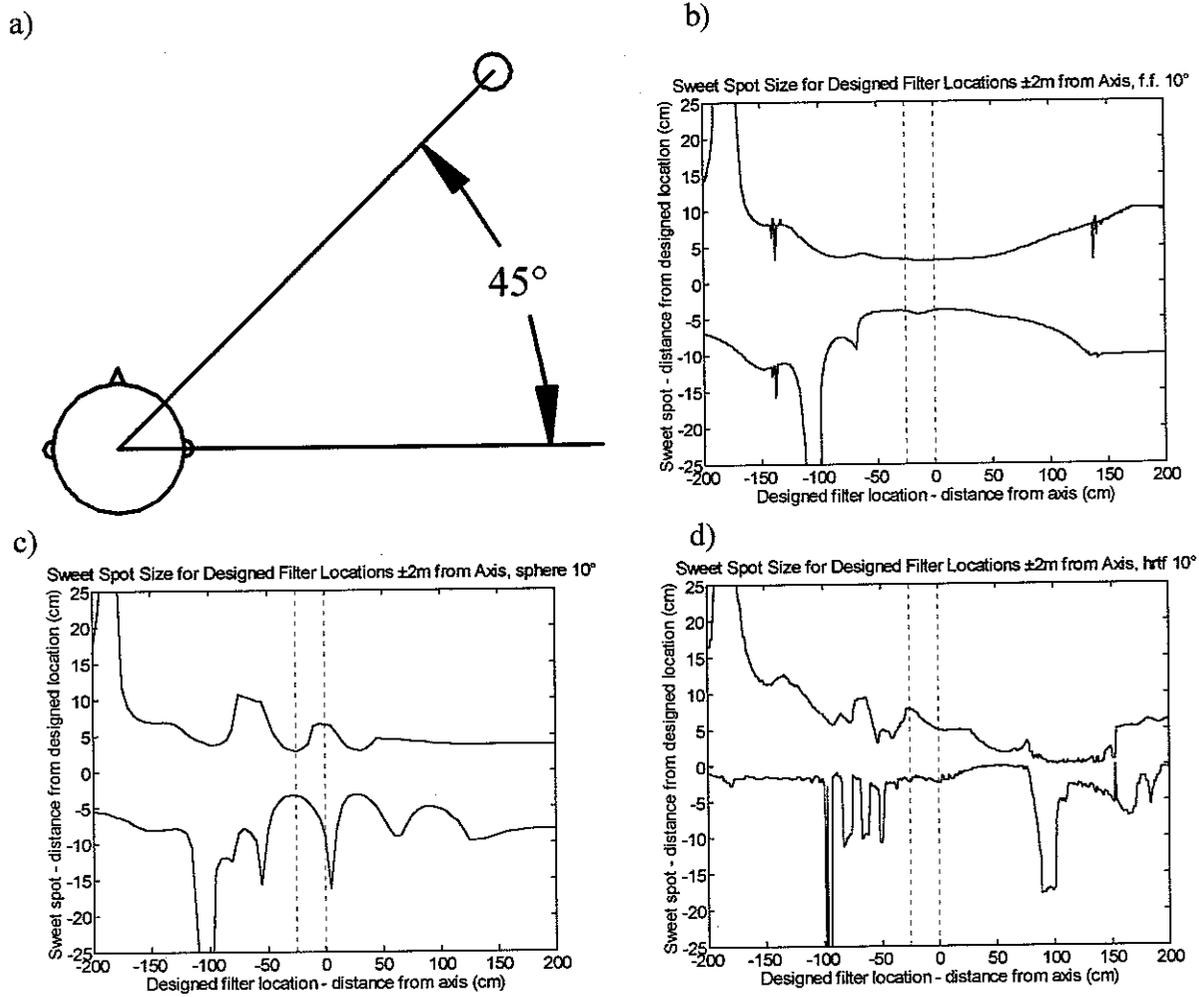


Figure 11. Boundaries of the sweet spot for a range of head locations  $\pm 2\text{m}$  off-axis as calculated from a  $10\mu\text{s}$  shift in ITD of the single virtual source represented by the small circle in a) when using b) the free field model, c) spherical head model, and d) the KEMAR dummy. Dashed lines show the positions examined subjectively in section IV.

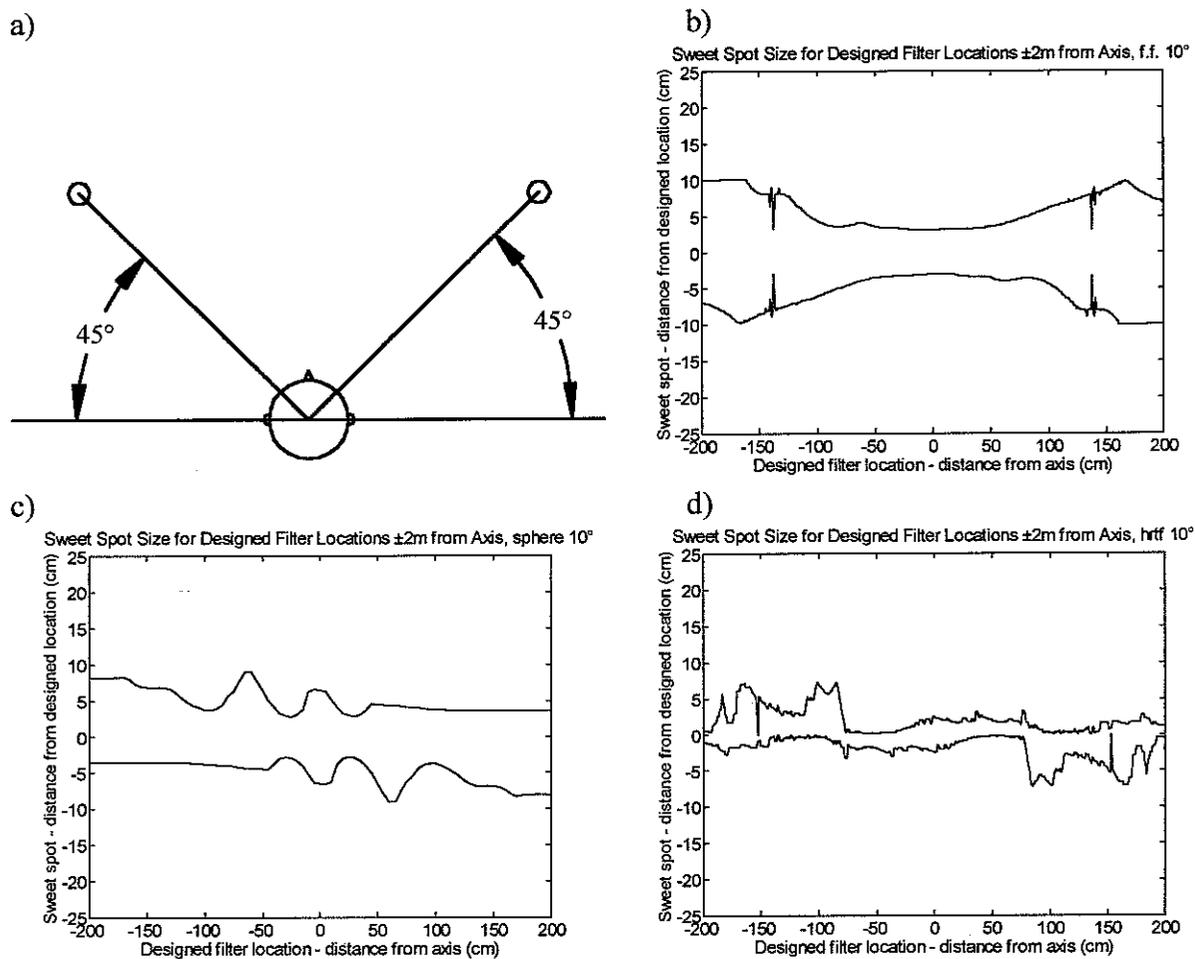


Figure 12. Boundaries of the sweet spot for a range of head locations  $\pm 2m$  off-axis as calculated from a  $10\mu s$  shift in ITD of either of the two virtual sources represented by the small circles in a) when using b) the free field model, c) spherical head model, and d) the KEMAR dummy.

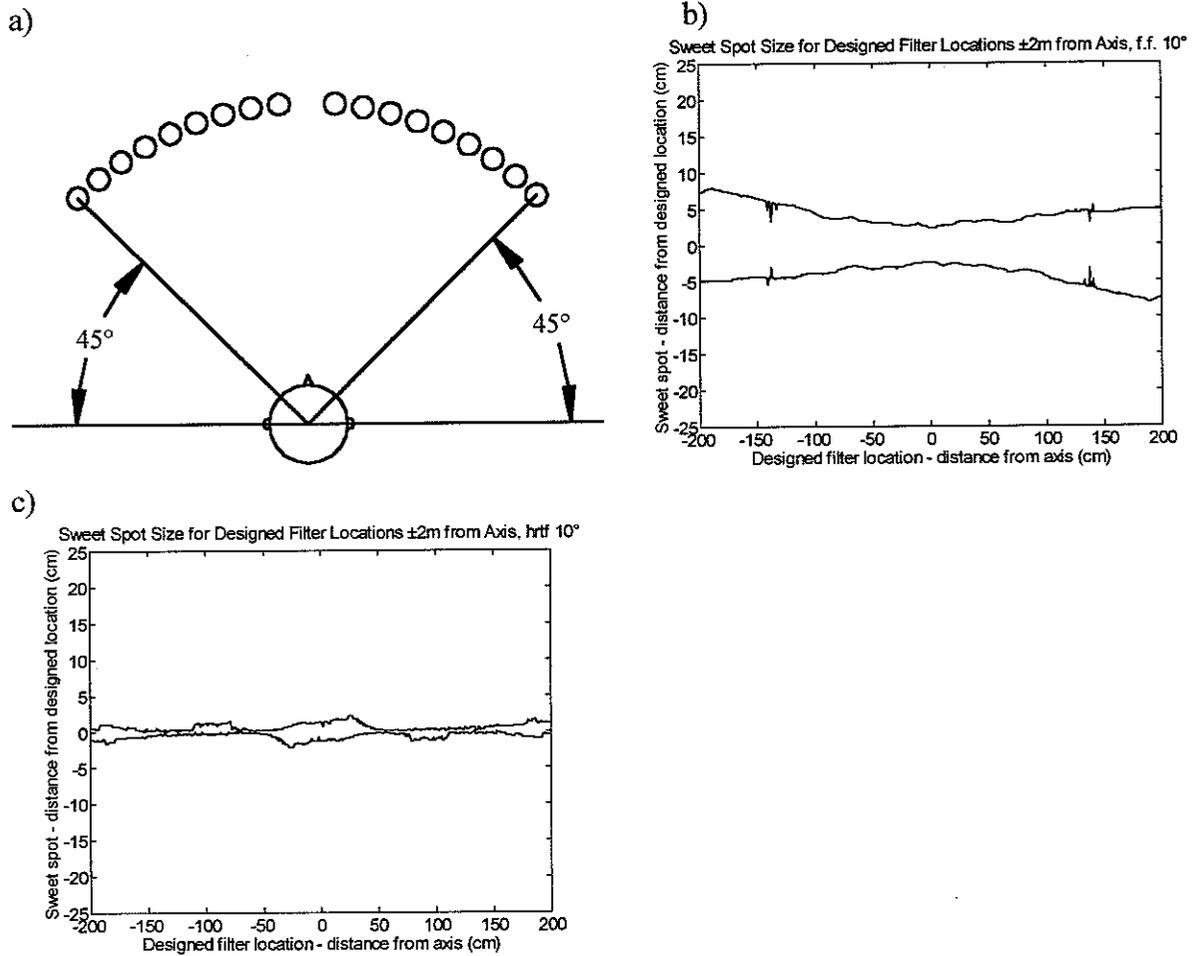


Figure 13. Boundaries of the sweet spot for a range of head locations  $\pm 2\text{m}$  off-axis as calculated from a  $10\mu\text{s}$  shift in ITD of any of the 18 virtual sources represented by the small circles in a) when using b) the free field model and c) the KEMAR dummy.

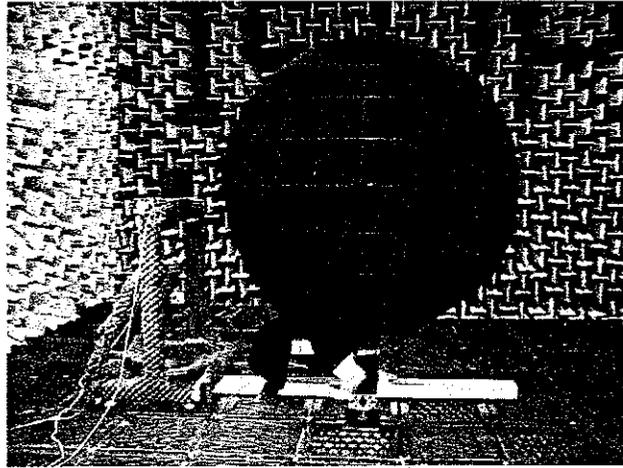


Figure 14. Set-up for subjective experiment inside anechoic chamber.

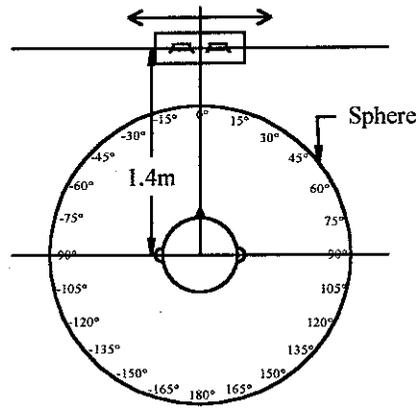


Figure 15. Plan view of the experimental arrangement with the subject inside the sphere, angles marked, and loudspeakers on a moveable slide.

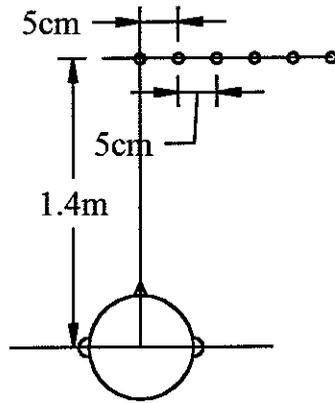


Figure 16. Plan view depicting the 6 designed virtual acoustic imaging filter positions that were examined subjectively. The small circles represent the point at the middle of the two loudspeakers for the 6 positions.

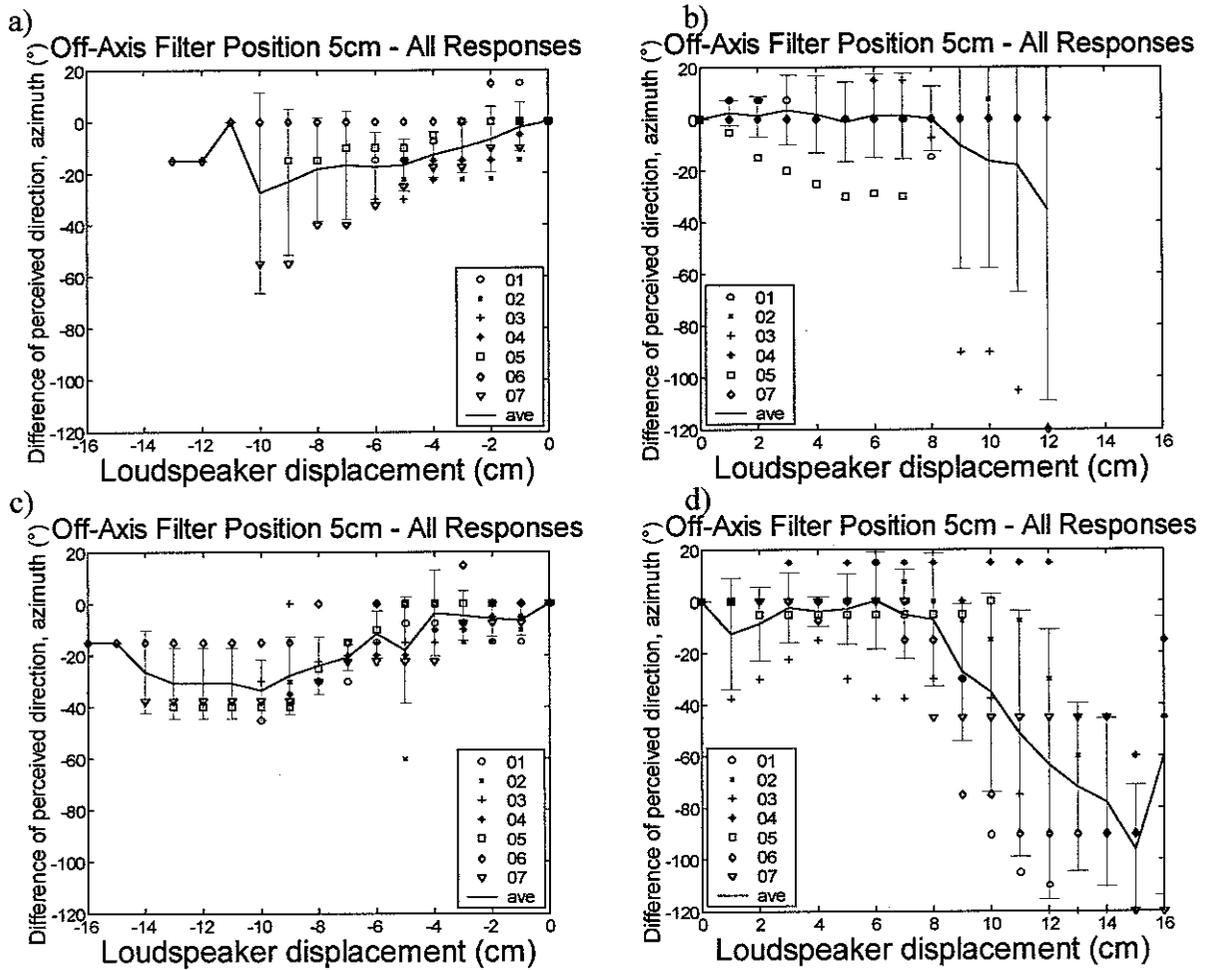


Figure 17. Subjects' responses for the 5cm off-axis designed virtual acoustic imaging filter position. Different symbols denote different subjects. Responses are shown with a) loudspeakers moving away from designed filter location, to the left, b) loudspeakers moving away from designed filter location, to the right, c) loudspeakers moving toward the designed filter location, from the left, and d) loudspeakers moving toward the designed filter location, from the right.

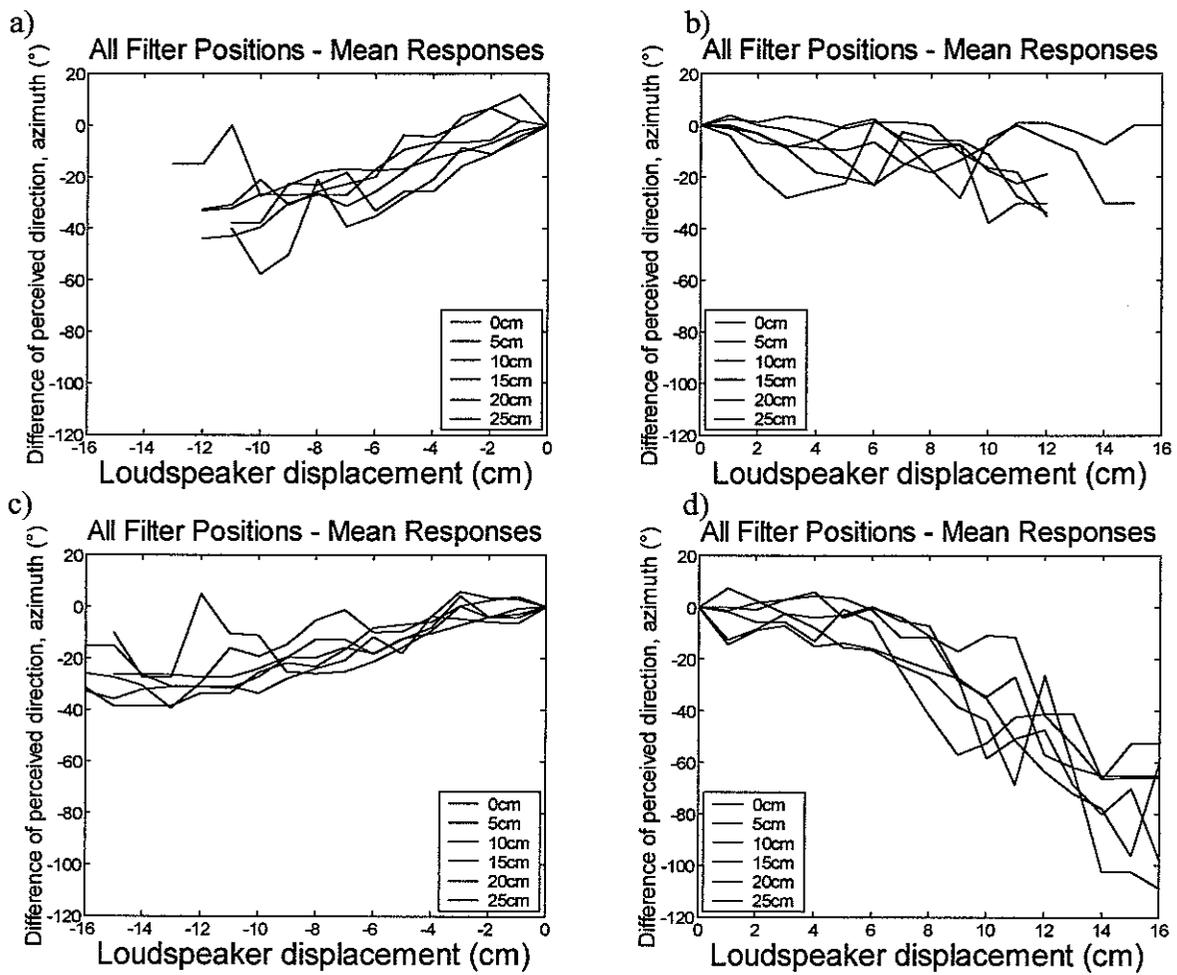


Figure 18. Average subject responses for the designed virtual acoustic imaging filter positions 0-25cm off-axis. Different lines represent different designed filter positions. Responses are shown with a) loudspeakers moving away from designed filter location, to the left, b) loudspeakers moving away from designed filter location, to the right, c) loudspeakers moving toward the designed filter location, from the left, and d) loudspeakers moving toward the designed filter location, from the right.