Ewan A. Macpherson
University of Waterloo
Waterloo, Ontario, Canada

Presented at the 87th Convention 1989 October 18-21 New York

AES

AN AUDIO ENGINEERING SOCIETY PREPRINT

A COMPUTER MODEL OF BINAURAL LOCALIZATION FOR STEREO IMAGING MEASUREMENT

by

Ewan A. Macpherson*
Audio Research Group
University of Waterloo
Waterloo, Ontario
N2L 3G1 Canada

Abstract

A binaural localization model has been developed for measuring the stereo imaging properties of recording and playback techniques. Ear signals from a dummy head are passed through a model of the inner ear, and inter-aural time and amplitude differences are extracted and processed to give image location and extent. Experiments show that the model emulates human performance under anechoic and reverberant conditions.

0. Introduction

Stereo reproduction of audio material produces images of sound sources between the two loudspeakers by creating, at the listener's ears, signals which attempt to simulate the inter-aural time and amplitude differences characterizing a real source at the desired position. Recording and playback arrangements vary in their ability to reproduce these binaural localization cues, and since there is usually no reference to which the ear signals can be compared, the resultant imaging must be evaluated by actual listening tests rather than by direct physical measurements. A system combining a dummy head with a computer model of human binaural localization has been developed to act as an objective "listener" for such tests.

The model takes as input recorded ear signals from a dummy head, and passes them through a model of the inner ear to obtain signals equivalent to the auditory nerve activity which is available to the human auditory system for binaural analysis. These are then processed in an appropriate manner to provide estimates of the source azimuth, which are combined in a weighted

*) The author is a student in the Guelph-Waterloo Program for Graduate Work in Physics.

average to provide a final localization judgement. An implementation of the precedence effect allows reflections and inter-channel delays to correctly influence the result. It must be noted that the approach taken in constructing this model is based on a binaural localization model described by Pocock in [1] and [2]. A number of alterations and extensions have been made to produce a model with improved performance and greater utility as a measurement tool.

Since the localization is based entirely on binaural differences, the model is restricted to localizing frontal sources in the horizontal plane - the most commonly encountered situation in stereophonic listening. It is also restricted to dealing with signals which are either impulsive or continuous (such as noise or periodic signals) since the localization of sounds composed of both is much more complex [3]. It also cannot perform other tasks of the auditory system such as separation of multiple sound sources, identification of the type of source signal, or detection of signals in background noise, which would require unavailable models of functions of the auditory system which are not well understood.

This paper describes the processing performed to derive azimuth estimates from the recorded ear signals, and presents the results of experiments which were done to verify the correct performance of the model and to examine its use in several imaging measurement applications.

1. The Basis of Auditory Localization

Auditory localization is possible because there are characteristics of the signals at the listener's ears which vary with the location of the source relative to the position and orientation of the head. For sources in the median plane, the signals at the left and right ears are identical (assuming a symmetrical head), and the source position is encoded in spectral alterations of the signals caused by diffraction at the head and especially the pinnae. While this requires a priori knowledge of the source spectrum in order to be an effective cue, localization of sources in the frontal horizontal half-plane does not suffer from such a restriction. This is due to the fact that in this case the source azimuth is encoded in differences between the left and right ear signals, and these binaural differences are independent of the original source signal.

The differences are of two kinds, the first of which is the inter-aural time difference (ITD). If a source is displaced to one side of the median plane, it is clear that the signal at the ear on that side will be advanced in time relative to that at the other ear. ITD varies monotonically with azimuth for sources between +90 (full left) and -90 degrees (full right), and also has a weak frequency dependence, which is due to phase velocity effects caused by diffraction around the head. Fig. 1, which displays ITD vs. azimuth curves for two frequency bands based on measurements of a Knowles Electronics manikin (KEMAR), shows these features clearly. (The means by which these ITD values were

obtained is explained in Section 2 of this paper.) The ITD provides unambiguous localization information for a sinusoidal signal only if it is smaller than half of one period. Since for most heads the maximum ITD lies between 600 and 700 μs, the first ambiguity occurs at approximately 800 Hz and appears at smaller azimuths with increasing frequency. The auditory system's sensitivity to ITDs in pure tones vanishes between 1400 and 1600 Hz, due to the time constants of the neural transduction in the inner ear.

The second type of binaural localization cue is the inter-aural amplitude difference (IAD); displacement of a source to one side of the median plane will result in a signal at the nearer ear with an amplitude higher than that at the other at most frequencies, due to the combined shadowing and baffling effect of the head. Fig. 2(a) shows the source-to-ear frequency responses for the left and right ears of a KEMAR manikin for a source at an azimuth of +30 degrees, and Fig. 2(b) the ratio of these transfer functions. As can be seen, the IAD increases with frequency from below 5 dB for frequencies below approximately 400 Hz, to 10 dB and more above 4 kHz. This is due to the increased head shadow/baffle effect at higher frequencies, and means that IAD cues are unreliable for such low-frequency signals. IADs also generally increase with source azimuth although, as can be seen in Fig. 2(c) (again measured on a KEMAR), the IAD vs. azimuth curves can be quite non-monotonic for higher frequencies due to diffraction effects.

The unreliability of ITD cues at high frequencies and of IAD cues at low frequencies led to the duplex theory of localization, which states that high frequencies are localized solely on the basis of IADs and low ones on ITDs [4]. However, it has been recognized for some time that complex high frequency signals are localized as easily with ITD cues as are low frequency sinusoids. The conclusion is that "... while the theory is true enough for [pure tones], it is almost totally irrelevant, for it cannot be appropriately generalized to real-world situations involving complex waveforms" [5]. The binaural localization theory assumed for the construction of this model is therefore: ITD measurements (based on phase or envelope) are made in all frequency bands and, if reliable, have equal potency as localization cues at all frequencies; IAD measurements are also made in all bands, but have a potency which increases with frequency; the inter-aural difference cues are transformed to azimuth estimates by comparing them to pre-learned "maps" of ITD and IAD vs. azimuth.

A final aspect of localization to be discussed involves listening under reverberant conditions. In most listening environments, the direct signal from a sound source is followed by numerous reflections (from floor, ceiling, and walls, for example), which reach the listener from directions quite different from that of the direct sound. The auditory system's ability to base localization judgements primarily on the direct sound and to suppress reflections is known as the precedence effect. These delayed signals do have an audible effect on the sound, but not on its perceived location. The precedence effect improves localization accuracy in reverberant environments, and also accounts for some phenomena of stereo reproduction involving time

delays between the two signal channels. For this reason, an implementation of the precedence effect, described in the following section, is considered to be an important aspect of the model.

2. The Binaural Localization Model

As mentioned previously, the binaural localization model involves simulation of the signal processing which takes place in the peripheral auditory system, and analyses the resulting signals to extract localization cues. This approach was considered to be appropriate since the non-linear aspects of the mechanical-to-neural transduction, which takes place in the inner ear, can result in localization phenomena which would be very difficult to model in a "black box" sense. It was also desired, wherever possible, not to give the model access to information unavailable to the human auditory system in order to avoid the possibility of unrealistic localization ability. Space does not permit a review of the physiology of the peripheral auditory system or of localization psychoacoustics, but for further information on the physiology and psychology of hearing, the books by Blauert [6], Gelfand [7], Moore [8], and Pickles [9] are recommended. In this section the processing steps outlined in the flowchart shown in Fig. 3 are explained, and the steps involved in forming a localization judgement of a particular source are used as an illustrative example.

Data Acquisition:

The first step in making a localization or imaging measurement is to measure the signals produced at the ears of the dummy head by the source in the environment in question. In order to do this, the KEMAR manikin previously mentioned was fitted with Bruel & Kjaer 4134, 1/2 inch, pressure microphones located at the entrance to the ear canals. It was decided to use a maximum-length sequence (MLS) technique [10] to measure the source-to-ear impulse responses (IRs) because of its good signal-to-noise ratio and the flexibility afforded by the availability of the IRs. If a localization of impulsive sounds is desired, the IRs can be used directly as the input to the model, and if some other source signal is desired, the ear signals can be derived by convolution with the previously measured IRs. Pseudo-anechoic measurements can also be made by editing the IRs to remove reflections. Fig. 4 shows the IRs produced by a source located 30 degrees to the left (at a sampling rate of 32 kHz). Note the earlier arrival and higher level of the left-ear signal. The processing of these ear signals is used as an example throughout the remainder of this section.

Basilar Membrane Filtering:

The frequency-selectivity of the auditory system is due in part to the mechanics of the cochlea, which cause each point along the basilar membrane to respond maximally at a different frequency, resulting in a frequency-place mapping. The motion of

points on the basilar membrane in response to the signal measured at the eardrum is generated by passing the ear signals through appropriate filters. A set of sixteen filters is used, with critical (peak response) frequencies ranging from 400 Hz to 12.8 kHz in 1/3rd octave steps. The filters were generated by using a simple expression for impulse responses which approximate those of filters which roll off at 60 dB/oct above the critical frequency, and 24 dB/oct below it [11]. The resulting responses of some of these filters, which approximate measured basilar membrane response curves, are shown in Fig. 5(a). Note that the middle ear is ignored and the eardrum signals supplied directly to the inner-ear model. This is justified by the fact that the middle ear is extremely linear and its gain control response time is much longer than the localization integration time for impulsive signals and cannot affect the results for steady periodic or noise signals. Some examples of basilar membrane responses in high and low frequency bands and to impulses and periodic signals are shown in Figs. 5 (b), (c), (d), and (e).

Hair Cell Response:

The transverse motion of the basilar membrane is converted into nerve impulses by the hair cells of the organ of Corti. Bending of their cilia in one direction excites the cells and increases the probability of firing, while bending in the other direction is inhibitory and hence the probability of firing is reduced. Thus the hair cell model input is a "soft" half-wave rectified version of the basilar membrane response, and since the firing probability is increased by rarefactions, it is the positive-polarity portions of the signal which are clipped. The output is a signal which is equivalent to the average rate of firing for a large number of cells rather than the output of a single cell, which would be a series of irregularly spaced spikes related stochastically to the excitation; modelling each of the 30,000 hair cells would clearly be impractical.

The hair-cell action is simulated by an electrical circuit model described by Pocock [2] and similar to that suggested in Schroeder [12]. The model used is shown in Fig. 6(a). There is synchrony between basilar membrane motion and hair cell firing for frequencies up to 5 kHz, with the response shifting from phase-following to envelope-following as frequency increases. However, in order to model the observed lack of sensitivity to inter-aural phase differences for frequencies above 1.5 kHz, the hair cell firing rate functions for critical frequencies higher than this are smoothed to remove any phase-locking. The hair-cell model responses to the basilar membrane signals of Fig. 5 are shown in Fig. 6. Note the phase-locking in the low-frequency band evident in Figs. 6(b) and 6(d), and the envelope-following in the high-frequency band in Figs. 6(c) and 6(e).

Critical Band Filtering:

The mechanical filtering action of the basilar membrane is not sufficiently sharp to account for the ability of the auditory system to discriminate between signals in different frequency

bands (additional neural processing must be involved), and no general model of IAD determination from neural firing rate signals was available. Therefore rather than attempting to model the required neural processing directly, the effect is simulated by passing the eardrum signals through filters which have the appropriate frequency selectivity and then measuring the energy in the left and right channels. The filters used were constructed using the expression for the magnitude response of critical band filters given in [13] and centred on the critical frequencies of the basilar membrane filters. 1/3rd octave bandwidths were used, which is a valid approximation above 400 Hz. Fig. 7(a) displays the magnitude responses of a number of these filters, and the signals shown in Figs. 7(b) and 7(c) are the responses to the ear signals of Fig. 4 in the 400 Hz and 6.4 kHz bands respectively.

Analysis Point Selection and Precedence Effect:

The ability of the auditory system to determine inter-aural time differences is usually modelled by performing a running cross-correlation on the neural signals and identifying the lag at which it is maximized (for example, [11]). Continuous computation of this function is inefficient (for a software implementation) since it gives reliable information only when corresponding peaks in the neural firing rates from each ear fall inside the correlation window. To determine the points in time at which this is true, all the significant peaks in the neural rate signals are located, and a list is formed of all right-left pairs which occur less than 1 ms apart. It is only at these points that processing is done to extract the time and amplitude difference information.

This procedure also allows a simple implementation of the precedence effect since ITD and IAD measurements can be said to occur at a specific time. It is currently believed that: sounds arriving up to 1 ms after the initial sound will significantly affect localization; those delayed between 1 to 6 ms will be completely suppressed; and those delayed between 6 ms and the echo threshold (assumed to be 20 ms for the model) will be partially suppressed [14], [15], [16]. To model the precedence effect, the analysis points are assigned weights according to their time of occurrence relative to the first. Those occurring between 1 and 6 ms after the onset are removed from the list since arrivals in this range are completely suppressed. In the case of noise or periodic signals the precedence effect is not invoked, and all analysis points are assigned equal weightings. Fig. 8 shows the analysis points determined for the neural rate signals of Figs. 6(b) and 6(c). In the low-frequency band (Fig. 8(a)), the peaks not chosen as analysis points are evidence of the suppression due to the precedence effect algorithm.

Inter-aural Time Difference Measurement:

As previously mentioned, a cross-correlation is performed to determine the ITD between a pair of neural rate peaks. A 2 ms segment (centred at the analysis point) of the neural rate signal from the left channel is offset over a range of +/- 1 ms and the overlap with the corresponding segment of the right channel signal

is computed. The ITD is indicated by the offset at which the overlap is maximized. It is possible that there will be more than one local maximum in the cross-correlation, and in such a case the largest is chosen and the measurement assigned a reliability based on the difference in height of the two highest peaks. As processing proceeds, the ITD measured at each point in each frequency band is stored along with its reliability value, which is used in the weighted averaging of the estimates. Figs. 9(a) and 9(b) show ITD measurements made from the neural rate signals of Figs. 6(a) and 6(b) respectively (400 Hz and 6.4 kHz bands).

Inter-aural Amplitude Difference Measurement:

To measure IADs, the energies in segments of the critical band filter outputs are calculated for each ear. These segments are centred at points following the analysis points at a delay equal to that of the maximum of the impulse response of the filter. In principle, the segments could be shifted inter-aurally based on the ITD measurement, but it was preferred to keep the ITD and IAD measurements independent, and no adverse effects have been apparent. For impulsive stimuli, 2.5 ms segments are used, and for steady-state signals they are 10 ms long. In the case of impulses, the segment length could be adjusted for each frequency band to include an integral number of periods of filter ringing, but the constant length has proved adequate.

The IAD is calculated as the ratio of the left channel energy to the right channel energy expressed in dB. Given the sharpness of the filters and the fact that the measurement is of an energy ratio, the result should be independent of the spectrum of the source. However, if there is little energy in a particular critical band, the IAD may be based on background noise, which does not provide localization information. For this reason, the IAD measurements are assigned a weighting based on the sum of the segment energies in the right and left channels. The IAD measurement for one segment of the critical band signals of Fig. 7 is illustrated in Fig. 10(a) (400 Hz band) and Fig. 10(b) (6.4 kHz band). Note the significantly lower IAD found in the low-frequency band.

Calibration:

Using the processing steps described above, the model can be made self-calibrating, in that no other ITD or IAD measurement techniques are necessary to form the map curves used in making azimuth estimates from inter-aural difference values. A calibration is made by measuring the source-to-ear impulse responses for azimuths from -90 degrees (full right) to +90 degrees (full left) in 5 degree increments, and then processing them to extract the ITD and IAD cues. These are then used to form curves of ITD and IAD versus azimuth in each frequency band. The curves in Fig. 1 and Fig. 2(c) were derived by this method.

Azimuth Estimation from ITDs:

Once the ITD and IAD measurements are completed for each analysis point in each frequency band, azimuth estimates are derived by matching the measured values to the standard "maps". Using linear interpolation between table entries, the azimuth corresponding to the measured ITD is found and stored. All azimuth estimates inherit reliability factors from the ITDs, but if the measurement is outside the range of the map, the estimate is recorded as +/- 90 degrees with a reduced reliability, and thus at least maintains the sidedness suggested by the ITD. Fig. 11(a) illustrates the process of azimuth estimation from the ITDs extracted from the 400 Hz band in response to the signal of Fig. 4 (source at +30 degrees), and Fig. 11(b) illustrates the corresponding process for the 6.4 kHz band.

Azimuth Estimation from IADs:

In a similar manner to the ITD-based estimation, azimuths are derived from the IAD measurements by matching them to a standard map. The estimate is found by interpolating between table entries and assigning estimates of +/-90 with reduced reliability to IADs outside the map range. Unlike the ITD map, the IAD curves for certain critical bands are not monotonic with azimuth, and therefore a measured IAD may match the map at several points. An IAD value lying within 2 dB of a local maximum or minimum in the curve is also treated as a match. In order to deal with this, all the possible estimates are computed and stored, and an average over all the candidate estimates is calculated. Then for each point, the estimate which is closest to the overall average is selected as the correct one. Estimates derived from IADs which give multiple possible azimuths are weighted in reliability by the inverse of the number of map matches. The "lookup" process in the 400 Hz and 6.4 kHz bands for the signal of Fig. 4 is shown in Figs. 12(a) and 12(b).

Azimuth Estimate Processing:

The final step in the processing is to provide a single azimuth estimate and confidence interval based on the complete set of time- and amplitude-based estimates generated, and to indicate the azimuthal distribution of estimates. The first is done by performing a weighted average of all the estimates using the following factors in weighting: accumulated reliability factors, precedence model weightings, and frequency-dependent time-intensity trading ratios. This gives the mean of the data, and allows a calculation of the weighted variance, from which the 95% confidence interval for the estimate can be determined. This is a straightforward but rather arbitrary method of determining a single image azimuth, but seems to provide good agreement with expected results. To determine the appropriateness of more sophisticated techniques would require comparison with the results of tests involving human listeners, which have not been practical for our research group to perform.

In addition to the weighted average, plots of the estimates' distribution can also be generated, of which the three types shown in Fig. 13 have proved to be useful. The first, seen in Fig. 13(a), is a plot of the azimuth estimate made at each analysis point, and indicates the type of cue on which the estimate is based (ITD or IAD). The plot in Fig. 13(b) is similar, but is a histogram of the estimates in the azimuth- frequency plane. In order to form a smooth histogram, each estimate is represented by a cosine-shaped hump 9 degrees wide at the base. Here the type of cue is not visible, but the weighting of the individual estimates is applied. The data are normalized by the value of the largest point after all the estimates have been superimposed. Fig. 13(c) is also a histogram, but the estimates from all the frequency bands are combined, and the data normalized by the peak which would have occurred had they all been at the same azimuth. The sharp peak in this azimuth histogram indicates a very "crisp", well-defined image. This is due to the fact that most estimates fall close to the true source azimuth of +30 degrees, as indicated in Figs. 13(a) and 13(b). The overall weighted average +28 degrees is equal to the centroid of the azimuth histogram.

3. Experimental Verification of the Model's Performance

Before the model was used in any stereo imaging measurements, its correct performance was verified by making measurements with stimuli for which the appropriate human response was known or could be assumed. The tests involved localizing impulses from a single source in environments which either had no reflections, had early reflections, or had late reflections (as defined below), and from a stereo loudspeaker arrangement under anechoic conditions.

Single Source, Anechoic:

In the first of these, the signals employed were the same as those used in the calibration of the model. A single small loudspeaker was used as the source, the source-to-ear impulse responses measured, and the data edited to remove reflections and create a pseudo-anechoic measurement. This was done for source azimuths from -60 to +60 degrees in 10 degree increments. As shown in the previous section, the model predicts that very sharply located auditory events will be produced by such stimuli, and therefore only the overall average azimuth estimate has been plotted in Fig. 14(a). In this figure, it can be seen that the azimuth estimated by the model agrees with the actual source position within 5 degrees, with much better agreement for smaller azimuths. In all cases, the 95% confidence intervals (indicated by the error bars) cover the true source azimuth. The asymmetry in the confidence intervals can be attributed to asymmetries in the dummy head itself. Impulses presented anechoically are arguably the easiest signals to localize, and here the model performed very well.

Single Source, Early and Late Reflections:

As a test of the efficacy of the precedence effect implementation in allowing reasonable localization under reverberant conditions, single source localization experiments were performed in two reflective environments. In the first, the manikin was surrounded on three sides by large square wooden panels, which provided strong reflections arriving from various directions within 1 to 2 ms of the direct signal. The source was placed approximately 2 m away, and measurements were made at a number of azimuth positions. This is referred to as the "early" reflection case since most of them arrive before or during the "dead time" invoked by the precedence effect model.

Fig. 14(b) shows the results of the localization of impulses under these conditions. The diamond symbols show the azimuth estimates made with the manikin facing in the 0 degree reference direction, and the squares those made with the manikin facing the source itself. In the latter case, the changing source-reflector geometry did not affect either the accuracy of localization or the size of the estimate confidence intervals, while these were adversely effected when the manikin faced in the reference direction.

In the second case, the "late" reflection situation, the manikin was placed approximately 2.5 m from the walls near one corner of a hard-walled laboratory approximately $7 \times 6 \times 4$ m in size. Measurements were made with the source displaced laterally along a line 4 m in front of the manikin. In this case, most lateral reflections arrived in the reduced sensitivity period following the dead time.

The results of localization under these conditions are shown in Fig. 14(c), in which, as in the early reflection case, squares indicate that the manikin was facing the source, and diamonds that it was facing in the 0 degree direction. As in the previous case, localization was quite accurate independent of the source position in the room when the manikin faced it, while accuracy was reduced at larger azimuths when the manikin's orientation was fixed.

From the results of these two experiments it can be seen that the implementation of the precedence effect included in the model does allow localization to function in a manner at least qualitatively similar to that of humans in a reverberant environment. Additional experiments using delayed signals produced from a second loudspeaker instead of natural reflections have given good agreement with the results of similar experiments using human listeners [17]. Thus we are satisfied that the model does correctly emulate human performance under these conditions.

Intensity stereo:

The fourth experiment performed to verify the performance of the model was the localization of image sources produced by presenting impulses in stereo from a pair of loudspeakers. The usual stereo configuration was used, the speakers and listener

forming an equilateral triangle with sides 2 m in length. To avoid any problems with unmatched loudspeakers, the impulse response measurements from a single loudspeaker placed in each of the two positions were superimposed to create the input to the model. The data were again made pseudo-anechoic by editing the impulse responses to eliminate reflections. The intended source azimuth was manipulated by changing the level of the signal sent to each speaker in accordance with the "law of sines" [18]. As shown in [18] simply changing the speaker feed levels creates amplitude and time differences at the listener's ears which are approximately the same as those which would be produced by a real source at the desired image position. Therefore it was expected that the model would localize correctly (i.e., emulate human performance) in this case as with the single-source experiments.

Measurements were taken with the loudspeaker levels adjusted to position the image at azimuths from -30 degrees (right speaker) to 0 degrees (centre) to +30 degrees (left speaker) in 5 degree increments. The agreement between the intended and estimated image azimuths is clearly seen in Fig. 15(a). The slightly larger confidence intervals for non-central images are due to the fact that the law of sines is derived using a low-frequency approximation, and therefore the estimates in high-frequency bands disagree slightly with those of low-frequency. At +30 and -30 degrees, the signal comes entirely from a single loudspeaker, and therefore these estimates are the same as those made at the corresponding azimuths in Fig. 14(a), and display the same confidence-interval asymmetry. In Figs. 15(b) and 15(c), the histograms show a sharp central image indistinguishable from that of a single real source. This final test of the model is considered a good one because it shows the model capable of correctly localizing a virtual or image source as well as a real one.

4. Applications of the Model in Imaging Measurement

Having verified that the model did indeed emulate human localization performance, some applications to the measurement of the stereo imaging properties of a number of recording, playback, and signal processing systems were examined. The experiments involved: stereo presentation with speakers wired in antiphase, stereo presentation of signals subjected to a crude form of stereo synthesis, the playback of recordings made with coincident and spaced microphones, and the effect of off-centre listening on imaging.

Antiphase Stereo:

The first experiment was intended to show the effect on a central image of inverting the polarity of the signal fed to one speaker of a stereo pair. In such a case it is easy to show that, for a symmetrical head, the signals at the left and right ears are also polarity-inverted versions of each other (but are quite different from the ear signals produced when the speakers are in phase). This polarity inversion amounts to a frequency-

independent 180 degree inter-aural phase difference in all frequency bands, which is equivalent to a strongly frequency-dependent inter-aural time difference (phase delay) in each band. IADs are not affected by the inversion and remain (for a symmetrical head and equal speaker feeds) 0 dB in all bands.

The effect of these inter-aural differences is revealed in the plots of Fig. 16, which shows the result of localization on impulses presented anechoically in antiphase. As can be seen, the IAD-based estimates are still clustered around 0 degrees, but the ITD-based ones show the estimated azimuth increasing significantly at lower frequencies. The reason for this is that the ITD corresponding to a 180 degree phase shift is inversely proportional to frequency, and therefore there is very little effect at higher frequencies. For a continuous signal, the ITD-based estimates would be expected to fall symmetrically about 0 degrees, but for the impulse signals used, the half-wave rectification of the hair cells causes the start of firing to be delayed consistently in one ear, hence the pulling of the image to the left in this case. These results are in agreement with Lipshitz's observation [18] that the "phasiness" associated with antiphase presentation of white noise is not present when the signal is high-pass filtered.

Synthesized Stereo:

A second similar experiment involved the stereo presentation of impulses which had undergone complementary comb-filtering in the left and right channels to create a diffuse image similar to those created by simple "stereo synthesis" devices. The left channel signal was delayed by 0.25 ms and added to the original, while the right was inverted, delayed by the same amount, and then added. Thus spectral nulls and peaks alternate every 2 kHz in each channel with peaks in the left occurring at the same frequency as nulls in the right and vice versa. The effect of playing back these signals through stereo speakers is to produce signals at the listener's ears having ITDs and IADs which vary from band to band, especially at low frequencies, where the bandwidth is significantly smaller than that of the combing.

The result on the imaging can be seen in Fig. 17, which shows the model output for impulses presented anechoically after undergoing the above processing. A displacement of the estimates around 0 degrees which depends on frequency and cue type (ITD or IAD) can clearly be seen in Fig. 17(a), and in 17(b) this is seen to cause a smearing of the image around the central position, which corresponds to the diffuse images perceived when signals are generated by this technique.

Stereo Microphone Techniques:

In the above section we considered simple manipulations of the signals fed to the stereo loudspeakers which affect the resulting images. Another important consideration is the manner in which the signals are initially recorded. There is some controversy over the best stereo microphone techniques to use

under various circumstances, and by using the model it is perhaps possible to quantify the differences between them. To this end, the imaging possible with coincident hypercardioid, coincident figure-8, and spaced omnidirectional microphones was investigated.

The experiments were performed with the microphones placed near the centre of the laboratory described previously, and used a small loudspeaker 2 m from the centre of the microphone array as the source to be recorded. This was positioned at azimuths from 0 to 70 degrees in 10 degree increments. Sound-absorbing foam on the floor and ceiling was used to reduce reflections from these surfaces, but no attempt was made to control lateral reflections. The recordings were made by measuring the speaker- to-microphone impulse responses with the MLS technique used for the ear signal measurements. In order to simulate playback under anechoic conditions, these recordings were convolved with the pseudo-anechoic speaker-to-ear impulse responses for sources at +30 and -30 degrees (which had already been obtained) and the resulting left- and right- channel signals summed to give the overall record-playback source-to-ear impulse responses. To reveal any differences between the imaging of impulsive and continuous stimuli, additional recordings were generated by convolving the record-playback IRs with a 200 Hz square wave signal. Doing these experiments "in software" proved to be significantly easier, less time-consuming, and more flexible than physically setting up the appropriate equipment.

Coincident Microphones:

The first such experiment employed a Calrec Soundfield microphone synthesizing a pair of coincident figure-8 microphones with patterns angled at 90 degrees. This gives an in-phase azimuth coverage for frontal incidence from -45 to +45 degrees, and since the figure-8 patterns provide channel levels in accordance with the law of sines [18], the image azimuths should correspond linearly to the source azimuths, but scaled to lie within the angle subtended by the speakers. Thus a source at +45 degrees is expected to image at +30 degrees, and one at +22.5 degrees at 15 degrees. Fig. 18(a) shows the results of the localization of the playback signals generated with this microphone configuration. The diamond symbols represent the image location of impulses and the squares those of the square waves. For source azimuths less than 45 degrees, the expected relationship between image and source positions holds approximately, and is at least monotonic. Beyond 45 degrees, the images fall back toward the centre and the confidence intervals increase. This is to be expected since the recorded signals are in antiphase and approaching equal level at 90 degrees incidence. It should be noted that the impulses and the square waves image fairly consistently, the differences perhaps being due to the lack of a precedence effect in the continuous signal case.

A similar experiment was performed with the Soundfield microphone synthesizing two hypercardioids angled at approximately 110 degrees. In this case, the in-phase frontal coverage is from -55 to +55 degrees, but apart from this difference the imaging characteristics should be similar to those of the figure-8's. The

localization results of Fig. 18(b) show reasonably smooth movement of the image as source azimuth increases, although the square wave images again seem to have been affected by reflections. Agreement between the impulse and continuous image positions is evident, as is the diffuse nature of images of sources beyond 55 degrees. Images in this region are not pulled back towards the centre as strongly as those in the case of the figure-8 microphones, since the rear lobes of the hypercardioid microphone patterns are much smaller than those of figure-8's.

Spaced Microphones:

The experiments with spaced microphones employed the same measurement technique, but a pair of Bruel & Kjaer 4133, 1/2-inch, free-field microphones were used to make the recordings. these were mounted vertically to provide an omnidirectional (although non-flat) response in the horizontal plane, and the measurements performed with spacings of 20, 50, and 100 cm. The characteristics of the recorded signals which change with source azimuth are the inter- channel delay and, to a lesser extent for close spacing, the channel levels. The images created are the result of quite different physical and psychoacoustic phenomena than for coincident microphones [18]. Thus, as expected, the image localization results were quite different from those obtained with the coincident microphones.

The differences are evident in Fig. 19(a), which shows the results for the 20 cm spacing. While the impulse image moves evenly with increasing source azimuth, that for the continuous signal is pulled at most half-way towards the left speaker, and for most source positions lies between 5 and 10 degrees. The discrepancy between the two image paths is in part a function of the precedence effect, allowing the first-arriving signal to dominate in the case of the impulses, and its absence in the case of the square waves. In addition, such a close microphone spacing does not result in significant inter-channel level differences for a source 2 m away, and therefore IADs will produce azimuth estimates near 0 degrees in the square-wave case, since the periodic signals from the two speakers are superimposed at the ears.

The results for spacings of 50 and 100 cm (admittedly an extreme case for the small distance to the source) are shown in Figs. 19(b) and 19(c), and show a similar difference between the image paths of the impulses and square waves. However as the spacing is increased, two effects appear. The first is that at sufficiently large azimuths, the signal from the speaker reproducing the delayed channel arrives within the precedence effect dead time, and thus the localization is dependent primarily on the first-arriving signal. Because of this, the image of the impulses tends to collapse into the leading speaker, as can be seen in Figs. 19(b) and 19(c). This is perhaps a manifestation of the "hole-in-the-middle" sometimes experienced with spaced-microphone recordings. The second effect is apparent in the 100 cm spacing results, in which the square-wave image is seen to be pulled towards the leading speaker for large source azimuths, and is due to the inter-channel level differences produced when

the ratio between the two source-to-microphone distances is significantly different from unity.

Off-centre Listening Effects:

Having examined some signal processing and recording techniques, the final application of the model which was attempted was the measurement of differences in imaging due to different loudspeaker designs. Specifically, experiments were done to determine the effect of off-centre listening on the localization of an (intended) central image. The speakers used were small 2-way box speakers (PSB Avanté) and dbx Soundfield 100s, which were designed to expand the range of listening positions resulting in reasonable imaging. For the box speakers, the experiment was done "in software", but since the design of the Soundfields requires that they be placed close to a wall and because the speakers are mirror images, their measurement had to be performed physically. The experiment consisted of supplying each speaker with the same signal (indicating a central image) and making the localization judgement at a number of points displaced from the midline along a line parallel to that passing through the speakers, as shown in Fig. 20(a). As with the previous experiments, both impulses and 200 Hz square waves were used in order to expose any differences between the imaging of impulsive and continuous signals.

Figure 20(b) shows the imaging results for the box loudspeakers. The azimuth estimate generated by the localization model has been converted to displacement of the image from the point midway between the speakers. As can be seen, both the impulse and periodic signal images collapse rapidly into the nearer speaker (located at 100 cm) as the listener moves away from the midline. This is due both to the precedence effect and the increased level of the nearer speaker at the listener's position. The smaller displacements of the square-wave image for listener positions from 50 to 70 cm are probably due to interference between the signals from the two speakers at the ears, although this has not been verified. The results for the Soundfields are quite different, as Figure 20(c) reveals. The polar patterns of these speakers are such that moving away from the centre position brings the listener more onto axis of the distant speaker, thus correcting the level imbalance. This cannot, however, prevent the collapse of impulse images into the nearer speaker, which is caused by the precedence effect, since the arrival time differences are not corrected. The technique is effective for the periodic signal images, which are kept quite close to the central position. Informal listening tests confirmed these results.

5. Conclusions

The results of the experiments detailed in the previous two sections have shown that the present binaural localization model is of use in measuring stereo imaging. Although exact correspondence between the judgements made by the computer model and those of human listeners should be confirmed by formal listening tests, the results are at least qualitatively those that are expected. This is true for the single source localization

under both anechoic and reverberant conditions and for the imaging measurements involving stereo reproduction. Using the model, the precise imaging available with pan-pot and coincident microphone recordings was distinguishable from the results obtained with spaced microphones, and the partial efficacy of a special loudspeaker design in reducing the image collapse caused by off-centre listening was demonstrated. Clearly such discriminations could be made using human listeners, but the model offers the benefits of results which are objective, repeatable, and quantified.

There are several ways in which the model could be improved in terms of both sophistication and speed. The weighted averaging employed in making the final azimuth estimates could be replaced if more data were available on the interaction of localization cues in different frequency bands. Work is being done on replacing the 95% confidence interval by some value derived from the azimuth histogram as a measure of image width, and some means of correctly dealing with signals containing both transient and steady state components may be attempted. At present the model is being run on a IBM-compatible computer with an 80386 microprocessor, an 80387 math coprocessor, and a 16 MHz clock, and requires 6 to 7 minutes to process one measurement. The bulk of this time is devoted to the basilar membrane and critical band filtering operations, which are done using FIR filters and convolutions performed by optimised assembly language FFT routines [19]. If these filters were implemented recursively and the operations performed by a dedicated DSP chip, the processing time would be significantly reduced. Such an implementation may be attempted.

6. Acknowledgements

I would like to thank John Vanderkooy and Stanley P. Lipshitz (Audio Research Group), and Phillip Bryden (Department of Psychology) of the University of Waterloo for their assistance and advice, and Earl Geddes and Henry Blind of Ford's Diversified Products Technical Center for their support of the project. Both the research and the author were supported by funding provided by the Ford Motor Company of America and the Natural Sciences and Engineering Research Council of Canada.

7. References

[1] M.W. Pocock, "A Computer Model of Binaural Localization," presented at 72nd Convention of the Audio Eng. Soc., preprint no. 1951, (1982).

[2] M.W. Pocock, *A Computer Model of Binaural Localization*, M.App.Sci. Thesis, University of Waterloo, Waterloo, Ontario, (1983).

[3] B. Rakerd and W.M. Hartmann, "Localization of Sound in Rooms, III: Onset and Duration Effects," *J. Acoust. Soc. Am.*, vol. 80, no. 6, pp. 1695-1706, (1986).

[4] S.S. Stevens and E.B. Newman, "The Localization of Actual Sources of Sound," *Am. J. Psych.*, vol. 48, pp. 297-306, (1936).

[5] D. McFadden and E.G. Pasanen, "Localization at High Frequencies Based on Inter-aural Time Differences," *J. Acoust. Soc. Am.*, vol. 59, no. 3, pp. 634-639, (1976).

[6] J. Blauert, *Spatial Hearing*, (MIT Press, 1983).

[7] S.A. Gelfand, *Hearing: An Introduction to Physiological and Psychological Acoustics*, (Marcel Dekker Inc., 1981).

[8] B.C.J. Moore, *An Introduction to the Psychology of Hearing*, (Academic Press, 1982).

[9] J.O. Pickles, *An Introduction to the Physiology of Hearing*, (Academic Press, 1982).

[10] D.D. Rife and J. Vanderkooy, "Transfer-Function Measurement with Maximum-Length Sequences," *J. Audio. Eng. Soc.*, vol. 37, no. 6, (1989).

[11] J. Blauert and W. Cobben, "Some Considerations of Binaural Cross-Correlation Analysis," *Acustica*, vol. 39, pp.96-104, (1978).

[12] M.R. Schroeder, "Models of Hearing," *Proc. IEEE*, vol. 63, no. 9, pp. 1332-1350, (1975).

[13] R.D. Patterson, "Auditory Filter Shape," *J. Acoust. Soc. Am.*, vol. 55, pp. 802-809, (1974).

[14] P.M. Zurek, "The Precedence Effect and Its Possible Role in the Avoidance of Interaural Ambiguities," *J. Acoust. Soc. Am.*, vol. 67, no. 3, pp. 952-964, (1980).

[15] B. Rakerd and W.M. Hartmann, "Localization of Sound in Rooms II: The Effects of a Single Reflecting Surface," *J. Acoust. Soc. Am.*, vol. 78, no. 2, pp. 524-533, (1985).

[16] P.M. Zurek, "The Precedence Effect," in *Directional Hearing*, ed. W.A. Yost, (Springer-Verlag, 1987).

[17] H. Wallach, E.B. Newman, amd M.R. Rosenzweig, "The Precedence Effect in Sound Localization," *Am. J. Psych.*, vol. 62, no. 3, pp. 315-336, (1949).

[18] S.P. Lipshitz, "Stereo Microphone Techniques ... Are the Purists Wrong?," *J. Audio Eng. Soc.*, vol. 34, no. 9, pp. 716-744,(1986).

[19] D.D. Rife, *DRA 387FFT Reference Manual*, (DRA Associates, 1988).

Figure Captions

Fig. 1 Inter-aural time difference versus source azimuth in two frequency bands used by the model. Measurements made on a KEMAR manikin.

Fig. 2 Inter-aural amplitude differences for a source 30 degrees to the left. The graphs are (a) source-to-ear transfer functions for left and right ears, (b) IAD versus frequency determined by the transfer functions in (a), (c) IAD versus azimuth in three frequency bands. Measurements made on a KEMAR manikin.

Fig. 3 Flowchart showing the processing stages employed by the model in obtaining a localization judgement.

Fig. 4 Signals created at the ear canal entrances of a KEMAR manikin by an impulse generated at an azimuth of 30 degrees left. The graphs show (a) the signals on an expanded scale and (b) the same signals on the scale used in the time-function graphs which follow.

Fig. 5 The basilar membrane modelling. The graphs show (a) frequency responses of six of the sixteen basilar membrane filters used by the model, (b) responses of the left and right filters in the 400 Hz band to the signal of Fig. 4, (c) responses of the filters in the 6.4 kHz band to the same signal, (d) responses of the filters in the 400 Hz band to a 200 Hz square wave generated at an azimuth of +30 degrees, and (e) responses of the filters in the 6.4 kHz band to the same signal.

Fig. 6 The hair cell modelling. The figures show: (a) RC circuit modelling the hair cells, firing rate is current I after smoothing; (b), (c), (d), and (e), neural firing rates in response to BM motions of Fig. 5 (b), (c), (d), and (e) respectively.

Fig. 7 The critical band modelling. The graphs show (a) frequency responses of six of the sixteen critical band filters, (b) responses of the left and right critical band filters at 400 Hz to the signal of Fig. 4, (c)

responses of the filters in the 6.4 kHz band to the same signal.

Fig. 8 Analysis point selection. The figures show (a) the analysis points selected in the 400 Hz band in response to the signal of Fig. 4, (b) the analysis point in the 6.4 kHz band in response to the same signal.

Fig. 9 Inter-aural time difference measurements. The figures are cross-correlation and ITD determination at one analysis point in (a) the 400 Hz band, and in (b) the 6.4 kHz band. Both are for the signal of Fig. 4, source at +30 degrees.

Fig. 10 Inter-aural amplitude difference measurements. The figures show the integration of the energy in analysis segments of the critical band signals in (a) the 400 Hz band and (b) the 6.4 kHz band. Both are for the signal of Fig. 4, source at +30 degrees.

Fig. 11 ITD-based azimuth estimation from the ITD measurements of Fig. 9. The figures illustrate the estimation procedure for the ITDs from (a) the 400 Hz band and (b) the 6.4 kHz band.

Fig. 12 IAD-based azimuth estimation from the IAD measurements of Fig. 10. The figures illustrate the estimation procedure for the IADs from (a) the 400 Hz band and (b) the 6.4 kHz band.

Fig. 13 Processing of the set of azimuth estimates for the signal of Fig. 4, source at +30 degrees. The figures are (a) estimate plot showing the azimuth and cue origin (ITD or IAD) of each estimate against band frequency, (b) frequency-azimuth histogram showing strength of weighted estimates in each band, and (c) azimuth histogram showing overall strength of weighted estimates. Numerical values are the overall weighted average and the 95% confidence interval.

Fig. 14 Single source impulse localization experiment results. Azimuth estimate is plotted against source azimuth for (a) anechoic case, (b) early reflections case, and (c) late reflections case.

Fig. 15 Intensity stereo imaging results for impulses. The figures show (a) image azimuth vs. azimuth predicted by the law of sines, (b) frequency-azimuth histogram for central image, and (c) azimuth histogram for central image.

Fig. 16 Anti-phase stereo imaging results for impulses. The figures are (a) estimate plot for equal level speaker feeds, (b) frequency-azimuth histogram for equal levels, and (c) azimuth histogram for equal levels.

Fig. 17 Synthesized stereo imaging results for impulses. The figures are (a) frequency-azimuth histogram for central image, and (b) azimuth histogram for central image.

Fig. 18 Imaging measurements for coincident microphone recordings. Diamond symbols represent results for impulses and squares results for 200 Hz square waves. The plots are for (a) figure-8's at 90 degrees, and (b) hypercardioids at 110 degrees.

Fig. 19 Imaging measurements for spaced omnidirectional microphone recordings. Diamond symbols represent results for impulses and squares results for 200 Hz square waves. The plots are for (a) 20 cm spacing, (b) 50 cm spacing, and (c) 100 cm spacing.

Fig. 20 Effects of off-centre listening. The figures are (a) diagram showing the experimental arrangement, (b) results for simple box loudspeakers, and (c) results for dbx Soundfield 100 loudspeakers. Diamonds represent measurements using impulses and squares those using 200 Hz square waves.

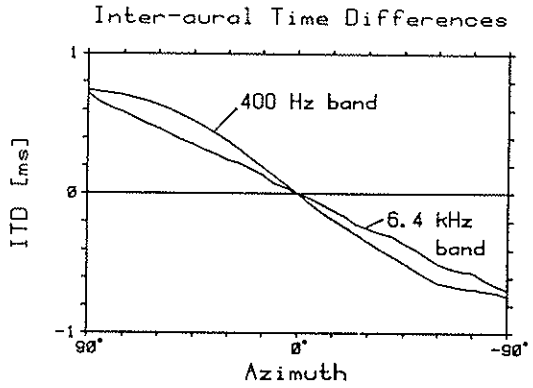


Figure 1

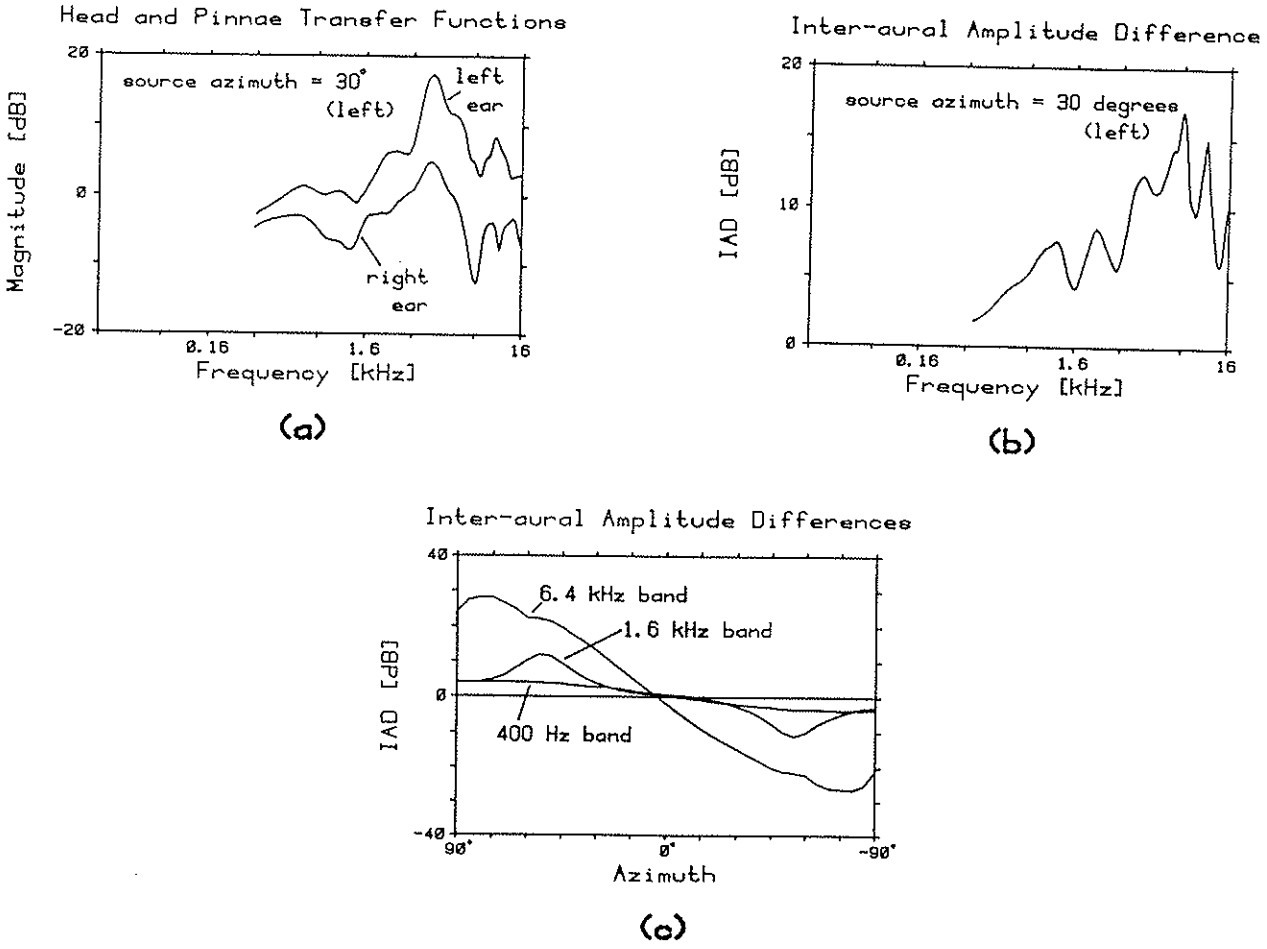


Figure 2

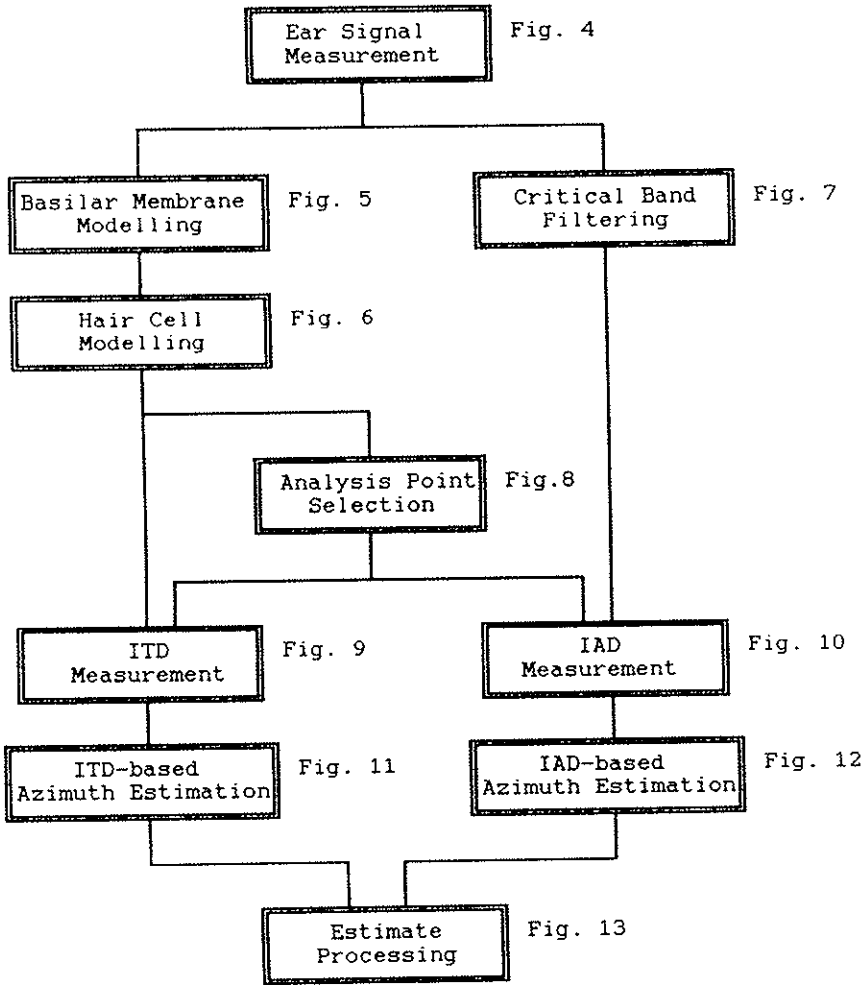


Figure 3

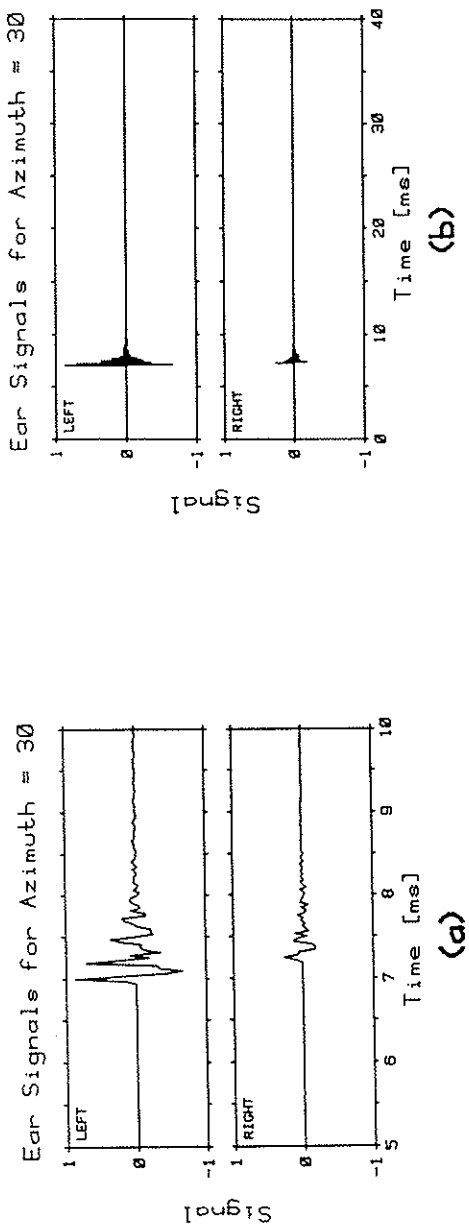
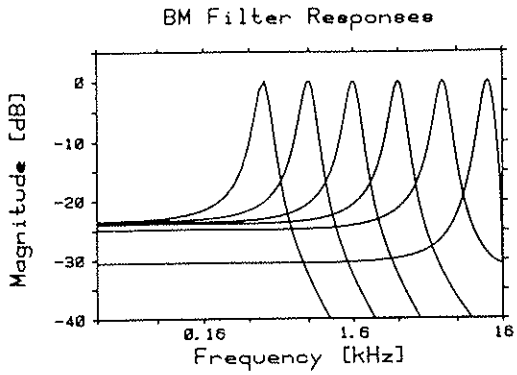
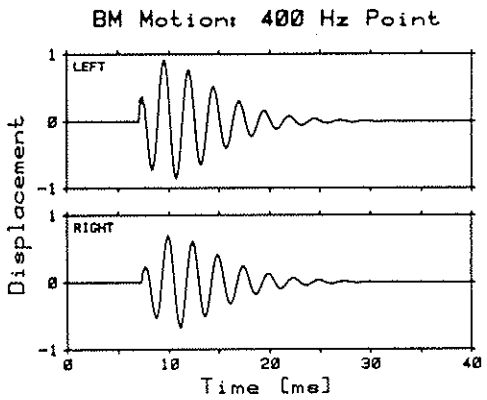


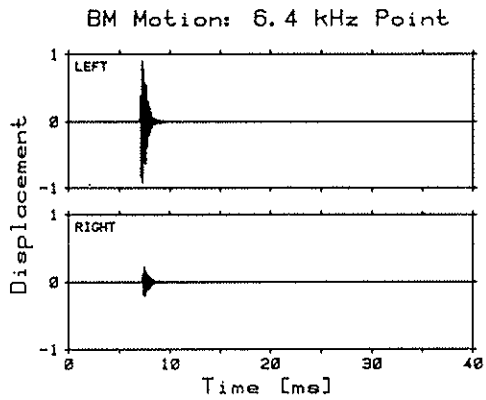
Figure 4



(a)

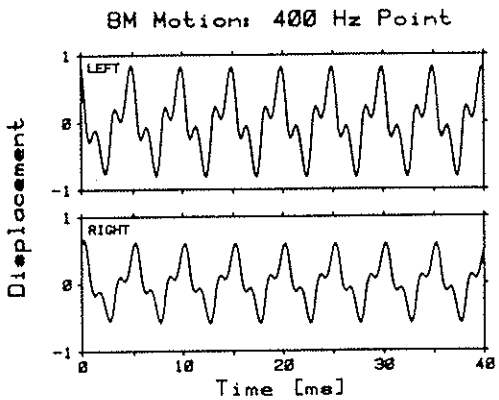


(b)

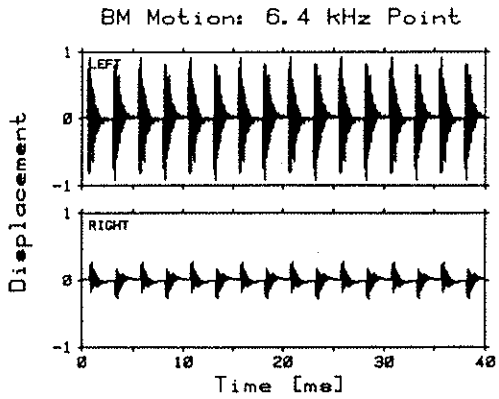


(c)

Figure 5

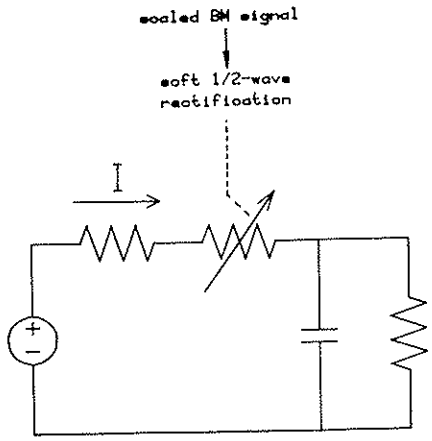


(d)

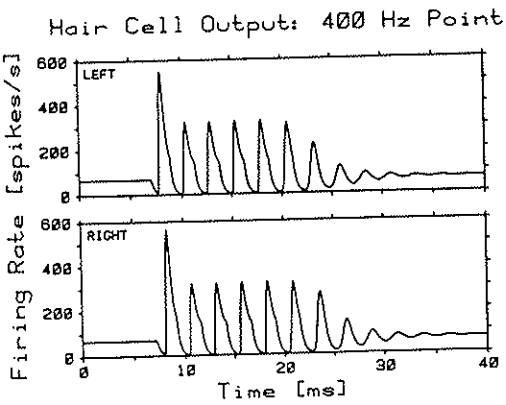


(e)

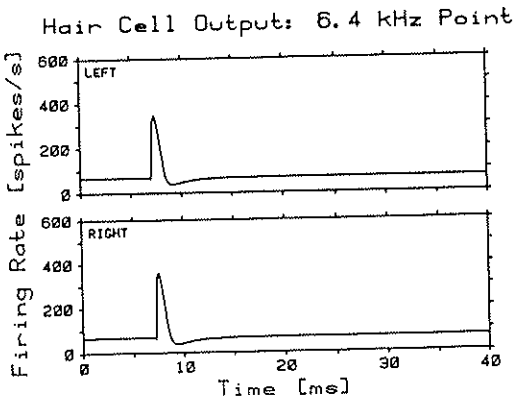
Figure 5



(a)

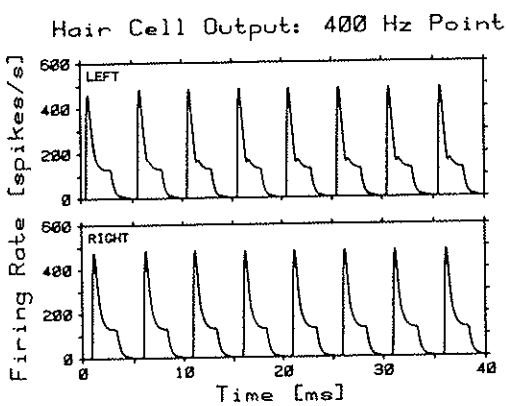


(b)

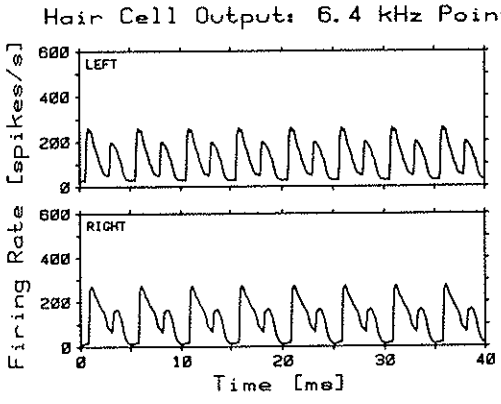


(c)

Figure 6



(d)



(e)

Figure 6

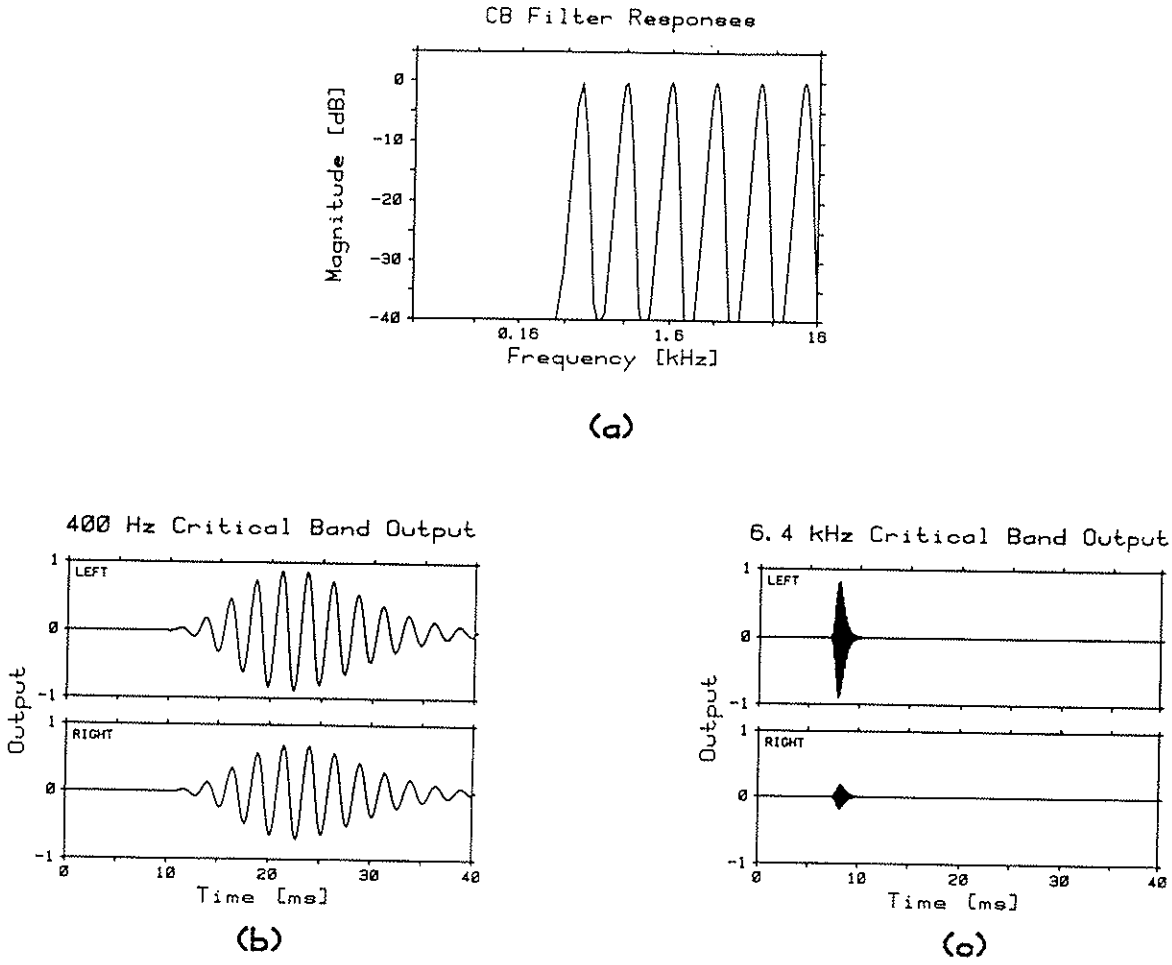


Figure 7

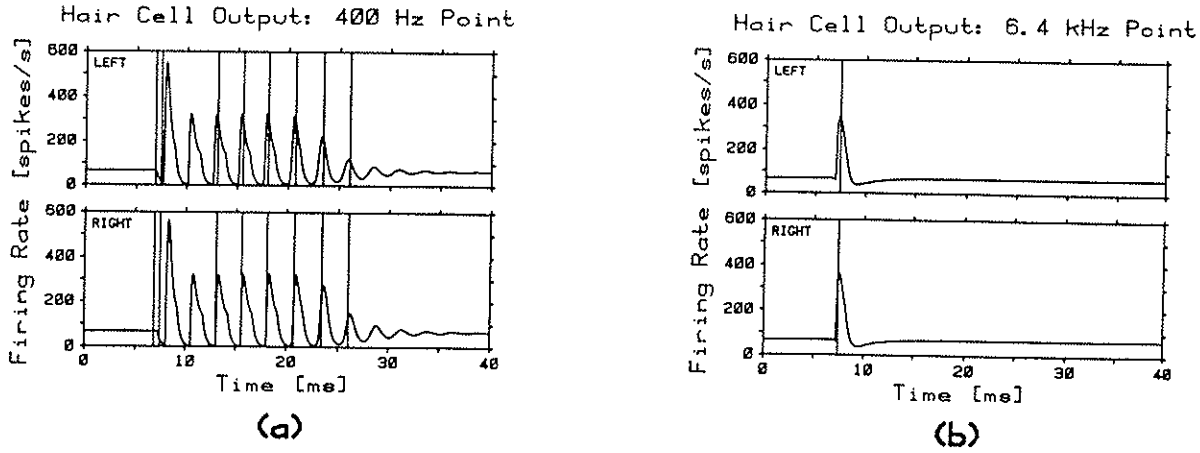


Figure 8

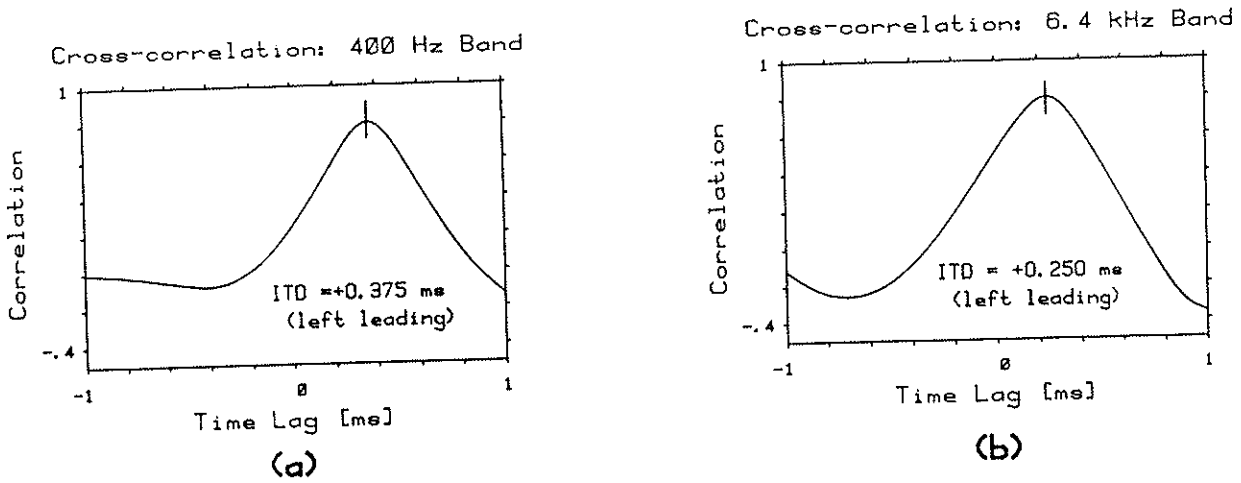


Figure 9

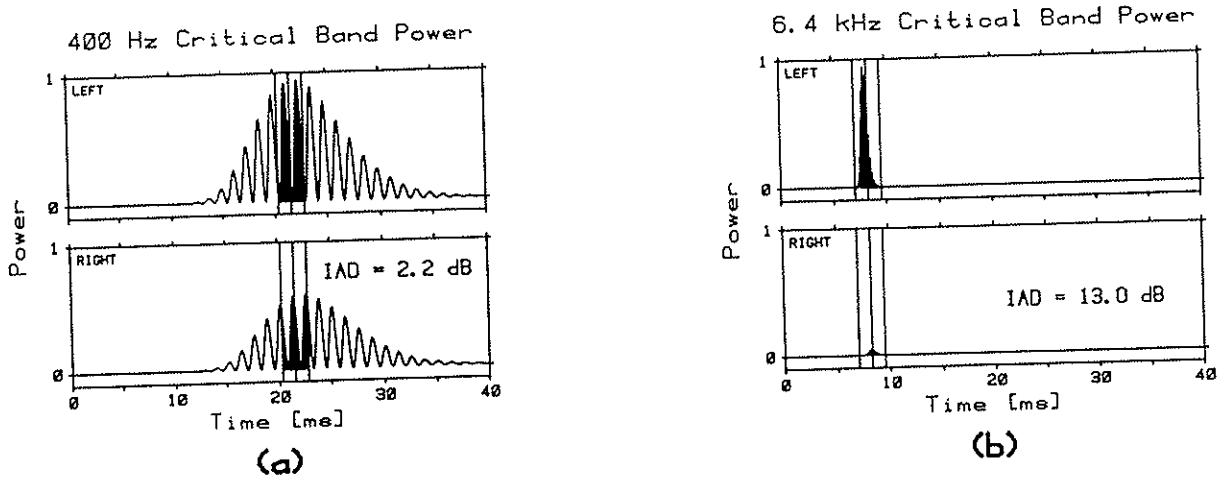


Figure 10

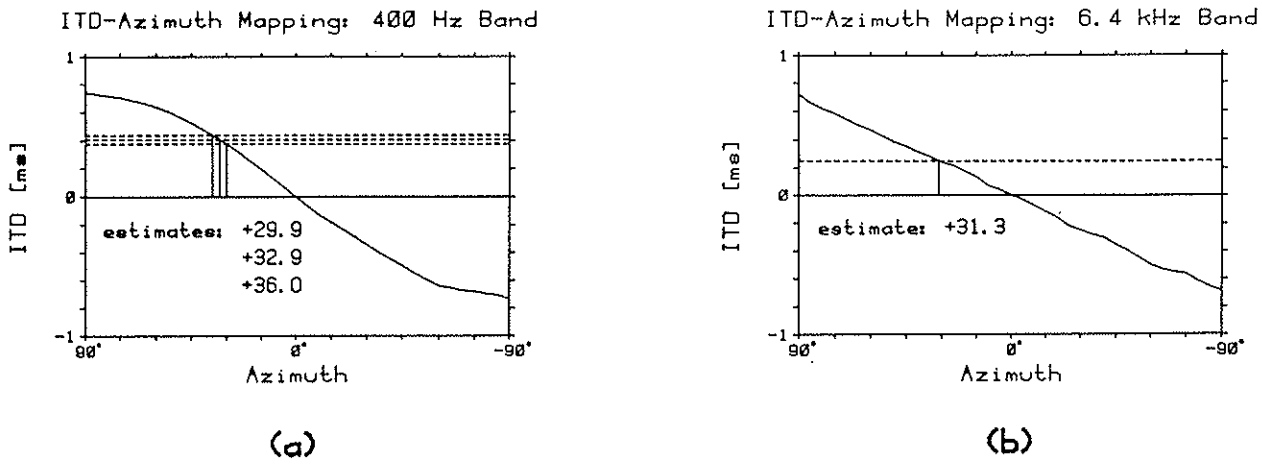


Figure 11

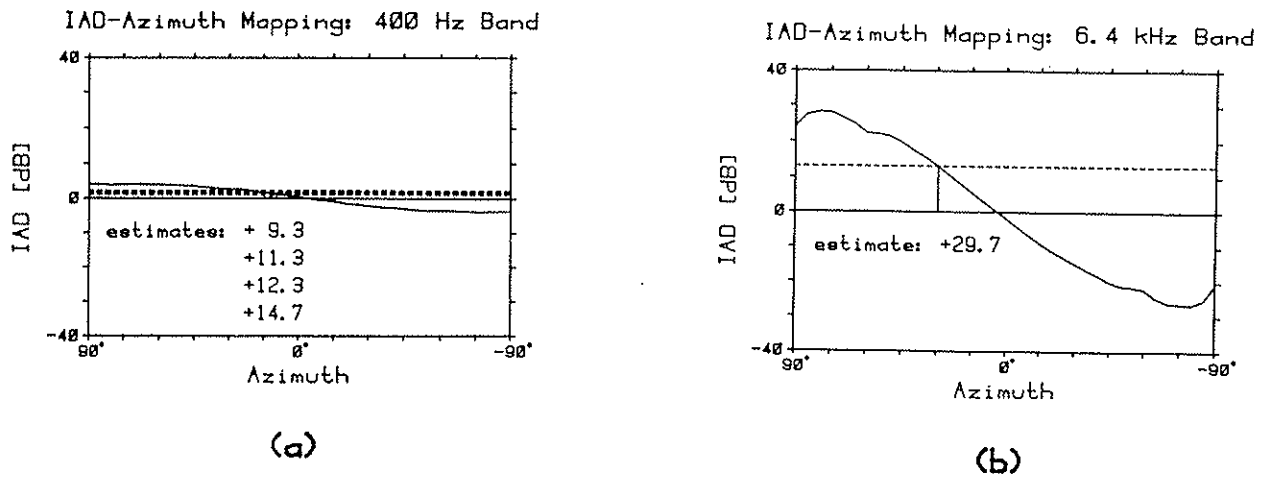
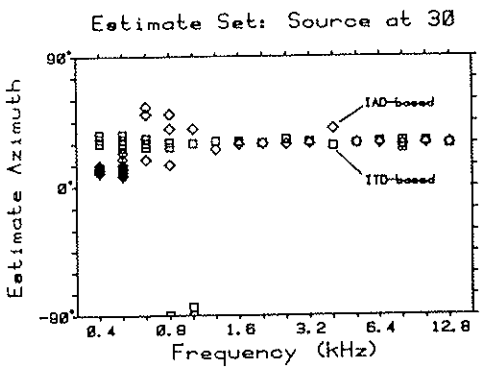
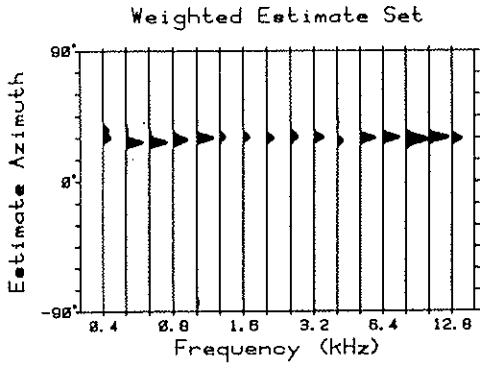


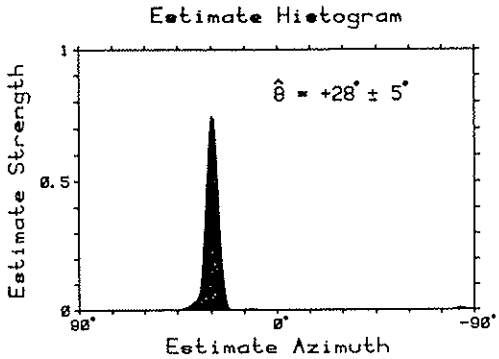
Figure 12



(a)

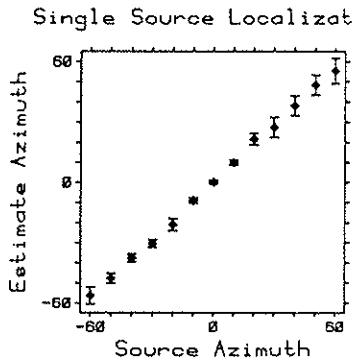


(b)

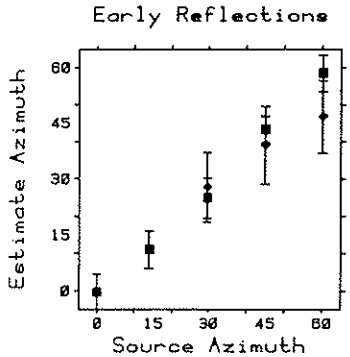


(c)

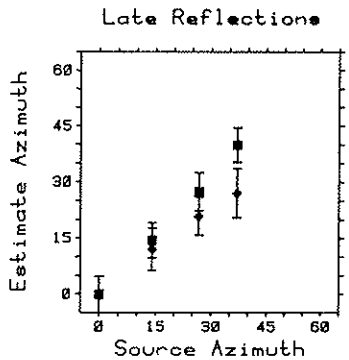
Figure 13



(a)

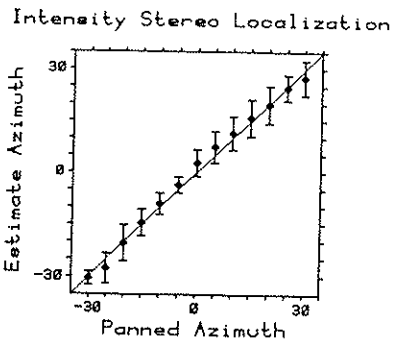


(b)

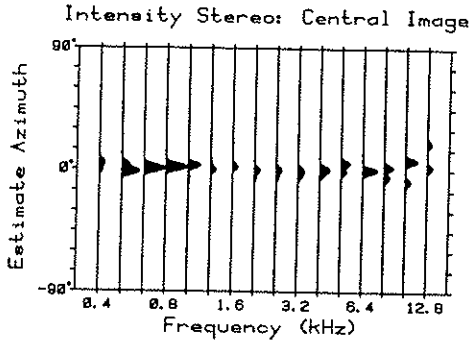


(c)

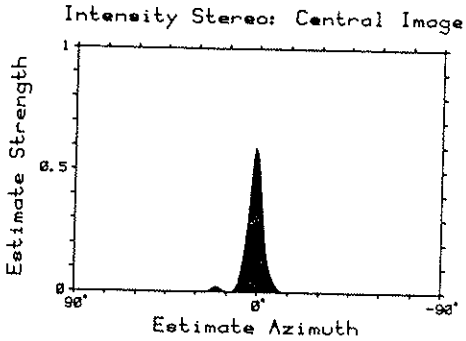
Figure 14



(a)

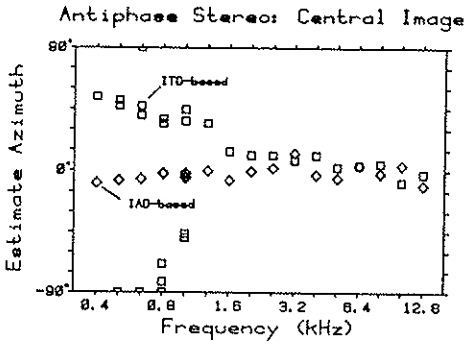


(b)

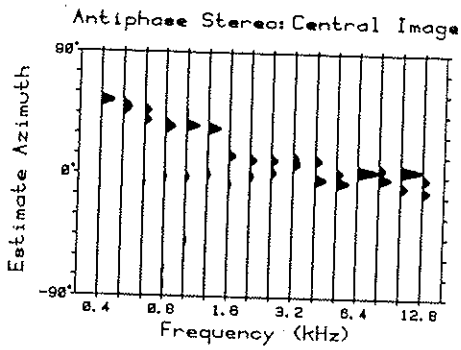


(c)

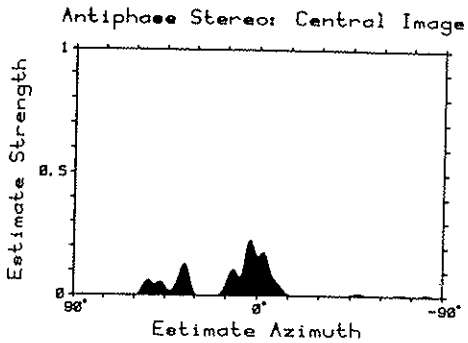
Figure 15



(a)



(b)



(c)

Figure 16

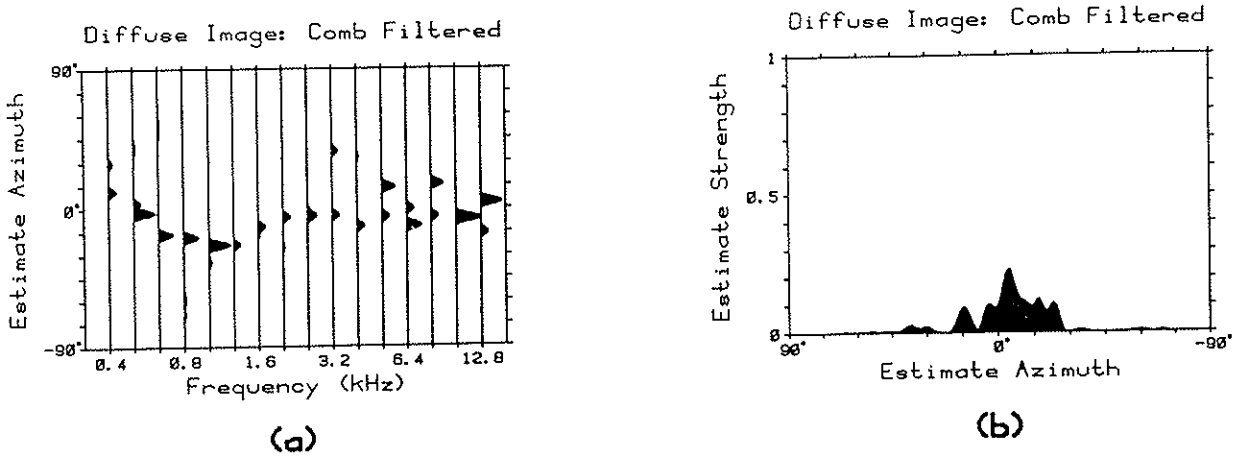


Figure 17

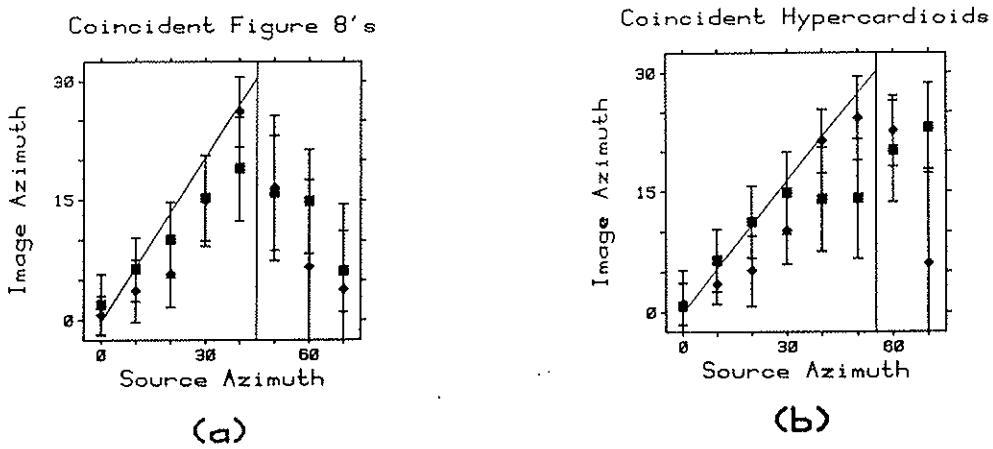
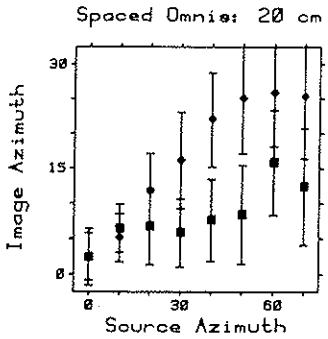
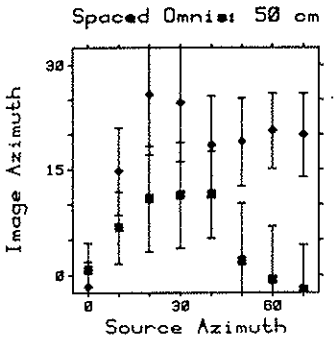


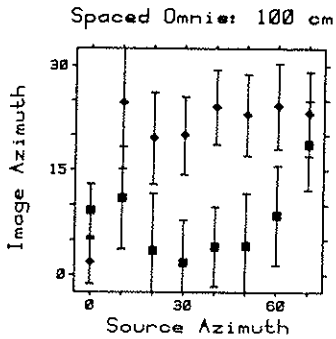
Figure 18



(a)

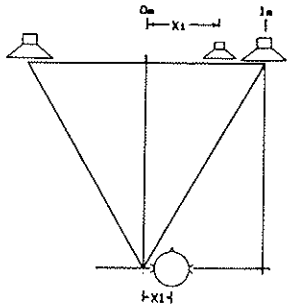


(b)

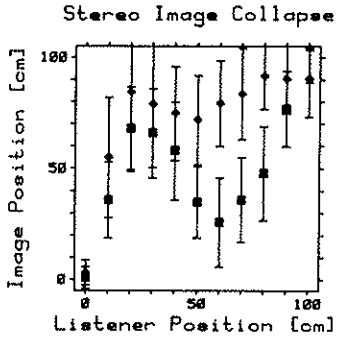


(c)

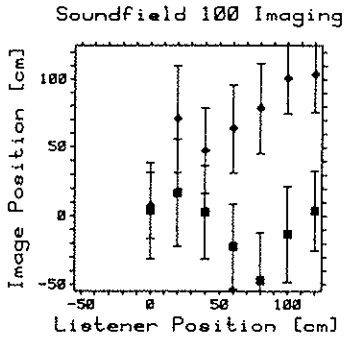
Figure 19



(a)



(b)



(c)

Figure 20