Characterization of Acoustic Head-Related Transfer Functions for Nearby Sources

by

Tara J. Brown

Submitted to the Department of Electrical Engineering and Computer Science

in Partial Fulfillment of the Requirements for the Degrees of

Bachelor of Science in Electrical Engineering

and Master of Engineering in Electrical Engineering and Computer Science

at the Massachusetts Institute of Technology

December 7, 2000

Author ______________________________
Department of Electrical Engineering and Computer Science
December 7, 2000

Certified by ______________________________
Nathaniel Durlach
Thesis Supervisor

Accepted by ______________________________
Arthur C. Smith
Chairman, Department Committee on Graduate Theses

Characterization of Acoustic Head-Related Transfer Functions for Nearby Sources
by
Tara J. Brown

Submitted to the
Department of Electrical Engineering and Computer Science

November 20, 2000

In Partial Fulfillment of the Requirements for the Degrees of
Bachelor of Science in Electrical Engineering
and Master of Engineering in Electrical Engineering and Computer Science

ABSTRACT

Head Related Transfer Functions (HRTF) for nearby sources were characterized for human subjects and the Knowles Electronics Manikin for Acoustic Research (KEMAR). The HRTFs were measured in several reverberant room conditions for sound source positions at 0, 45, and 90 degrees in azimuth and 15cm and 1m from the center of the head using maximum-length sequences. The effects of reverberation on interaural time difference, interaural level difference, magnitude response, and spectral content were examined. Results were consistent with acoustic theory. Reverberation was found to decrease ILD, cause comb-filtering, and to distort and add frequency-to-frequency variations to acoustic cues. These affects were most pronounced in room positions with the most reverberant energy.

Thesis Supervisor: Nathaniel Durlach
Title: Senior Research Scientist

1.0 Introduction

1.1 Head Related Transfer Functions

As a sound travels from its source to the eardrums of a human listener it is transformed by interactions with the listener's head, neck, torso, and outer ear. This transformation, which is unique to each listener, varies with the location of the sound source relative to the head, and is described by the listener's Head Related Transfer Function (HRTF). In the far field, more than one meter from the head, the HRTF is independent of distance and varies with azimuth and elevation. In the near field, within one meter of the head, the HRTF is dependent on distance as well as azimuth and elevation.

Important cues for sound-source localization are contained in the HRTF, including interaural time differences (ITDs) and interaural level differences (ILDs) (defined below) and monaural spectral content. The interaural difference cues are perceptually robust (i.e., they can be unambiguously determined, independent of source characteristics) while the monaural spectral cues are dependent on knowing the source spectrum. The ITD is the time difference in the arrival of the sound at the listener's left and right ears, and results mainly from differences in path length from the source to the two ears. ITD cues are primarily used to determine the angle between the vector from origin to source and the interaural axis. ILDs are the differences in sound intensity at the listener's right and left ears, and arise primarily from the acoustic interference of the head. ILDs are largest for frequencies whose wavelengths are small relative to the size of the head. In anechoic space, ILDs enable the listener to determine source distance from the head for near-field sources [1, 2, 3]. Perception of sound elevation and front/back direction is determined from the monaural spectral cues generated by the direction-dependent filtering of the pinnae [4, 5]. Due to interactions with the head and ears, some frequencies are amplified while others are attenuated in a manner dependent on source direction. The pinnae are essential for even reasonable localization performance in the median vertical plane, and significantly improve localization accuracy in the lateral vertical plane [4, 5].

Until recently, most research focused on HRTFs in the far field. Recent near field research has used spherical representations of the head with diametrically opposed ears to create mathematical models to estimate the naturally-occurring binaural cues for nearby sources. In addition, the Knowles Electronics Manikin for Acoustic Research (KEMAR), a physical model of the average human head, was used by Brungart to measure the most complete set of near-field HRTFs currently available [1, 2, 3]. This set includes measurements at every degree in azimuth and at distances from 12 to 100 centimeters from the center of the head.

In this study, HRTFs for nearby sources were measured on human subjects (for the first time) as well as for KEMAR in a reverberant room. The spatial properties of these HRTFs are compared to those predicted by the spherical head model. Individual differences across subjects are also examined. To understand how a room affects acoustic spatial cues, "pseudo-anechoic" HRTFs (in which only that part of the HRTF that corresponds to the direct signal reaching the ear is analyzed) are compared with measurements containing reverberation. Possible sources of error, such as artifacts produced by the measurement equipment and processing, are identified and analyzed.

1.2 Spatial Properties of HRTFs

Basic spatial properties of the HRTF can be determined from the spherical head model. The major characteristics common to the spherical model and KEMAR HRTFs include: the magnitude of the HRTFs increases with frequency when there is a direct path from the source to the ear, the high frequency responses of the HRTFs are attenuated when the ear is in the acoustic shadow of the head, the overall gain of the HRTFs increases as distance decreases when a direct path exists between the source and the ear, the overall attenuation of the HRTFs increases as distance decreases when the ear is shadowed by the head, the magnitude of the HRTF increases more rapidly at low frequencies than at high frequencies as the source approaches the head, and the "acoustic bright spot" seen in the sphere model is also present in the KEMAR measurements, although its structure is more complicated [1]. It follows from these major characteristics that in both the KEMAR and spherical head models, ILDs increase substantially both as a source approaches the head and as the frequency increases. The ILD is near zero in the median plane and generally increases as the source moves lateral to the head. In contrast, the interaural time difference (ITD) remains constant, roughly independent of distance.

The spherical head model provides a good first-order approximation to binaural cues such as ITD and ILD, but does not incorporate some of the features of the HRTF predicted by KEMAR, such as those due to the irregular shape of the head and the outer ear as well as those due to diffraction by the neck and torso. The spherical model fits the KEMAR measurements best at low frequencies (below 2 kHz). At high frequencies (i.e., above 5 kHz), KEMAR HRTFs exhibit direction-dependent peaks and notches that arise from the geometry of the pinna and are therefore not present in the spherical model. Because these high frequency characteristics are strongly dependent on ear geometry (which varies from person to person), it is unlikely that KEMAR's high frequency characteristics will match those of any particular human listener. As a result, KEMAR HRTFs may not predict how well listeners can judge source elevation.

One past study analyzed sound localization of nearby sources in anechoic space [1, 3]. These results were recently compared with results from a comparable study in echoic space [6]. Generally, errors were larger in reverberant space than in anechoic space; however directional perception of source distance was better in reverberant space than in anechoic space (i.e., the reverberation actually improves performance) [7, 8]. In addition, localization performance improved with practice in the reverberant case, but not in the anechoic case [8]. These results imply that localization cues are distorted by reverberation, but that listeners can gradually learn to overcome errors due to such distortion. This study analyzes and compares "anechoic" and "reverberant" HRTFs to estimate the form and magnitude of the distortions of spatial cues by reverberation. In order to determine the extent to which spatial properties of HRTFs are distorted by reverberation, HRTFs are recorded in echoic space. The impulse response of the HRTFs is then time windowed, so that only the HRTF resulting from the direct sound remains and the "anechoic" HRTF can be estimated. The spatial properties of the windowed HRTF are compared to those of the non-windowed HRTF (which contains room echoes) to determine the effects of reverberation. It was expected that the reverberation would degrade the binaural ITD and ILD cues, as well as the monaural spectral cues. Reverberation was expected to decrease the magnitude of the ILD since the reverberation will, on average, arrive from all possible directions and cause near zero ILD. Similarly, ITD information was expected to be disrupted since echoes will superimpose on the direct sound and alter interaural phase characteristics. Finally, spectral cues were expected to be affected by the addition of correlated, delayed sounds, which can cause comb-filtering effects on the spectrum.

2 Experimental Set-up, Equipment, Techniques, and Procedures

2.1 Experimental Set-up

For the HRTF measurements on KEMAR as well as the human subjects, a wooden chair was placed on a small platform approximately 1 foot high in the center of the room. The chair was placed on this platform to increase the distance between the microphones and the floor so that the first echo (off the floor) would arrive with as large a delay as possible. This allows the direct sound to die out before the first echo is recorded. The subjects sat on the wooden chair with their head against a headrest, which had been attached to the chair's back. The headrest was used to reduce head movement as much as possible. Microphones mounted in earplugs were inserted into the subject's ear canals. A head-tracking device was used to center the subject's head on the coordinate axis in the center of the room. The sound source was controlled by a computer connected via optical cables to a Tucker-Davis Technologies (TDT) signal processing system. The output of the D/A converter of the TDT system was fed into an amplifier and then into a BOSE mini-cube speaker. The raw acoustic responses were measured by the microphones in the subject's ear canals. The signal from the microphones was then sent thorough a microphone amplifier and into an A/D converter in the TDT system. These raw signals were analyzed to find the HRTFs for the listener and room.

The maximum sound source level that would not cause clipping for each distance and angle of interest was used. Using a level that is too high results in clipping while using a lower level decreases the signal-to-noise ratio and hence degrades the measured HRTF. Trial and error was used to determine what level was appropriate for each position being measured. KEMAR was placed in the chair in the center of the room and the microphones were inserted into his ears. Sounds were presented at several levels for each sound source position used in the study until the level at which clipping began to occur was empirically determined. Because the clipping level does vary slightly from subject–to-subject, a level slightly less (3-5 dB) than the clipping level found for KEMAR was used for each tested position.

2.2 Measurement Techniques

Initially, the Golay code HRTF measurement technique was considered for this study, but after examining the advantages and disadvantages of various techniques it was decided to use binary maximum-length sequences (MLS) instead.

The Golay measurement technique presents a pair of matched acoustic signals known as Golay codes to the ears. Golay codes have the property that the sum of the autocorrelation functions for each code is only nonzero at zero lag [9]. The Golay codes are a complementary pair of radix-two digital sequences that are generated by a simple recursive relation. The system's impulse response is determined by cross-correlating the responses to each of the codes in the pair with the codes themselves and summing the result. Due to the properties of the Golay codes, the final result of this operation yields an estimate of the impulse response.

Taking measurements on humans presents a challenge because the measurement method must be able to handle small variations from a time invariant system since humans are not capable of remaining perfectly motionless during measurements and HRTFs are strongly dependent on head position. The effects of time variance on system responses obtained using Golay codes and MLSs have been examined by Zahorik [9]. The results of Zahorik's research shows that the Golay code method can have particularly damaging artifacts in the impulse response with even very small subject movements. Specifically, if the system being measured is time variant, the Golay code responses are no longer complementary. The sum of the responses to the codes no longer cancels perfectly to yield the true impulse response, but rather has (incorrect) non-zero elements near a lag of the L/2 (L = the length of the Golay code). These non-zero elements are the artifacts introduced by the time variant system. As the length of the Golay codes or the number of averages taken increases, the time lapse between the presentation of two codes increases, therefore increasing the chance that the subject has moved during the measurement and increasing the probably that artifacts contaminated the impulse response. When measuring HRTFs in a reverberant space, longer codes are required to capture the entire room response and avoid time aliasing and more averages may be required to increase the signal-to-noise ratio. This again increases the chance of artifacts in the impulse response. For this study, human subjects are being used and effects of reverberation are being studied, so the artifacts introduced by time variant systems when using Golay codes could cause significant data contamination. In contrast, according to the analysis by Zahorik [9] the MLS method is more robust under these conditions.

Binary MLSs are periodic pseudorandom sequences of +1's and -1's with period length $L = 2^n-1$, where n is an integer, generated by n-stage shift registers. To obtain the desired system frequency response, the system response to the sequences has to be cross-correlated with the same MLS that was used for exciting the system. MLSs' most important property is that their Fourier transform has the same magnitude for all frequency components [10].

2.3 Measurement Procedure

The procedure used to record the HRTFs was to first read a pre-generated MLS of 32767 (2^{15}) bytes containing -1's and +1's into MATLAB. Two sequences were concatenated. This concatenated sequence was then played through a BOSE cube speaker and the response was recorded by Knowles Electret microphones positioned at the entrance to the subjects' blocked ear canals. The concatenated sequence was played ten times and the results were averaged. The response to the second half of the sequence was then cross-correlated with the original sequence to obtain the impulse response. The resulting impulse response was 750 ms long and was sampled at 44.1 kHz. This contains all the reverberation of the room, as well as the direct-sound HRTF.

In echoic space, the measured impulse responses contain the transfer characteristics of the head, the room, and the sound delivery and measurement systems. Therefore, the reverberation, the echoes, and the characteristics of the sound delivery measurement system must be removed to derive estimates of the "anechoic" HRTFs. In order to remove the effects of the delivery and measurement system, the MLS technique is used to measure the impulse response of the left and right microphones in free space in the center of the room being used for experiments. These recordings provide insight into the gain characteristic of the presentation and recording system.

After using temporal windowing to remove the effects of reverberation, the portion of the HRTF contains the transfer characteristics of both the listener's head and the sound delivery and measurement system. Presenting the MLSs to the left and right microphones in the free field and time windowing the resulting impulse responses (as described above) provides a measurement of the impulse response of the delivery and measurement system. The inverse of this transfer characteristic has been calculated from these response measurements separately for the left and right microphone. Post-compensation for the sound delivery and recording system is accomplished by applying the inverse filters to the impulse responses from the raw MLSs. While this approach can, in theory, remove the effects of the measurement system, in practice it can also distort the phase response of the transfer function. In addition, the magnitude response of the measurement system is nearly flat, varying smoothly over approximately 10 dB between 300Hz and 20 kHz. For the analysis performed in this study, post-compensation was not used because of the importance of keeping the room effects and not distorting the room response and phase information when analyzing the effects reverberation on HRTFs.

2.4 Verification of Measurement Technique

Measurements were taken with the left and right microphones in free space in the center of the room used for the HRTF measurements. The only correction made to these measurements was to compensate for the difference in sensitivity of the left and right microphones. This correction entailed multiplying the raw response of the right microphone by a constant (1.22). Measurements in the center of the room indicated that the first-arriving echo is from the floor and that this echo will arrive no earlier than 5 ms after the direct sound (for all source locations tested). Measurement taken from the ear canals of KEMAR confirmed that the amplitude of the direct sound is negligible 5 ms after the arrival of the direct sound (see *Figure 1*). Therefore, temporal windowing can be used to remove the effects of reverberation. Time waveforms were examined for each subject to determine the appropriate time window for their results. A time window of 9.5 ms was found to be adequate for all subjects and all sound source positions because the first echo arrived between 9.75 ms and 11 ms, depending on source position and subject.

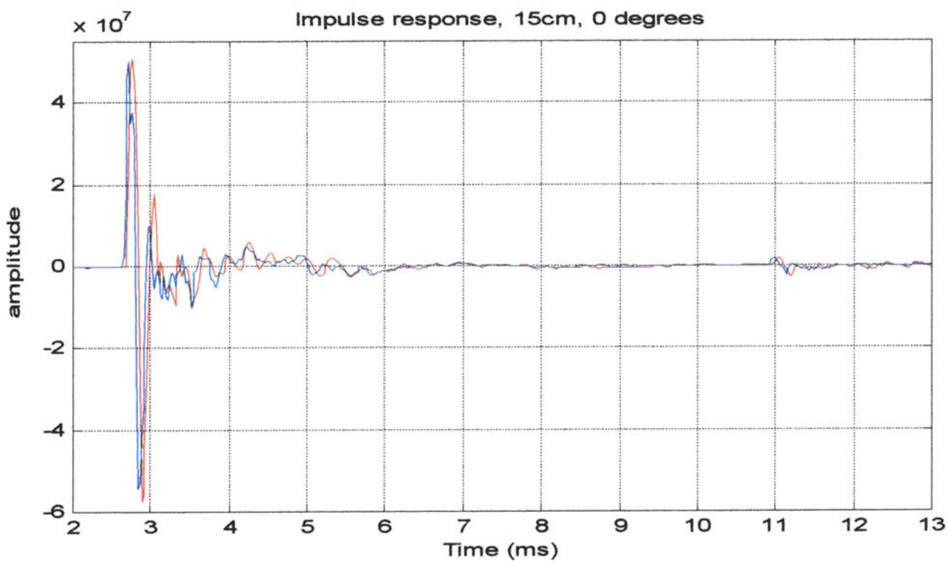


Figure 1: Impulse response for KEMAR with the sound source at 15 cm and 0 degrees from center of head (blue is left ear, red is right ear). The large initial impulse corresponds to the "direct sound" impulse response. Note first reflection at approximately 11ms (roughly 8ms after the onset of the direct sound impulse response).

Figure 2 shows the magnitude of the DFT of the measurements taken with the microphones in free space in the center of the room for 3 sound source levels (as well as with the source off, to show the noise floor). *Figure 3* shows the same measurement

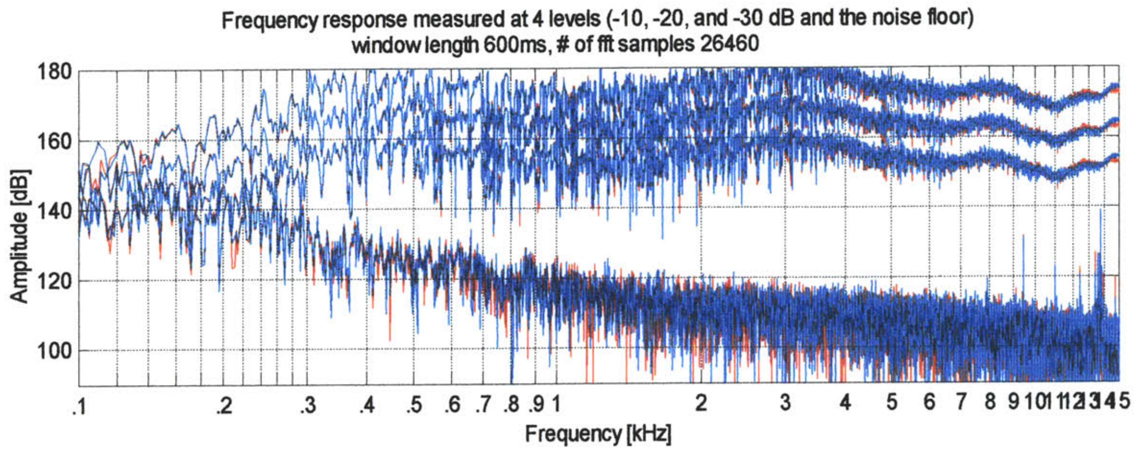


Figure 2: Frequency response measured in the center of the room with sound source 1m from microphones. Red is right microphone response and blue is left microphone response. Relative sound source levels top to bottom are –10 dB, -20 dB, -30 dB, the noise floor.

windowed with a time window at 10.5 ms to remove the reverberation. In both *Figure 2* and *Figure 3*, it is evident the measured response magnitude grows linearly with presented source amplitude for frequencies above 200 Hz. The noise floor (the bottom-most curve on each figure) at 300 Hz is 25 dB below the measurement at -30 dB. Above 3 kHz, the noise floor is 50 dB below the measurement at –30 dB. In other words, above

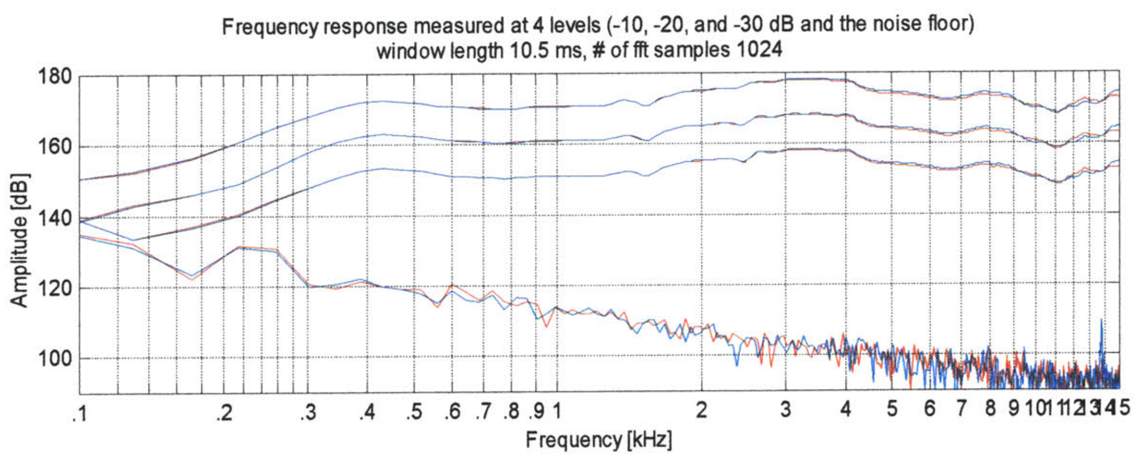


Figure 3: Frequency response measured in the center of the room with sound source 1m from microphones. Time window of 10.5 ms used to remove reverberation. Red is right microphone response and blue is left microphone response. Relative sound source levels top to bottom are –10 dB, -20 dB, -30 dB, and the noise floor.

200 Hz, the noise floor amplitude is insignificant compared to the amplitude of the HRTF measurements and above 300 Hz it is negligible.

The free-field frequency response also varies inversely with source distance, as expected. *Figure 4* shows the frequency response of the direct sound measured with constant sound source level and the source placed at several distances between 1 m and 15 cm from the microphones. The response amplitude is proportional to 1/distance for frequencies above 600 Hz. For frequencies less than 600 Hz, there still appears to be a linear relationship, but the results look noisier because there are fewer points plotted in the FFT (on this log-frequency scale). The frequency response is also quite flat above 300 Hz.

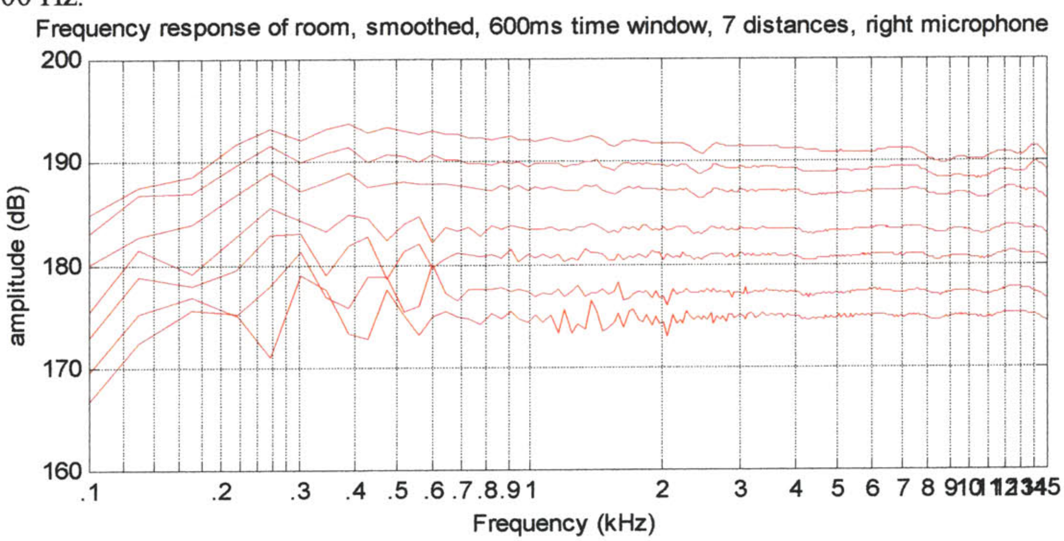


Figure 4: Frequency response of right microphone in free space in center of the room. Top to bottom distances measured at 15 cm, 19 cm, 25 cm, 38 cm, 50 cm, 75 cm, 1 m. All distances are measured at the same sound source level. Responses for the left microphone are similar.

2.0 Results and Discussion

3.1 Pseudo Anechoic KEMAR HRTFs

First we examine the pseudo-anechoic HRTFs for sources near a listener. These HRTFs are derived by taking measurements when the listeners are located in the center of the room and windowing out the reverberation. HRTFs were measured for 15 human subjects and for KEMAR using the methods and set-up previously described. In this and most subsequent sections, a set of six locations (relative to the listener) are reported. These positions consist of two distances (15 cm and 1m) and 3 azimuthal angles (0, 45, and 90 degrees) in the horizontal plane (elevation 0 degrees).

3.1.1 Magnitude Responses

The pseudo-anechoic HRTF magnitude responses measured on KEMAR are shown in *Figure 5*. At zero degrees, the magnitudes at the left and right ears are the same up to frequencies of 6 kHz. Above 6 kHz, there are small differences in the locations and depths of the notches and peaks in the transfer functions for the left and

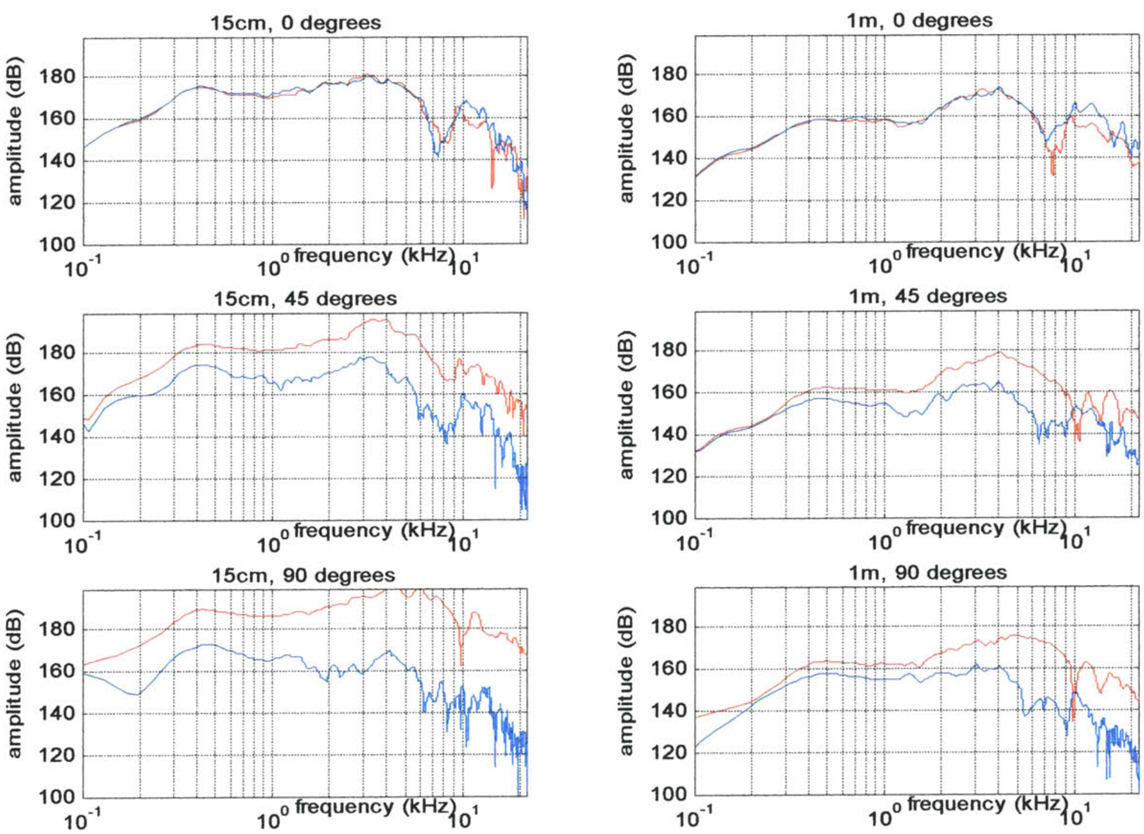


Figure 5: Magnitude response for KEMAR at right ear (red) and left ear (blue) for 6 sound source positions.

right ears, probably, due in-part to the asymmetries of KEMAR's pinna. The magnitude at the left ear is slightly greater than that at the left ear, particularly above 6 kHz. This may also be caused by a misplacement of the sound source slightly to the left of 0 degrees. The magnitude at the right ear increases when the sound source is moved from 0 to 45 degrees and again when the sound source is moved from 45 to 90 degrees. As the angle increases, the sound source moves closer to the right ear causing, the observed increase in magnitude response. At the same time, the response at the left ear decreases as the sound source moves from 0 to 45 and 45 to 90 degrees. As angle increases, the distance to the left ear increases and the head interferes with the transmission of sound to the far side of the head (i.e. there is an acoustic "head shadow"), causing the sound source magnitude at the left ear to decrease. The response at the right ear also increases as the sound source moves toward the center of the head (from 1m to 15cm). While the response at the left ear does increase, it does not increase as much as the response at the right ear. This slight increase at the left ear also results from the sound source moving closer to the left ear. The increase is smaller than that at the right ear because the relative change in distance from the source to the right ear is substantially larger than the relative change in distance to the left ear and because the "head shadow" becomes larger as the source moves closer to the head [12].

3.1.2 Interaural Level Differences

The ILD calculated from the measured KEMAR data for the at 6 sound source positions is shown in *Figure 6*. Below 6 kHz, the ILD at 0 degrees is approximately 0. The distance to the left and right ears is essentially the same for a source at 0 degrees, so the magnitude of the right and left frequency responses should be equal, yielding an ILD of zero. In the previous section, it was observed that the left and right magnitude responses were not exactly equal above 6 kHz; as a result the ILD is not equal to zero above 6 kHz. As discussed in the previous section, this is due to differences in the high frequency peaks and notches caused by differences in left and right pinna shapes and (possibly) a mis-placement of the speaker slightly left of center.

The ILD increases when the sound source is moved from 0 to 45 and from 45 to 90 degrees. This is expected because the right magnitude response increases and the left magnitude response decreases due to the distance from the sound source to the right ear decreasing and the distance to the left ear increasing. The ILD decreases when moving the sound source from 15 cm to 1 m. This occurs because a source at 1 m is nearly equidistant from the 2 ears (on a log distance scale) so the only significant ILD is caused by head shadow. Conversely, for a nearby source (at 15 cm), the ratio of the

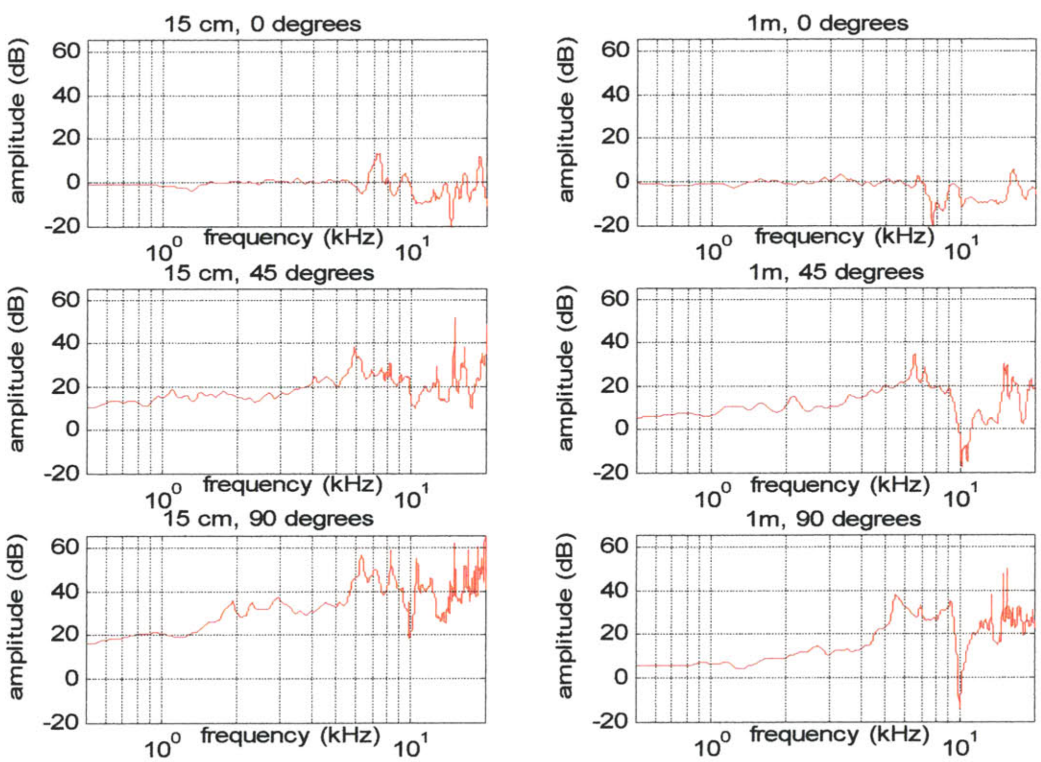


Figure 6: Interaural level difference for KEMAR at 6 sound source positions.

distances to the right and left ears is much less than 1, resulting in a large ILD that is essentially added to the head shadow [12]. The high frequency notches and peaks in the ILD are due to the high frequency notches and peaks in the right and left magnitude responses being misaligned. As seen in the previous section, the magnitude response notches and peaks do not occur at the same frequency, nor do they necessarily have the same magnitude at the left and right ears due to differences in pinna shape. This leads to ILDs that vary dramatically in magnitude and even sign at high frequencies.

3.1.3 Interaural Time and Phase Differences

The interaural time (ITD) and phase difference (IPD) are related such that the ITD is equal to IPD/(2π*frequency). The IPD is the phase difference between the right and left ears, calculated by dividing the angle of the FFT of the right ear impulse response by the angle of the FFT of the left ear impulse response. Physiologically, the brain computes IPD, which is ambiguous in ITD by whole cycle shifts (e.g., 0 ITD is represented the same way as $2\pi f$). A pure ITD that is constant across frequency leads to a purely linear IPD across frequency, except for discontinuities due to ambiguity. Thus phase ambiguity is resolved when multiple frequencies are present in the stimulus. To estimate true ITD, one simply looks for the constant line ITD trajectory as a function of frequency. The addition of a factor of +/- k*2 (where k is an integer) is necessary to compensate for jumps in phase that occur when the phase is greater than 2π or less than -2π, which cause discontinuities in the IPD waveform and to produce a consistent ITD value across frequency from inherently ambiguous IPD information.

Figure 7 shows the ITD calculated for KEMAR for 6 sound source positions using pseudo-anechoic HRTFs. Each color represents one integer value of k ranging from –5 to 5. At zero degrees, only k equal to 0 is necessary to obtain linear phase up to 4 kHz. The ITD value at which the linear IPD occurs at zero degrees is 0 s, as expected. At 45 degrees, k equal to 0 and 1 are necessary for a linear phase, and at 90 degrees, k equal to 0, 1, 2, and 3 are necessary to obtain linear phase up to 4 kHz. At 45 degrees, the ITD value at which linear IPD occurs is approximately 0.35 ms while at 90 degrees it is approximately 0.7 ms. The ITD is a function of sound source position. If the sound source is placed at 0 degrees, directly in front of the subject, the ITD is expected to be zero because the left and right ears are equal distance from the sound source and therefore the sound should reach both ears at the same time. The results shown in *Figure*

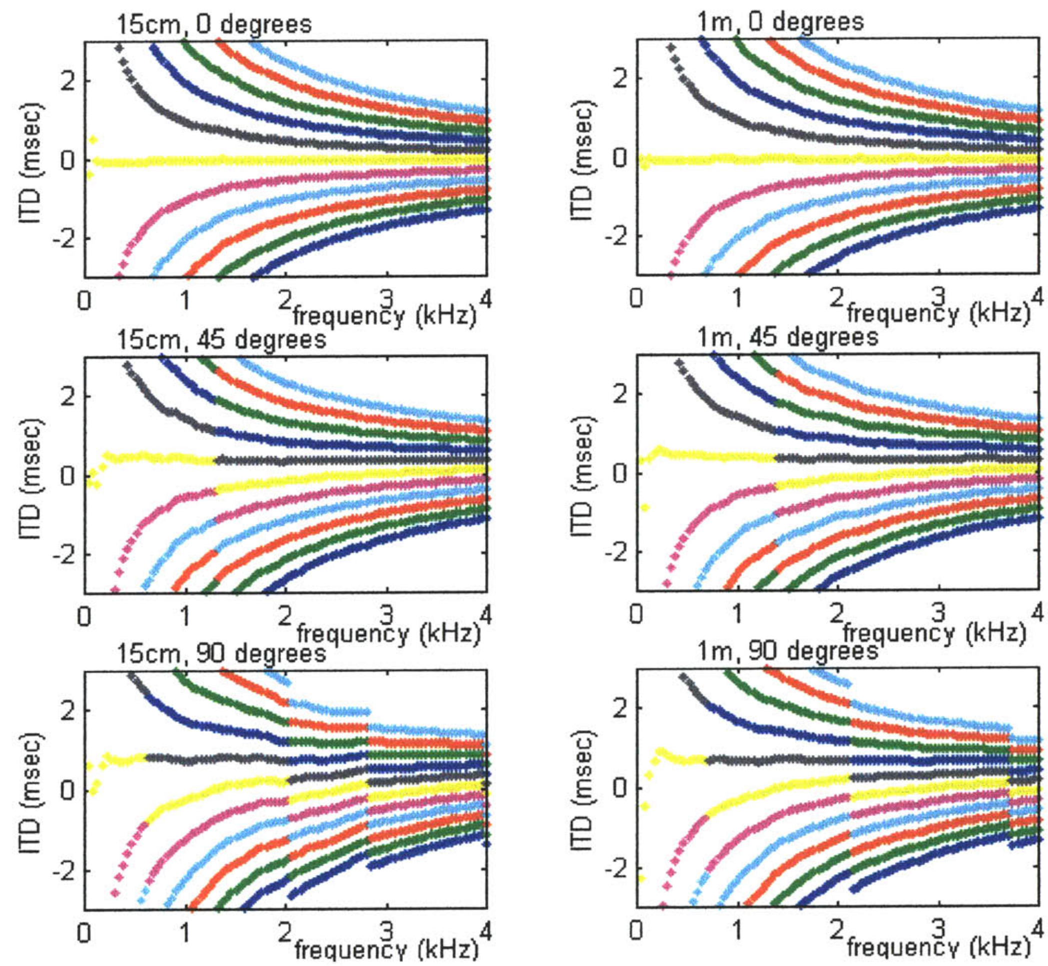


Figure 7: Interaural time difference for KEMAR at 6 sound source positions.

7 are consistent with this expectation, showing the mean value of the constant ITD curve is zero at zero degrees. As the sound source angle increases, the sound source moves towards the right ear and away from the left ear, the ITD is expected to increase because the difference in the sound source distance to the right ear versus that to the left ear increases. This difference reaches a maximum at 90 degrees and therefore the ITD reaches a maximum at 90 degrees. Given the average distance between the ears, this maximum expected ITD is approximately 1ms. Distance should not affect ITD greatly because the difference in path length to the ears does not change dramatically with distance. Again the data shown in *Figure 7* agrees with expectations: the mean of the constant phase trajectory increases with increasing angle but is roughly independent of distance.

3.2 Pseudo-Anechoic HRTFs across Human Subjects

Across subjects, the HRTFs have the same basic shape for each position, but the frequency and depth of high frequency peaks and notches varies. This pattern of individual differences is similar to that observed for far-field HRTFs [10]. Anecdotally, these individual differences appear larger than the measurement error in the HRTFs. HRTFs below 300 Hz are not reliable since at these frequencies, the noise floor becomes significant. Large trial-to-trial and subject-to-subject variations are observed at these low frequencies.

3.2.1 Magnitude Responses

Appendix 1 shows the pseudo-anechoic magnitude responses for the 15 human subjects participating in this study. The overall shape and magnitude for the left and right magnitude responses at a given sound source position and frequency are very similar across all subjects (including KEMAR). Although the overall shape is the same, at high frequencies, the peaks and notches vary in magnitude and in frequency location. Some subjects appear to have more symmetrical heads than others, as demonstrated by the fact that the left and right magnitude response are essentially identical at zero degrees.

3.2.2 Interaural Level Differences across subjects

ILD is similar for KEMAR and the human subjects. As can be seen from the magnitude plots in *Appendix I*, the difference between the magnitude response at the right and left ear increases as angle increases and decreases as distance increases. For several subjects, the left and right magnitude responses have larger differences at zero degrees, meaning a larger ILD, than for KEMAR, probably because human subjects tend to be more asymmetrical than KEMAR.

3.2.3 Interaural Time Differences across Subjects

The ITDs calculated from the HRTFs measured from 15 human subjects and KEMAR located in the center of the room with 6 sound source positions are shown in *Figure 8*. These ITDs were calculated by taking the cross-correlation of the pseudo-anechoic impulse responses. Since the delay is expected to be less than 1 ms in magnitude, only peaks at delays in this range were examined. The peak resulting from the cross correlation corresponds to the delay between the peak of the direct sound reaching the right and left ears. The results shown in *Figure 8* are consistent with the theory discussed above. The small negative ITDs calculated at 0 degrees indicate the distance to the left and right ears were not quite equal. This could have been caused by the subject's head being slightly asymmetrical or the speaker not being perfectly centered in front of the subject. Because the ITD is consistently negative at zero degrees, even with KEMAR (shown by the eighth bar from the left) who has a nearly perfectly symmetric head, it is most likely that the speaker was always placed slightly left of center with respect to the subjects' center.

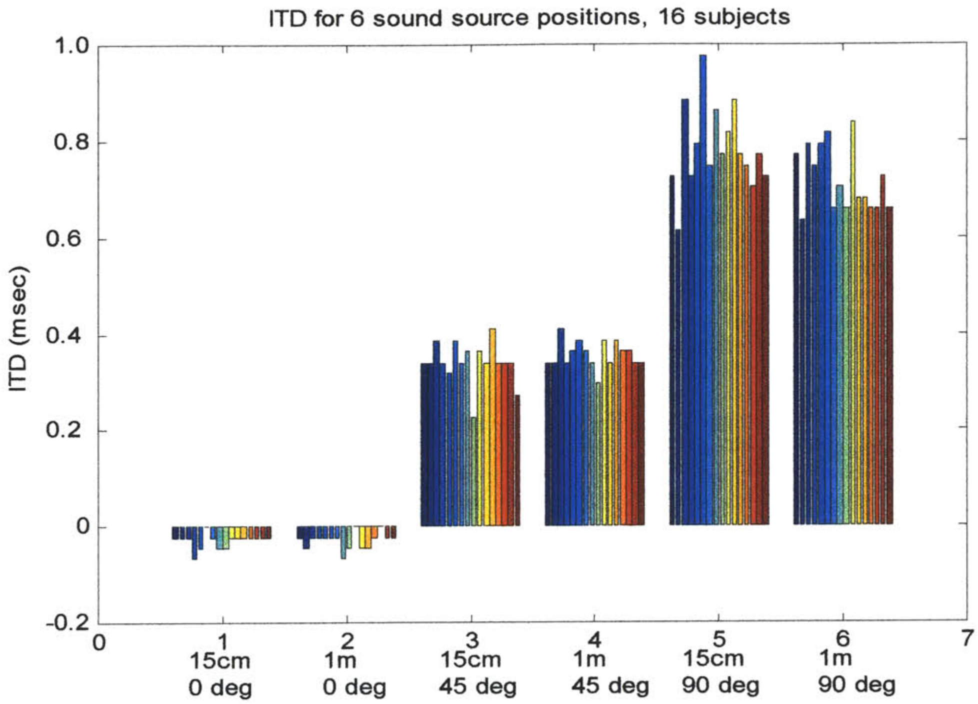


Figure 8: ITD for 15 human subjects and KEMAR placed in the center of the room for 6 sound source positions. Each color represents a different subject.

The ITD for a source at 45 degrees is approximately 0.35 ms for both 15 cm and 1 m source distances. This is expected, since the difference in the pathlength from the sound source to each ear is nearly the same at both distances (distance from the head really does not matter - only difference in pathlength) [11]. At 90 degrees, the ITD is approximately 0.7 ms, double that at 45 degrees. This is expected because the difference in the distance from the sound source to left and right ears approximately doubles. The maximum ITD calculated at 90 degrees is just under 1 msec. The larger variation in ITD seen at 90 degrees is due to varying head sizes across subjects, leading to different distances between ears. Such differences in head size have the largest effect on ITD when the speaker is placed at a location yielding the maximum difference in distance between the speaker and each ear (90 degrees). The data in *Figure 7* (the ITD as a function of frequency for KEMAR) and that in *Figure 8* (the mean ITD for the human subjects at different sound source positions) agrees, showing an average ITD of approximately 0 ms at 0 degrees, 0.35 ms at 45 degrees and 0.7 ms at 90 degrees.

3.3 Reverberant KEMAR HRTFs

As shown in *Figure 1*, a time window at 9.5 ms is sufficient to remove reverberation without removing the direct sound. Reverberant HRTFs were obtained from the same data as in the pseudo-anechoic case by using a 600 ms time window instead of a 9.5 ms time window. This time window is sufficiently long to include all measured reverberation that is above the noise floor.

3.3.1 Magnitude Responses

Figure 9 shows the reverberant magnitude response at 6 source locations for KEMAR. The magnitude response has the same general shape as in the pseudo-anechoic case but shows variations around the mean, particularly at the left ear where reverberant energy is greater and direct energy is weaker. As in the anechoic case, the magnitude responses of the left and right ears have the same mean at zero degrees, except at high frequencies where pinna effects increase and small asymmetries can have a large effect. The same acoustical effects seen in the anechoic case cause the right magnitude response to increase and the left magnitude response to decrease as angle increases and as distance decreases.

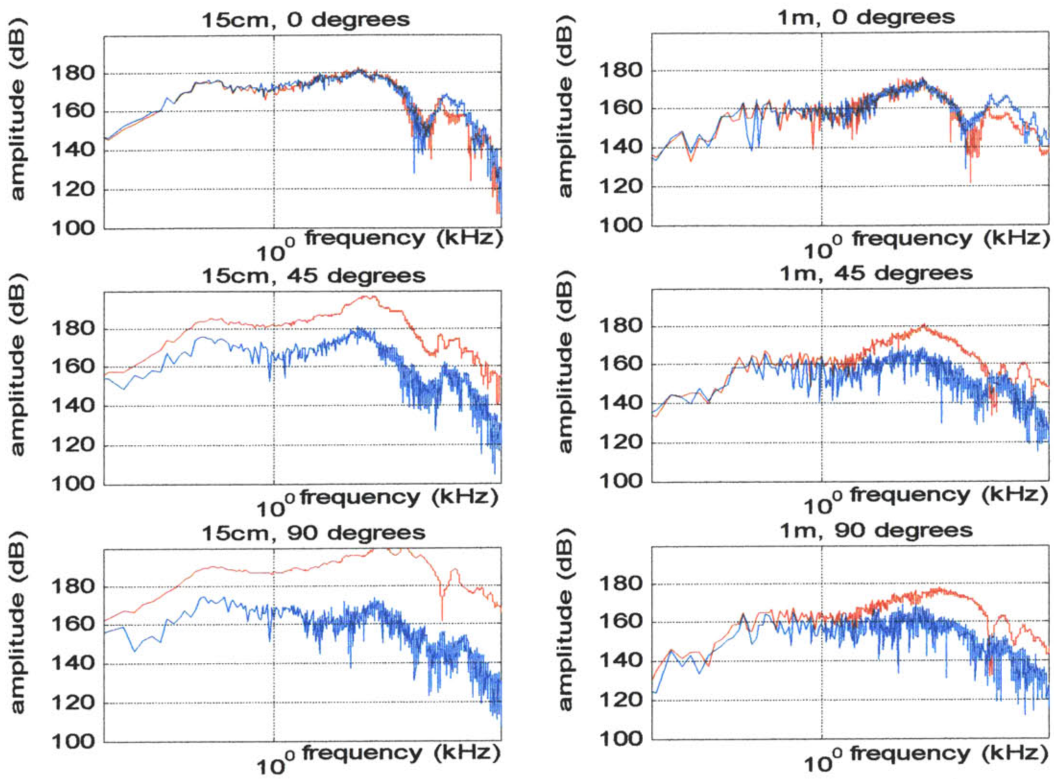


Figure 9: Magnitude of the frequency response with 600 ms time window for KEMAR at 6 positions. Red is the right ear and blue is the left ear.

3.3.2 Interaural Level Differences

The reverberant ILDs are shown in *Figure 10*. Just as in the anechoic case, the ILD has a mean of zero at zero degrees and increases with increasing angle and with decreasing distance. The high frequency peaks and notches are similar in location and depth to those in the anechoic condition, but are somewhat obscured by the large high frequency fluctuations around the mean ILD. There is much more variation around the mean in the reverberant case. For all positions, the variations from the mean increase with increasing frequency.

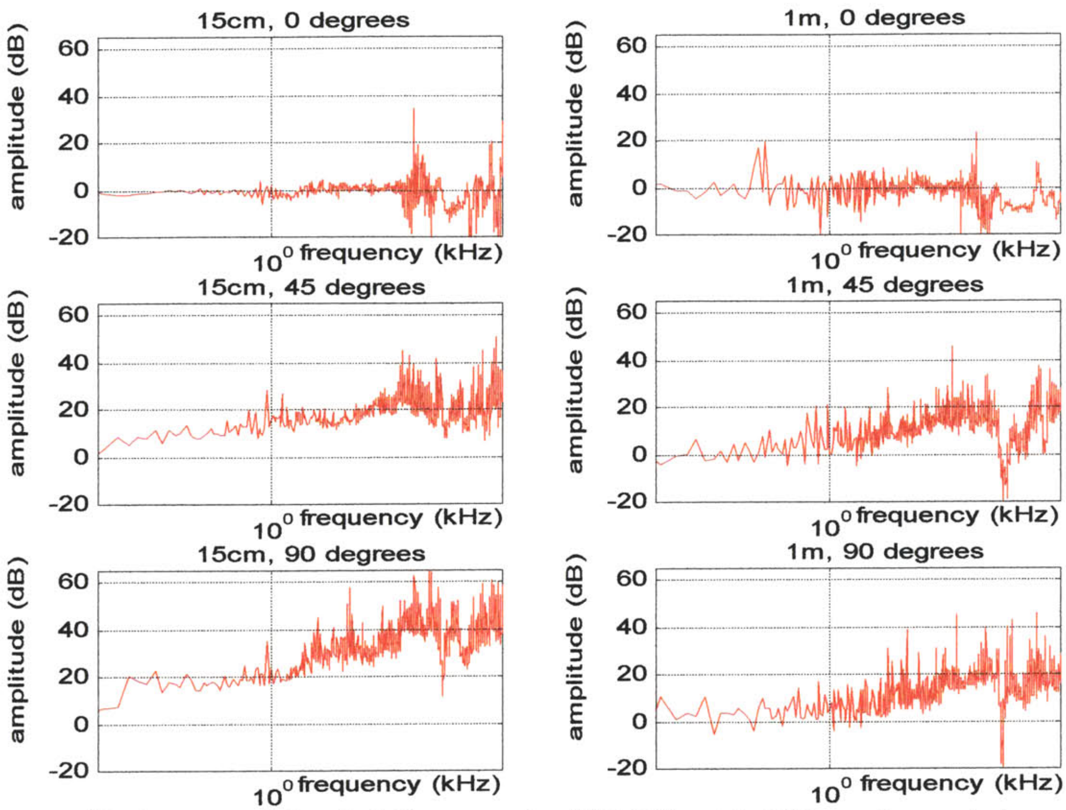


Figure 10: Interaural level differences for KEMAR with 600 ms time window and 6 sound source positions

3.3.3 Interaural Time and Phase Differences

As mentioned in section 3.1.3, the IPD can be calculated from the ITD. The reverberant ITD is shown in *Figure 11*. The reverberant ITD and IPD appear to have the same shape and nearly the same mean values as in the anechoic case (*Figure 7*); however, there is slightly more variation around the mean in the reverberant case. The ITD at 0 degrees averages 0 ms, at 45 degrees averages 0.37 ms, and at 90 degrees averages 0.69 ms at 1m and 0.82 ms at 15cm (see section 3.5.4.3 for summary of ITD mean and variance calculated from *Figures 7 and 11*).

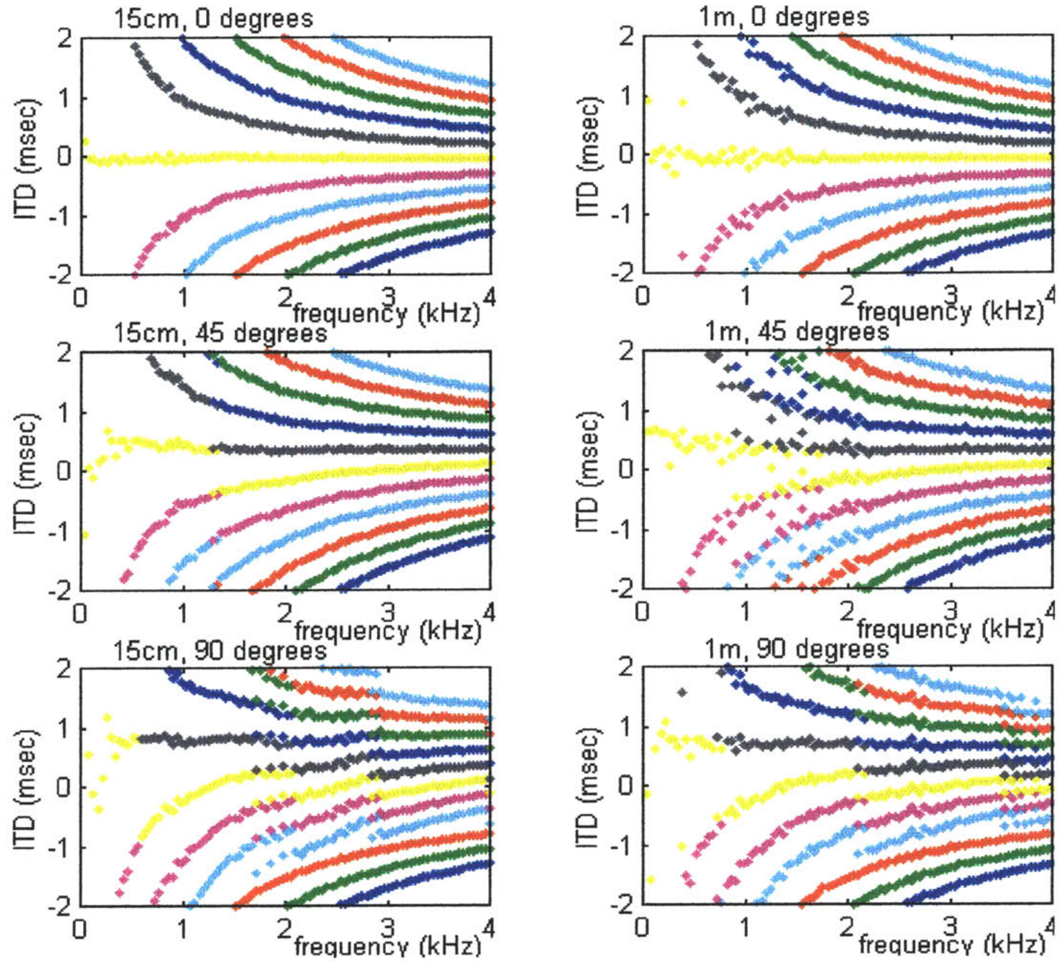


Figure 11: Reverberant interaural time difference for KEMAR with 6 sound source positions.

3.4 Effects of Reverberation on HRTFs

The time-windowed (9.5 ms time window) HRTF data was compared to reverberant data (600 ms time window) to examine the effects of reverberation on HRTFs. This comparison leads to the conclusion that reverberation essentially adds noise to the HRTF data. To facilitate calculating numerical values for comparing the reverberant and anechoic data, the reverberant data was smoothed by averaging each FFT point with the values from the four adjacent frequencies (i.e., the FFT was convolved with a flat, 5-point smoothing kernel). The overall shape of the anechoic and reverberant

HRTF data could then be compared after removing the large frequency-to-frequency variations introduced by the reverberation. Such processing enables quantitative comparisons between the anechoic and reverberant mean results, as well as of the magnitude of the frequency-by-frequency variation present in the reverberant results.

3.4.1 Magnitude Responses

Figure 12a shows the raw magnitude responses with (red) and without (blue) reverberation in the right ear while *Figure 12b* shows the same data for the left ear. The magnitude of the right anechoic and smoothed reverberant responses are roughly equal. The reverberation does add frequency-to-frequency variations to the right magnitude response, but does not change its mean. The amplitude of this "noise" is larger when the sound source is at 1 m than at 15 cm, consistent with the idea that the nearer source HRTF is more completely dominated by the direct sound, whereas the farther source HRTF has relatively more reverberant energy. At 15 cm and 45 and 90 degrees, the effect of reverberation at the right ear is barely noticeable.

At the left ear, for sound source positions at zero degrees, the mean amplitude of the frequency response is the same with and without reverberation, just as at the right ear. At zero degrees, the reverberation only adds noise to the magnitude response. In contrast, at 45 and 90 degrees, the magnitude of the left anechoic frequency response is roughly equivalent to left echoic responses for frequencies less than 5 kHz. For frequencies above 5 kHz, the difference between the left anechoic and echoic frequency responses increases as the sound source distance increases and as the sound source angle increases from 0 to 90.

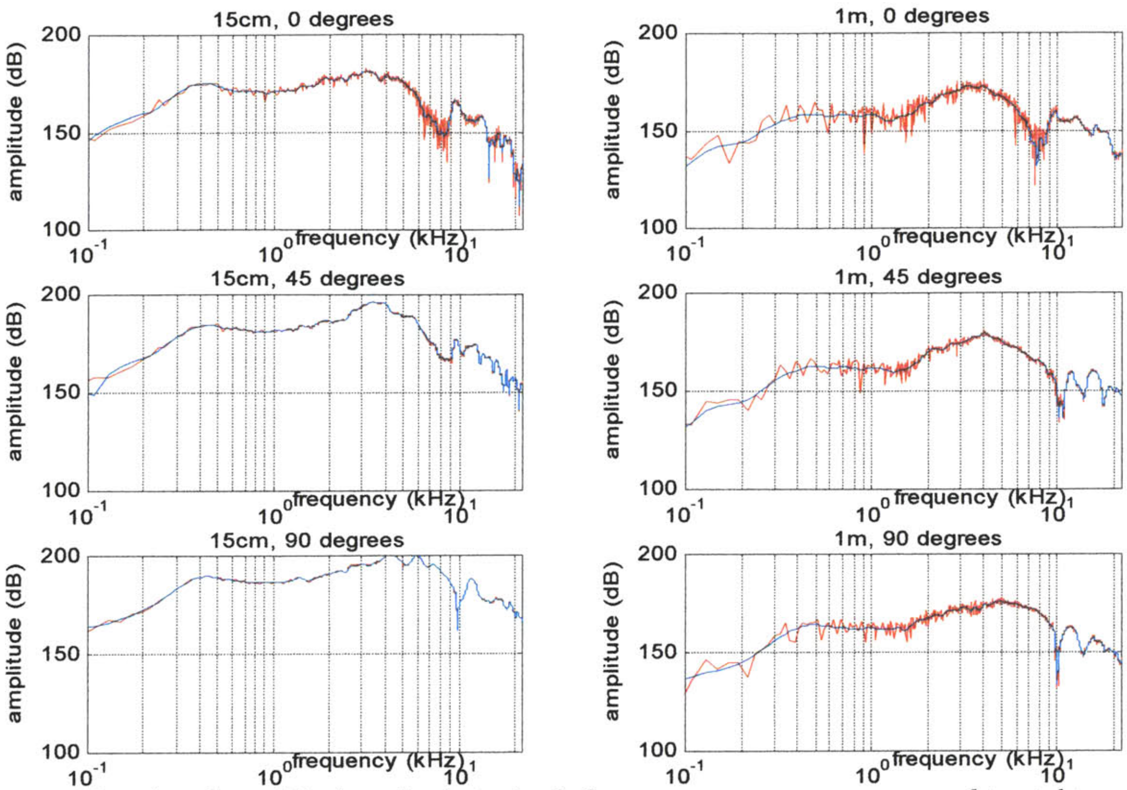


Figure 12a: Anechoic (blue) and echoic (red) frequency responses measured in right ear with KEMAR in the center of the room for 6 sound source positions.

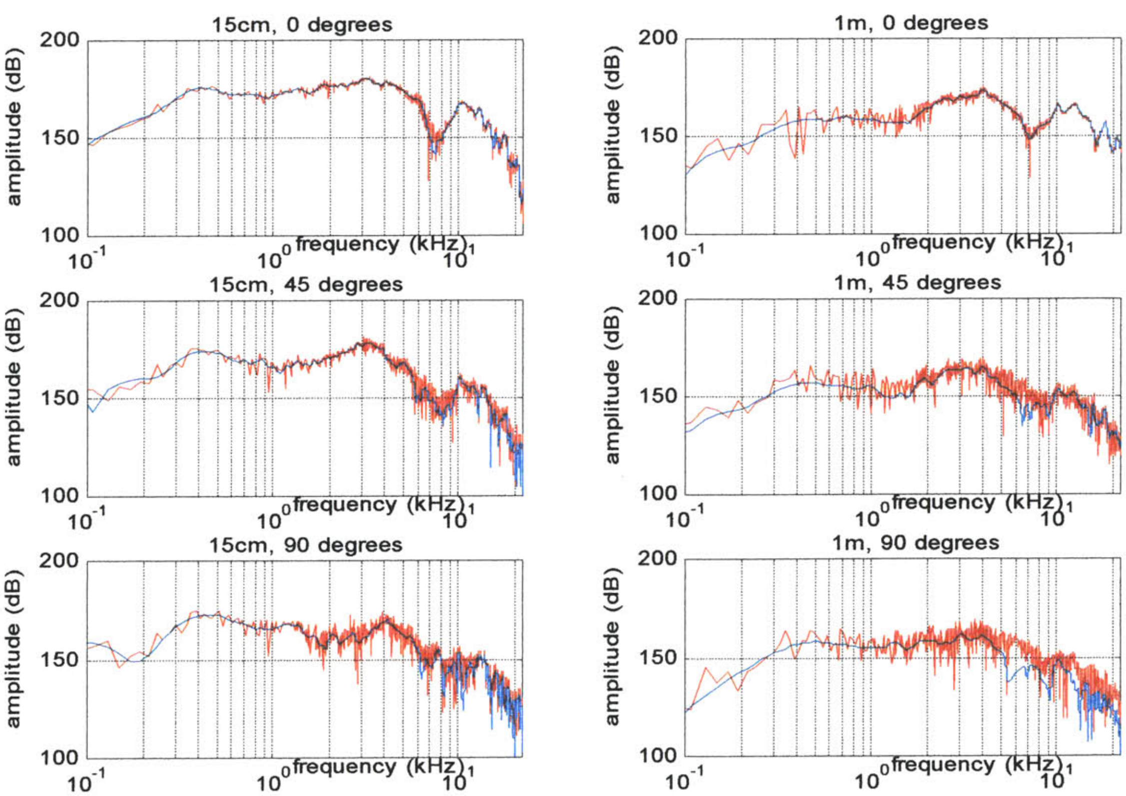


Figure 12b: Same as 12a, but at left ear.

These results can be understood after considering acoustic effects. As the distance from the head to the sound source increases and the angle increases, the sound source moves further from the left ear, causing the magnitude of the direct sound at the left ear to decrease. In particular, as the source moves farther from the head or increases in laterality, the reverberant response energy (which tends to be omni-directional and has a flatter spectral shape) tends to fill in the magnitude response and flatten out any notches in the anechoic (direct-sound) HRTF. For the same sound source movement, the magnitude of the reverberation does not change very much (the reverberation is roughly independent of source location). At the right ear, there is very little observable change in the reverberant magnitude with changes in the sound source location because the magnitude of the direct sound dominates the right ear response; in fact, the direct sound magnitude increases as the angle approaches 90. Thus, the relative importance of the reverberation decreases as the source moves to the right side of the head.

The root mean square (RMS) of the differences between the anechoic and smoothed reverberant magnitude responses at each frequency were computed to find the RMS difference caused by the reverberation. A summary of the RMS difference of the smoothed echoic right magnitude response from the anechoic right magnitude response for 15 subjects and KEMAR at 6 sound source positions is shown in *Figure 13a.* The RMS difference is relatively small for all sound source positions at the right ear. The fact that the RMS difference is small at the right ear is consistent with the fact that the reverberation has only a small impact on the total signal at the right ear (e.g., see *Figure 12a)*. For the right-ear signal, the RMS difference is largest at zero degrees. Variations

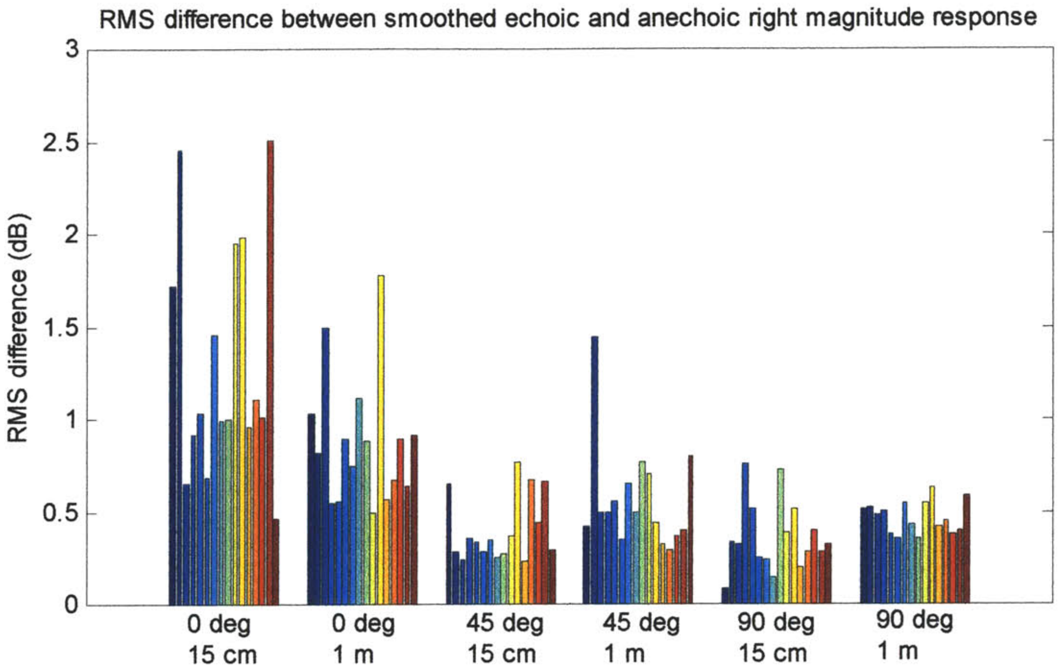


Figure 13a: RMS difference between the smoothed anechoic magnitude response and the echoic magnitude response for the right ear signal of 15 subjects and KEMAR at 6 sound source locations.

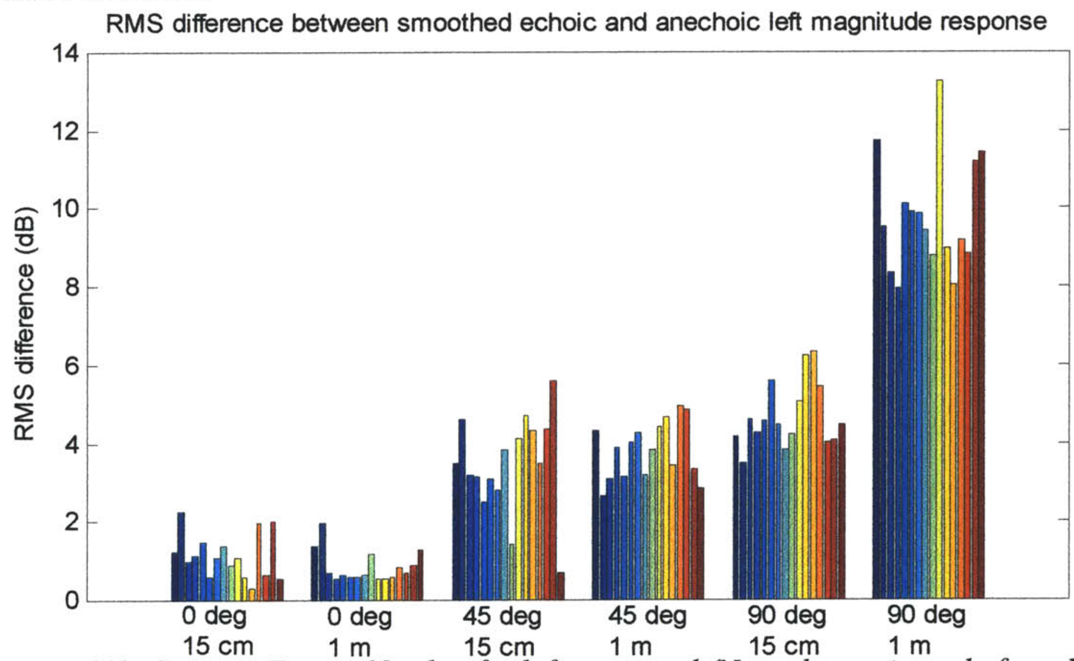


Figure 13b: Same as Figure 13a, but for left ear signal (Note change in scale from 13a).

across subjects in the right ear RMS difference are also greatest at zero degrees. The larger RMS difference at zero degrees is consistent with the observation that the reverberation causes the largest variations around the anechoic HRTF are largest at zero degrees in *Figure 12a*. For sources at 45-90 degrees, the RMS difference is generally greater at 1 m than 15 cm, which is also consistent with the larger variations around the mean shown at these sound source positions in *Figure 12a*.

The RMS difference between the smoothed echoic magnitude response and the anechoic magnitude response for the left ear signals of the 15 subjects and KEMAR are shown in *Figure 13b*. On average, the RMS difference is much larger at the left ear than at the right. The RMS difference at zero degrees at the left ear is similar to that at the right ear, ranging between 0.4 - 2.5 dB for all subjects. This is expected, since at zero degrees, the direct sound source magnitude and the reverberant energy magnitude at the left and right ears should be comparable. The similarity can also be seen in the plots in Figures 12a and 12b at zero degrees; the magnitude of the variations of echoic magnitude responses around the anechoic magnitude response at zero degrees are similar for the left and right ears. The RMS difference increases as angle increases and as distance increases, consistent with the fact that the direct sound level decreases at the left ear with angle or distance increases.

3.4.2 Interaural Level Difference

Comparing the reverberant ILD to the anechoic ILD shows that, for most positions and frequencies, the anechoic ILD looks like a smoothed version of the echoic ILD, although, once again, the similarity depends on the relative level of the direct and reverberant energy. *Figure 14* shows the anechoic ILD (red) plotted on top of the smoothed echoic ILD (blue). For the measurements at zero degrees, the anechoic ILD passes through the mean of the points in the echoic ILD. Since one might expect the reverberation to have roughly the same effect (increasing the magnitude slightly) at both

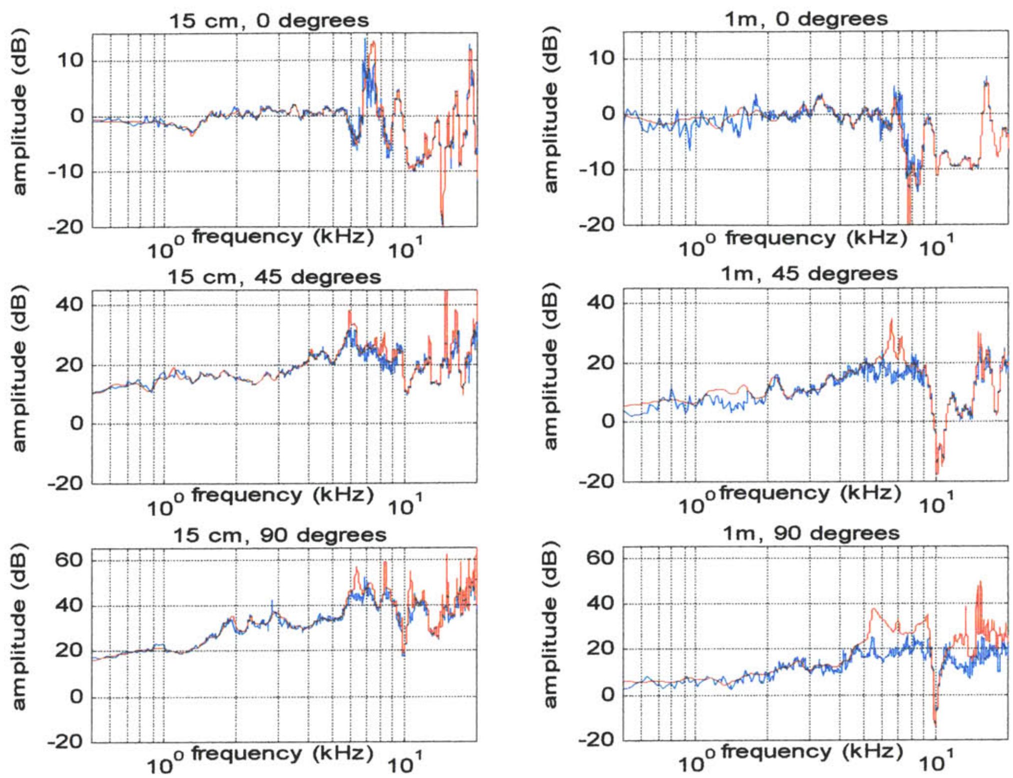


Figure 14: Anechoic ILD (red) and echoic ILD (blue) measured with KEMAR placed in the center of the room for 6 different sound source positions.

the left and right ears for a sound source at zero degrees azimuth, the net change should be to decrease the magnitude of the ILD towards zero. As the sound source angle increases, reverberation has little effect on the magnitude of the frequency response at the right ear because the direct sound energy is so great, but causes a noticeable increase the magnitude of the frequency response at the left ear, where the direct sound energy is small (see *Figures 12a and 12b)*. As a result, the ratio of the sound magnitude at the right ear to that at the left ear (the ILD) tends to decrease with reverberation. Even in cases where the right ear direct sound energy is small and the reverberation increases the right ear level, it will cause a similar increase in the left ear level, thus tending to reduce the total ILD. The magnitude of the echoic ILD is slightly less than that of the anechoic ILD when the sound source is placed at 45 degrees for frequencies over 5 kHz. This difference is larger at 1 m then it is at 15 cm. In particular, large peaks or valleys in the ILD are smoothed out for these positions (that is, the smoothed echoic ILD tends towards smaller magnitudes than the anechoic ILD). There is a more significant difference in the high frequency ILD at 90 degrees; again, the difference is larger at 1 m than at 15 cm.

The RMS difference between the anechoic and the smoothed echoic ILD was calculated for each subject and sound source position. The results of these calculations for the 15 subjects and KEMAR are shown in *Figure 15*. It is clear that the difference between the anechoic and echoic ILD grows rapidly with angle; however, it also varies with distance. These results are consistent with the trends observed in *Figure 14*. At 0 degrees, increasing distance increases the RMS difference between echoic and anechoic ILDs slightly. For angles other than 0 degrees, increasing distance increases the difference between the echoic and anechoic ILDs. This result is predictable because the

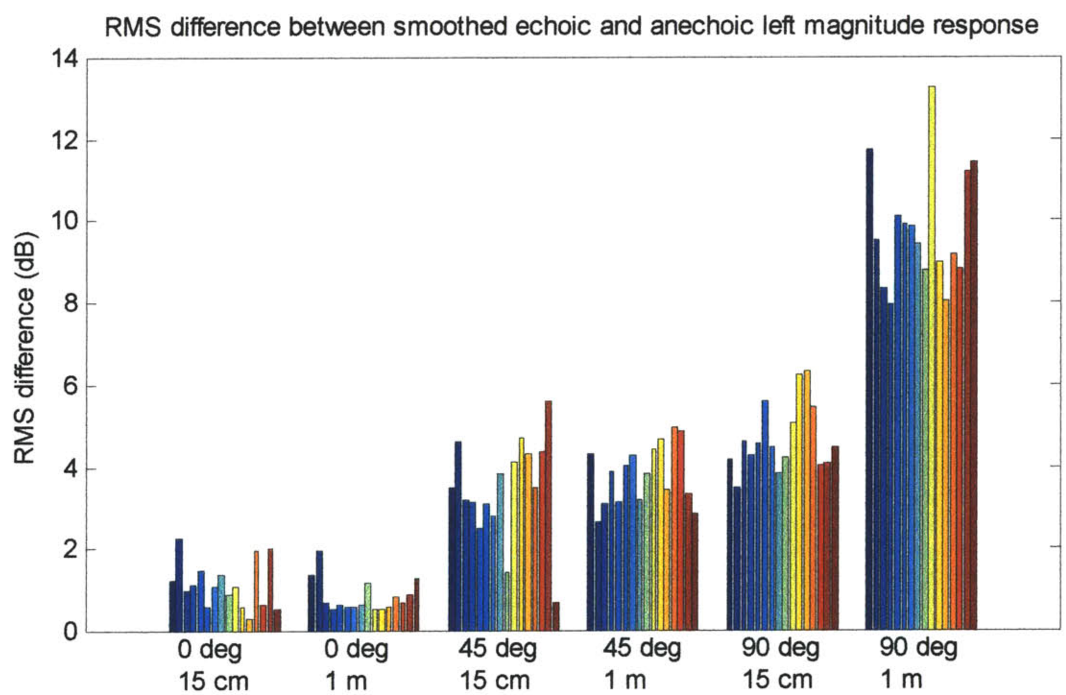


Figure 15: RMS difference between anechoic ILD and smoothed reverberant ILD across 6 sound source positions and 16 subjects (15 people and KEMAR). Each color represents one subject.

relative importance of the reverberation increases as the direct sound level decreases. The deviation seen between anechoic and smoothed reverberant ILDs at 0 degrees is small and is mainly due to the high frequency fluctuations in the echoic ILD.

3.4.3 Interaural Time Differences

Figure 16 shows the ITD with and without reverberation calculated for KEMAR in the center of the room for 6 sound source positions. The anechoic condition is shown in blue and the reverberant condition is shown in blue. The mean of the echoic ITD is the same as that of the anechoic ITD, but the echoic ITD has a larger variance around this

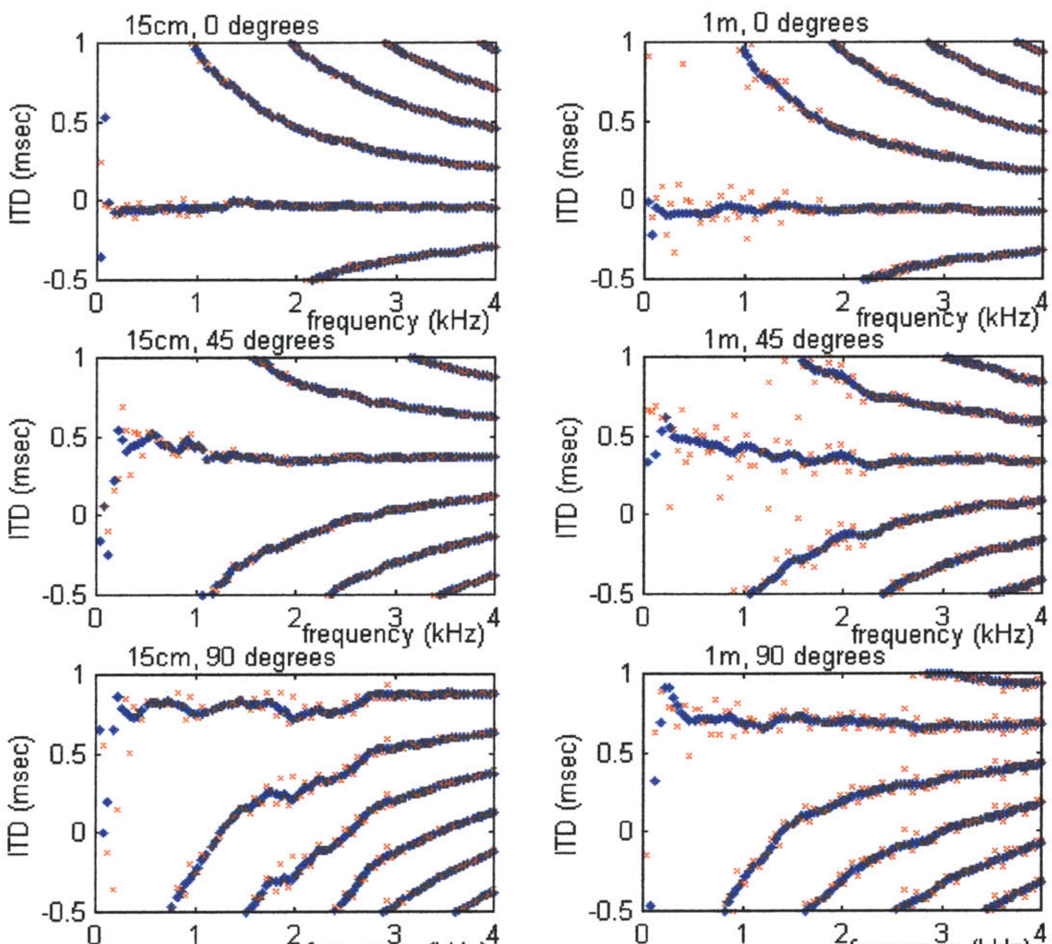


Figure 16: Interaural time difference for KEMAR with six sound source positions. The pseudo-anechoic case is shown in blue and reverberant case in red.

mean. The variance of the echoic ITD is smallest at 0 degrees. These results suggest that reverberation adds noise to the ITD cue, making it more difficult to localize sound using ITD in a reverberant room.

3.4.5 Summary of Reverberation Effects

The results in section 3.4 demonstrate that the level of direct sound influences the degree to which reverberation distorts anechoic localization cues. The relative level of reverberation is greatest for distant sources and sources to the far side of the head, where the direct sound energy is decreased. These results also show that the primary effect of reverberant energy is to cause random phase distortions in interaural time differences, and to decrease the magnitude of interaural level differences toward zero. Magnitude spectrum cues are relatively robust in the near ear, but can become substantially degraded in the far ear. At least to a first approximation, reverberant energy is relatively independent of the source location relative to the listener in the center of the room, so that the relative level of direct reverberant energy (and the relative influence of the reverberation) can be said to vary directly with the direct sound level. The results of section 3.4 are consistent with these expectations, in that the influence of reverberation appears to depend only on the relative level of the direct sound at each ear. However, these results were obtained for a listener positioned in the center of the room. In this configuration, moving the sound source will generally not result in the source being very near any wall. Instead, moving the source in any direction will tend to move it closer to one wall and, at the same time, away from another wall. Since the source is relatively far from all walls, and since the listeners' ears are also relatively far from all reflecting surfaces, changes in source location cause only small modulations in the total reverberant energy reaching the listener, and the reverberant energy level will be approximately invariant with source location.

3.5 Reverberant KEMAR HRTFs (Not Room Center)

Since the relative level of direct and reverberant energy determines the effect of reverberation, additional measurements were taken to examine how reverberant energy changes with the position of the listener in the room. Unlike the cases measured in the previous section, when a listener is positioned relatively close to one or two reflecting surfaces the relative energy contributed by reflections off the nearby surface(s) may dominate the reverberant energy at the ears. Thus, moving the sound source relative to the listener may cause large modulations in the energy of both the direct sound and the reverberation. In addition, since the reflections are now dominated by energy coming from one or two directions (hemifields), the reverberant energy may differ at the two ears. For instance, when a listener is placed with his left ear near a wall, most loud reflections will arise from the left. This in turn may cause a systematic shift in the ILD (at all frequencies) such that the ILD favors the left ear.

In order to determine whether the contributions of the reverberant energy are systematic and can vary with source location when a listener is near a reflecting surface, KEMAR was placed at 3 different positions near (40-100 cm from) the walls in the room. These positions were 1) with his left ear near a wall, 2) with his back near a wall, and 3) in a corner so that both his back and left ear were near walls. In all of these positions, the first echo occurs approximately 2 ms after the arrival of the direct sound, which is long before the direct sound has died out. Thus, reflections may actually interfere constructively or destructively with the direct sound, depending on the exact timing of the reflections relative to the direct sound. In order to analyze the effects of reverberation, the reverberant measurements taken near the wall can be compared to the pseudo-anechoic measurements taken in the center of the room.

3.5.1 Back-to-Wall Position

In the back-to-wall position the reverberant energy reaching both ears increases dramatically over that in the center position. The reverberant energy should be essentially equal at the left and right ears.

3.5.1.1 Magnitude Response

With KEMAR's back to the wall, his left and right ears are equal distance from walls. Therefore, the amplitude of the frequency response is expected to be roughly equal in both ears for sound source locations at 0 degrees. The echoic and anechoic magnitude responses at the left and right ears with KEMAR's back to the wall for 6 sound source locations are shown in *Figure 17*. Just as when KEMAR is in the center of the room, at zero degrees, reverberation does not change the mean of the magnitude responses at the left and right ears, but does increase the variation of the magnitude responses around the mean. At both 45 and 90 degrees and 15 cm, reverberation appears to decrease the mean of the echoic magnitude response from the anechoic magnitude response, while at 45 and 90 degrees and 1 m, the mean echoic and anechoic magnitude responses are about the same. The decrease in the mean of the echoic magnitude response from anechoic magnitude response at 45 degrees and 15 cm, as well as the frequency-dependent fluctuations in the left-ear magnitude response, imply constructive/destructive interference caused by comb-filtering.

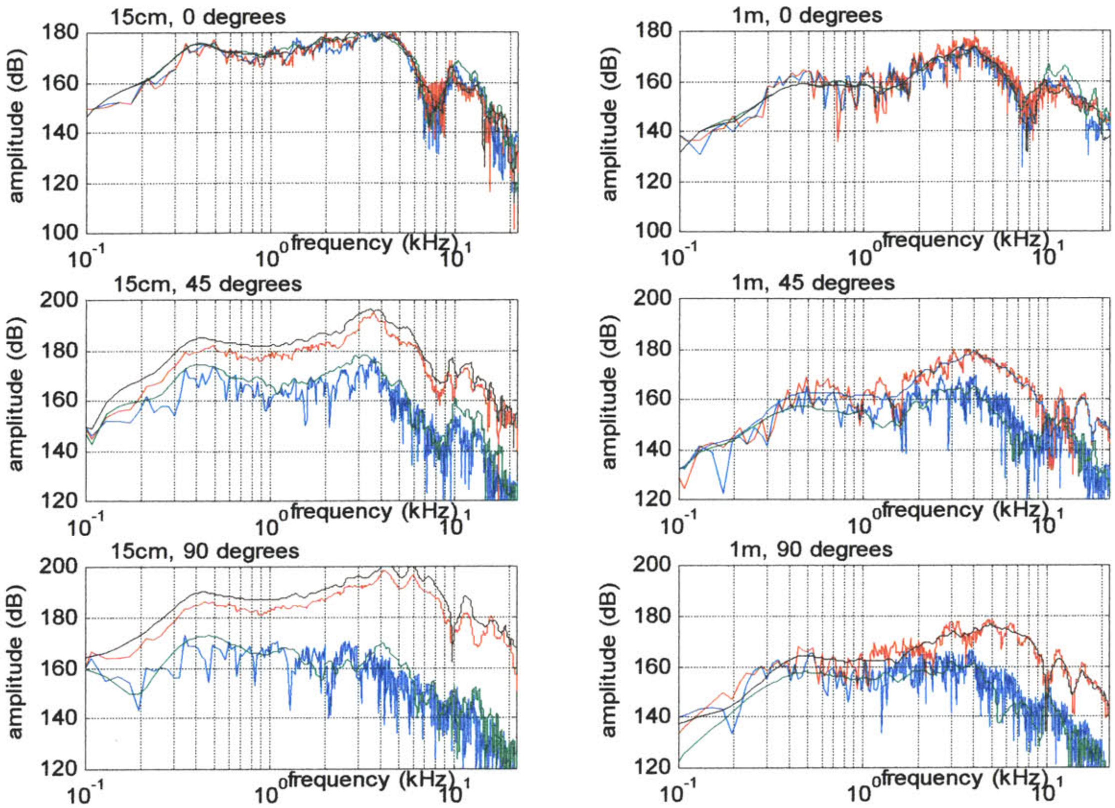


Figure 17: Magnitude response for KEMAR in back-to-wall position for six sound source positions. Red is right ear echoic, black is right ear anechoic, blue is left ear echoic, green is left ear anechoic.

The possibility of comb-filtering is examined in *Figure 18*, which shows the center-position left-ear impulse response in red and the back-to-wall left-ear impulse response in blue. The direct sound energy contained in the first large spike and its subsequent ripples should be approximately the same in both positions without comb-filtering. In the center-position, the first echo does not occur until at least 5 ms after the direct sound and is much weaker than the direct sound. In the back-to-wall condition, the first echo's impulse occurs before the direct sound has died out and is of the same order of magnitude as the direct sound. This means that it interferes with the direct sound impulse, and depending on its exact arrival time with respect to the direct sound, this interference may be constructive or destructive. The effect of the echoic impulse on the total impulse response can be seen by comparing the back-to-wall impulse response to the center–position impulse response. The large distortions in the frequency and amplitude of impulse response ripples after the first echo show that comb-filtering occurs.

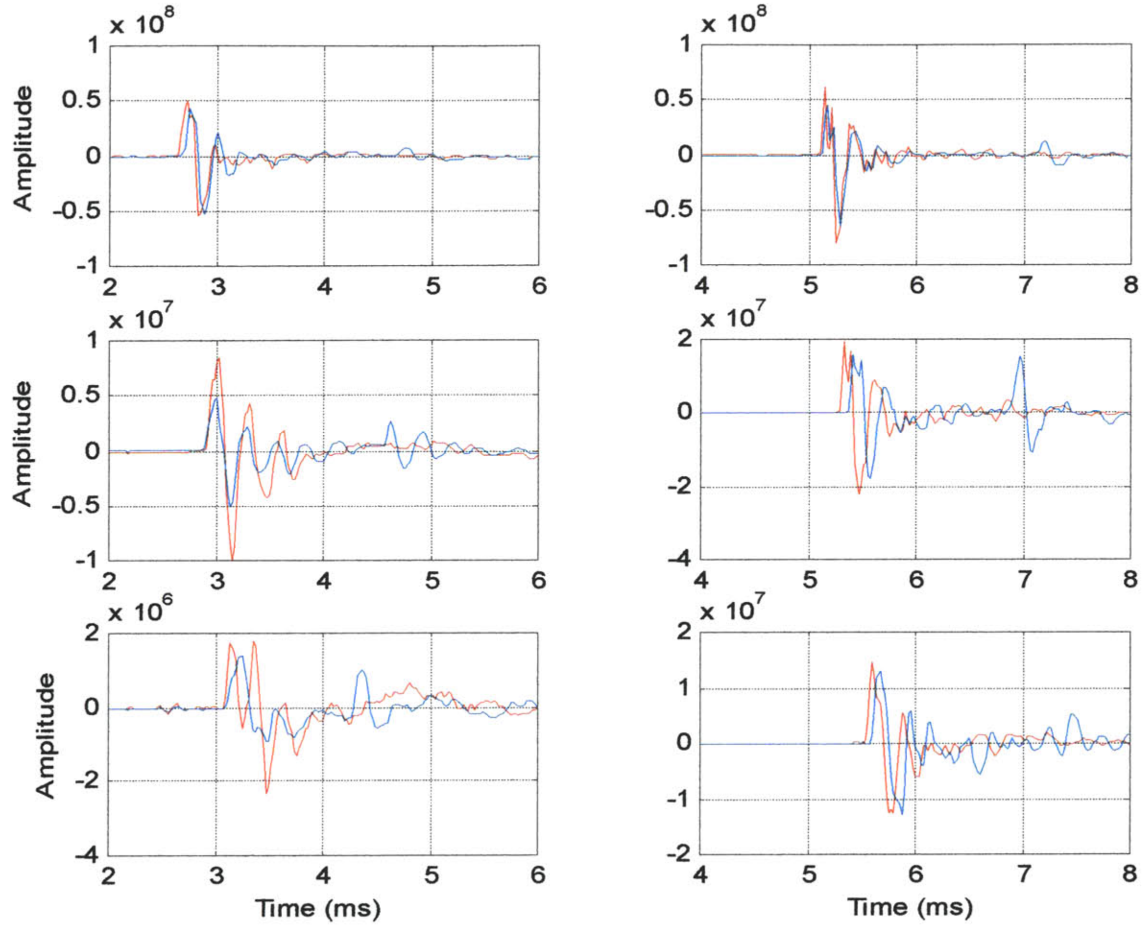


Figure 18: Center of room impulse response (blue) and back-to-wall impulse response (red) at the left ear with 6 sound source positions.

3.5.1.2 Interaural Level Differences

Figure 19 shows the anechoic ILD in the center of the room (blue), compared to the smoothed echoic ILD with KEMAR's back close to the wall (red). At 45 degrees and 15 cm, the ILD at low frequencies varies around the anechoic ILD due to frequency-dependent fluctuations in the left ear magnitude (see *Figure 17*). These fluctuations and the peaks and notches in echoic ILD with frequency are caused by constructive/destructive interference due to the addition of the direct sound impulse response plus the wall impulse response leading to a comb-filtering affect (as seen in *Figure 18)*. At the right ear, cancellation occurs that is not frequency dependent. At 90 degrees, the differences between the anechoic and reverberant ILDs are larger, but follow a similar pattern.

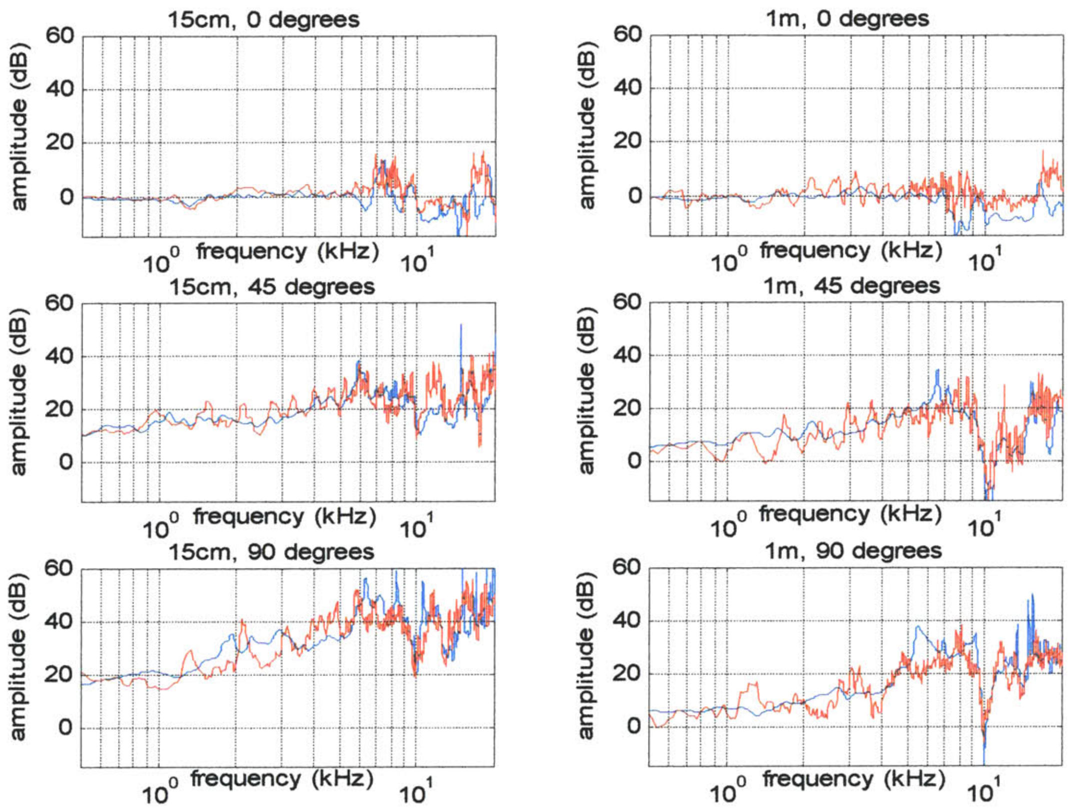


Figure 19: Anechoic ILD with KEMAR in the center of the room (red) and smoothed echoic ILD with KEMAR's back to the wall for 6 positions.

3.5.1.3 Interaural Time and Phase Differences

Figure 20 shows the reverberant and anechoic interaural time differences as a function of frequency for KEMAR in the back-to-wall position for 6 sound source positions. The reverberant case is shown in red and the anechoic case is shown in blue. On average (across frequency), the reverberant ITD equals the anechoic ITD; however the reverberant ITD oscillates around the smoother anechoic ITD. The amount of variation in the reverberant ITD around the anechoic ITD depends on source position, and is particularly small for a source at 0 degrees and 15cm. As the sound source angle increases towards 90 degrees, these variations increase significantly. The variations between reverberant and anechoic ITD are much larger when KEMAR is in the back-to-wall position than in the center-of-room position (*Figure 16*). This is expected because the reverberation is more intense in the back-to-wall position.

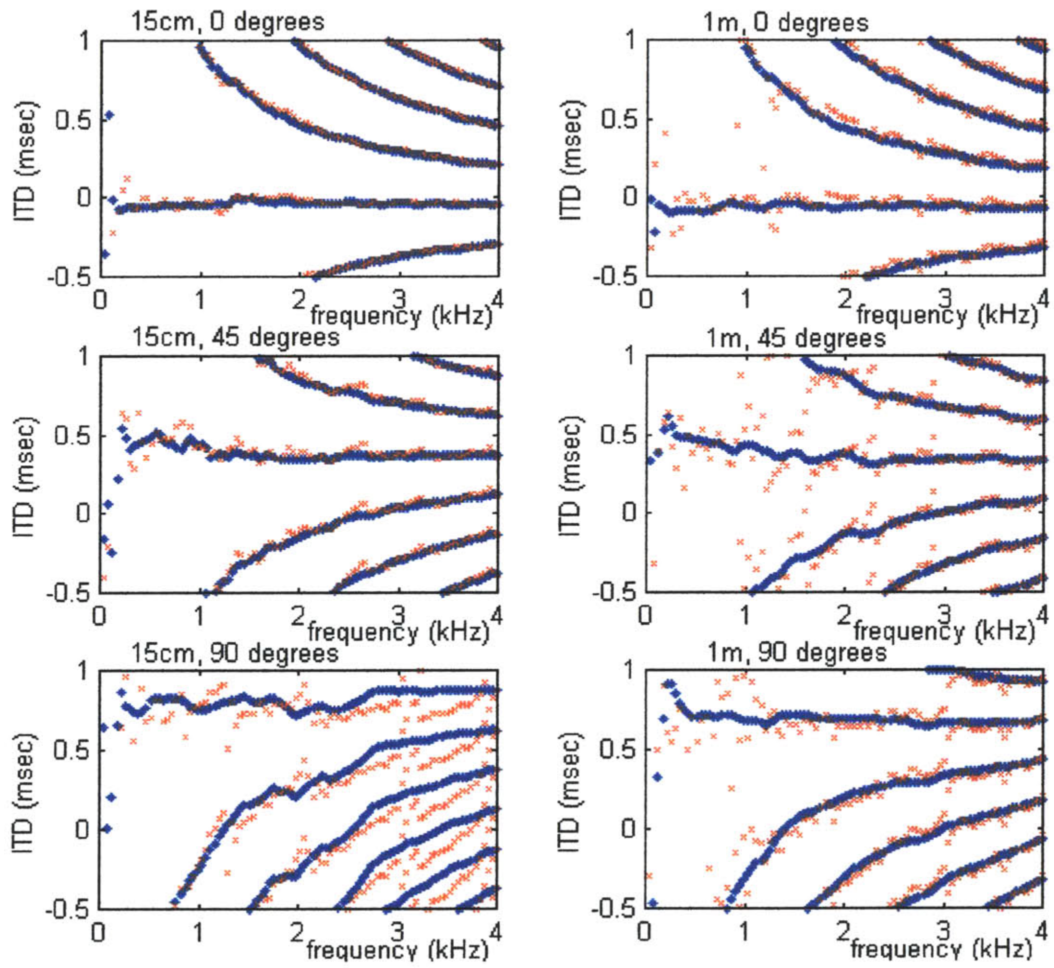


Figure 20: Interaural time difference for KEMAR in back-to-wall position with six sound source positions. The pseudo-anechoic case is shown in blue and the reverberant case in red.

3.5.2 Left-Ear-to-Wall Position

In the left-ear-to-wall position, reverberant energy at the left ear increases dramatically while the direct sound at the left and right ears remains unchanged.

3.5.2.1 Magnitude Responses

It is expected that the reverberant energy reaching the left ear will be very strong compared to the direct sound energy. The direct sound will be affected by a strong reflection off the left wall. This should lead to increasing left echoic magnitude response relative to the anechoic magnitude response as the source moves to the right of the head (which will decrease the magnitude of the direct sound). In contrast, the increase in energy at the right ear will be very small relative to the energy of the direct sound, since the reflections off the left wall must pass by the head before reaching the right ear. Therefore, there should be only a small difference between right ear echoic and anechoic magnitude responses.

Figure 21 shows the reverberant and echoic magnitude responses at the left and right ears with KEMAR in the left-ear-to-wall position for 6 sound source positions. At zero degrees, reverberation does not change the mean value of the magnitude responses, similar to the results found in the center of the room and in the back-to-wall position. At all positions, for both ears, reverberation increases the variance of the magnitude response with frequency. The frequency-dependent peaks and notches in the left-ear magnitude response are caused by comb-filtering as in the back-to-wall case. Because the reverberation is stronger and arrives at the left ear earlier in the left-ear-to-wall conditions than in the back-to-wall condition, there is more interference between the

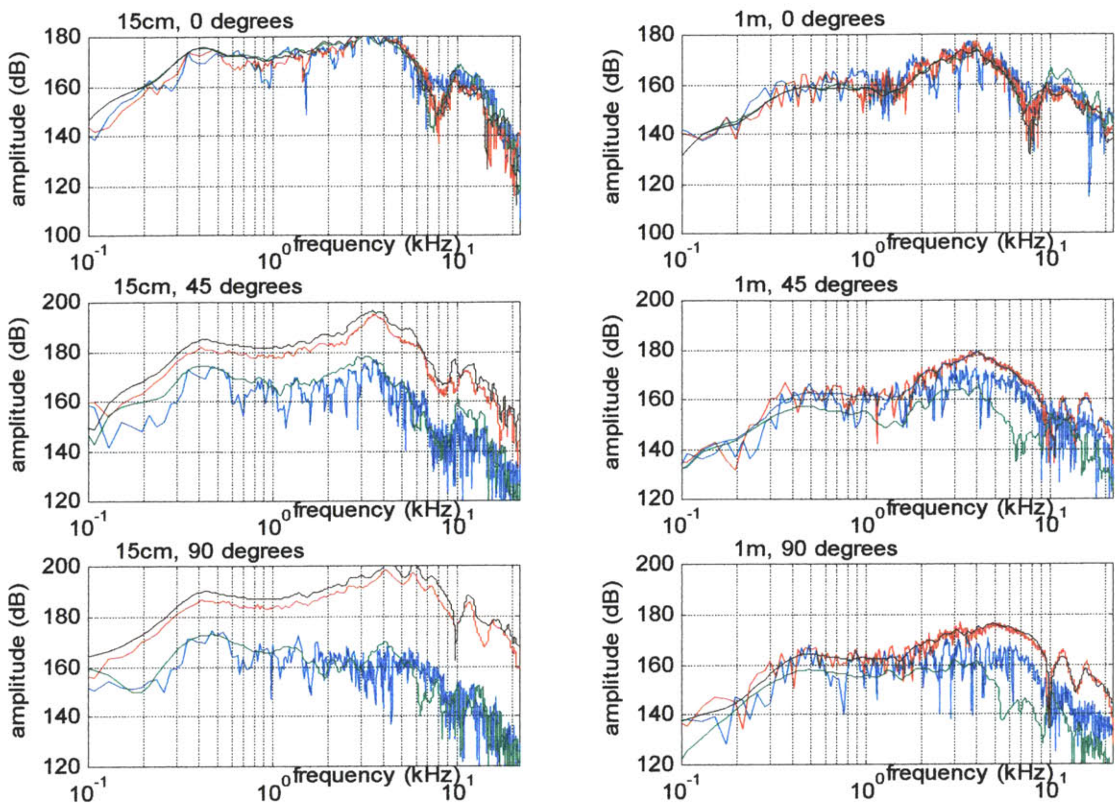


Figure 21: Magnitude response for KEMAR in left-ear-to-wall position for six sound source positions. Red is right ear echoic, black is right ear anechoic, blue is left ear echoic, green is left ear anechoic.

direct and reverberant sound and a larger comb-filtering affect. As expected, at 1 m, as the sound source angle increases towards 90 degrees, the magnitude of the left reverberant frequency response increases relative to the left anechoic frequency response. The overall magnitude of the left ear echoic and anechoic frequency responses decreases with increasing sound source angle while the right ear magnitude response increases with increasing sound source angle due to the direct sound decreasing at the left ear while increasing at the right ear. At a greater sound source distance, the magnitude response at both ears decreases due to decreasing direct sound energy. The variance of the echoic magnitude response is smaller for the right ear at 15 cm than at 1 m and than at all positions for the left ear.

3.5.2.2 Interaural Level Differences

As mentioned in section 3.5.2.1, it was expected that the reverberant energy reaching the left ear would be strong relative to the direct sound energy, while the increase in reverberant energy at the right ear will be very small relative to the energy of the direct sound. Thus, when KEMAR is placed with his left ear near the wall, we generally expect the ILD at 45 and 90 degrees to be smaller than the anechoic ILD. At 0 degrees, the direct sound should be equal at the left and right ears, but the reverberation should be much stronger at the left ear, leading to a slightly negative ILD.

Measurements with the left ear very near the wall (less than 50 cm from the wall) at zero degrees and at 15 cm were found to be very sensitive to sound source location, especially compared to measurements taken with other subject positions and sound source distances and angles. Positioning the speaker even a small fraction of a degree off-center caused large differences in direct sound magnitude at the left versus right ears. The slight positioning off-center was too small to be noticed by looking at the speaker position, but was noticeable when the impulse responses were examined. Very slight deviations from the center point lead to up to a 40% difference in the amplitude of the direct sound at the left versus right ears, favoring the right ear. This effect is caused by interference of the impulse response off the wall with the direct impulse response leading to the to comb-filtering of the left ear response which adds at some frequencies while subtracting at others.

Figure 22 shows the smoothed reverberant ILD with KEMAR's left ear near the wall (red) and the anechoic ILD with KEMAR in the center of the room (blue) for the same source positions (relative to the head). As expected, at 0 degrees, the ILD averages just a little less than zero, but it has a few sharp peaks above zero, most likely due to asymmetrical pinna effects. As the sound source angle increases, the difference between the echoic left-ear-to-wall ILD and the anechoic ILD increases. Increasing distance also increases this difference. These differences are largest at high frequencies, as they have been for all other subject positions. The increase in ILD difference as angle and distance increase is expected because the ratio of reverberation to direct sound energy at the left ear increases while that at the right ear does not change significantly. Comb-filtering of the left magnitude response due to interference of the echo and direct sound causes peaks at some frequencies in the echoic ITD and notches at others. The depth of modulation is greater at 1 m than at 15 cm due to the echo level being relatively close to the level of the direct sound at 1 m.

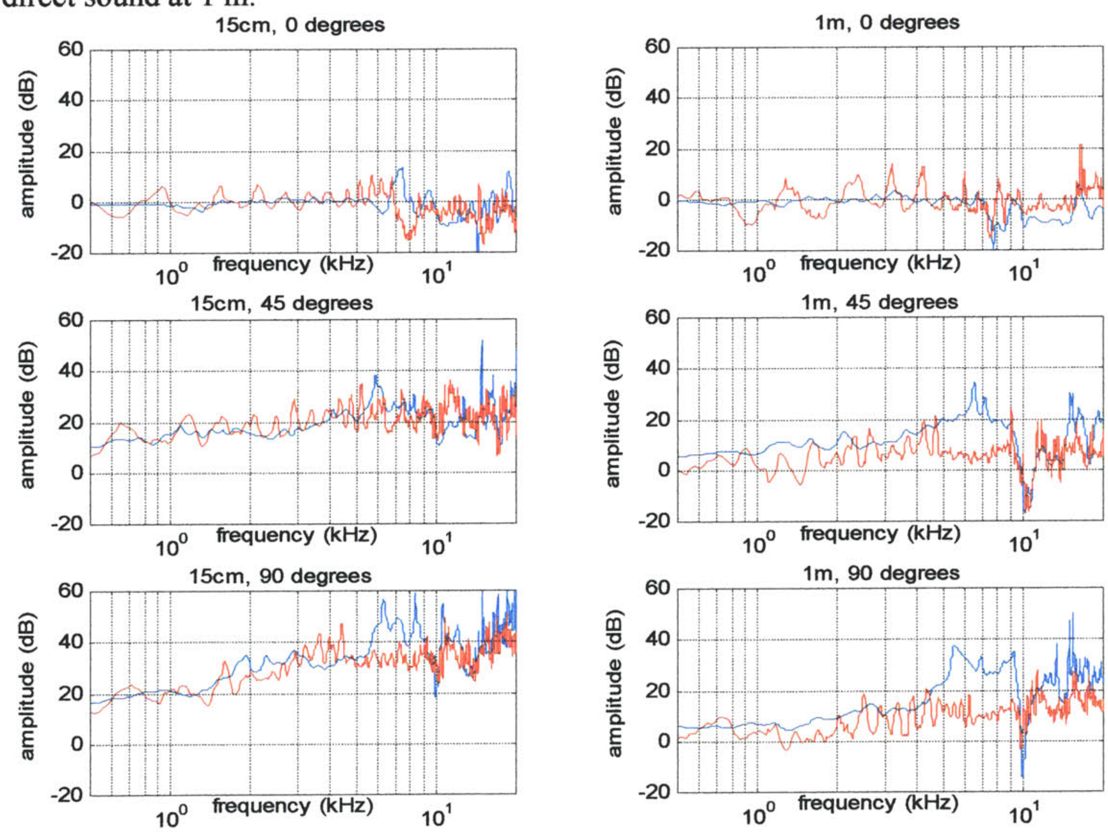


Figure 22: Anechoic ILD with KEMAR in the center of the room (red) and smoothed echoic ILD with KEMAR's left ear to the wall (blue) for 6 sound source positions.

3.5.2.3 Interaural Time and Phase Differences

Figure 23 shows the interaural time difference in the (reverberant) left-ear-to-wall position (red) and in the (anechoic) center position (blue) for 6 sound source positions. The mean ITD across frequency in the (echoic) left-ear-to-wall position is approximately equal to that in the anechoic case although the left-ear-to-wall ITD shows frequency-to-frequency variations. The magnitude of these variations increases as sound source distance and angle increase, to the point that in the 1 m, 90 degree condition, there is no single ITD which is consistent for all frequencies. As distance increases, the magnitude

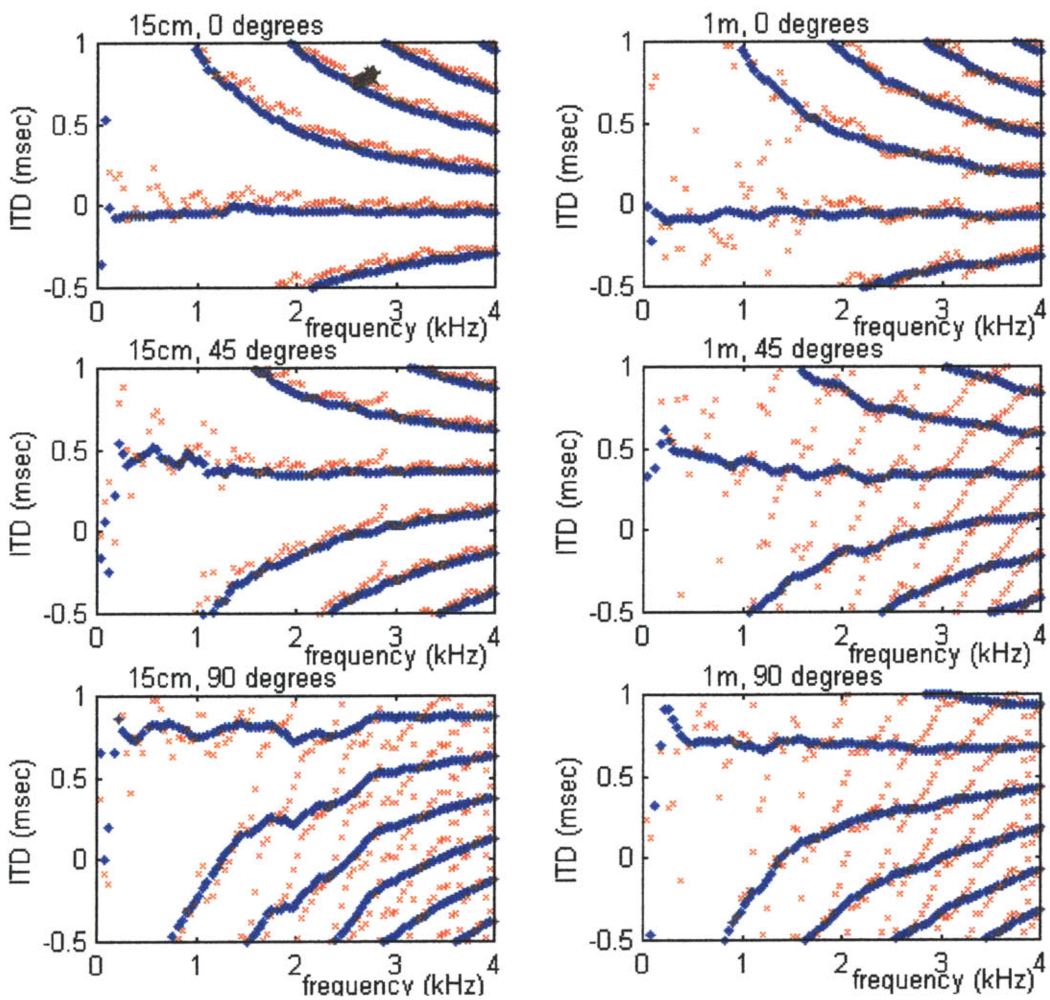


Figure 23: Interaural time difference for KEMAR in left-ear-to-wall position with six sound source positions. The pseudo-anechoic case is shown in blue and the reverberant case in red.

of the direct sound decreases, and as angle increases, reverberant energy increases. Both of these effects lead to the increased ITD variations

3.5.3 Corner Position

Finally, KEMAR was placed in a corner so that his back and left ear were both within 1 m of walls. In this position, the reverberation is greater than at the previous positions. The back wall will tend to increase the reverberant energy reaching both ears equally. The left wall position will tend to increase the reverberant energy reaching both ears as well, but will cause a greater increase at the left ear.

3.5.3.1 Magnitude Responses

The magnitude responses at the left and right ears with KEMAR in the corner position for 6 sound source positions are shown in *Figure 24*. In *Figure 24*, the right ear echoic response is represented by red, the right ear anechoic response is represented by black, the left ear echoic response is blue, and the left ear anechoic response is green. At zero degrees, the magnitude responses at the left and right ears are about the same. For the most part, at zero degrees, the cross-frequency mean of the echoic and anechoic magnitude responses are the same, but the echoic response shows large frequency-to-frequency variations. At zero degrees and 1 m, the average of the echoic response magnitude appears to be just a little greater than that of the anechoic. As angle increases, the difference between the left and right magnitude responses increases because the direct sound energy at the left decreases, resulting in an increase in relative reverberation level. The difference between the echoic and anechoic responses also increases as angle increases since the relative reverberation level increases with increasing angle. The reverberation increases the left magnitude response particularly at high frequencies. Comb-filtering is evident in the frequency-dependent peaks and notches in the left ear magnitude responses. There is more reverberant energy in the corner position than in the back-to-wall and left-ear-to-wall position so the interference leading to comb-filtering should be largest in the corner position. The frequency-by-frequency variations in the echoic magnitude responses are greater at the right ear with the speaker 1 m from the head than at 15 cm, where the direct sound at the right ear is more intense. The difference between the left and right magnitude responses is also much greater at 15 cm than at 1 m.

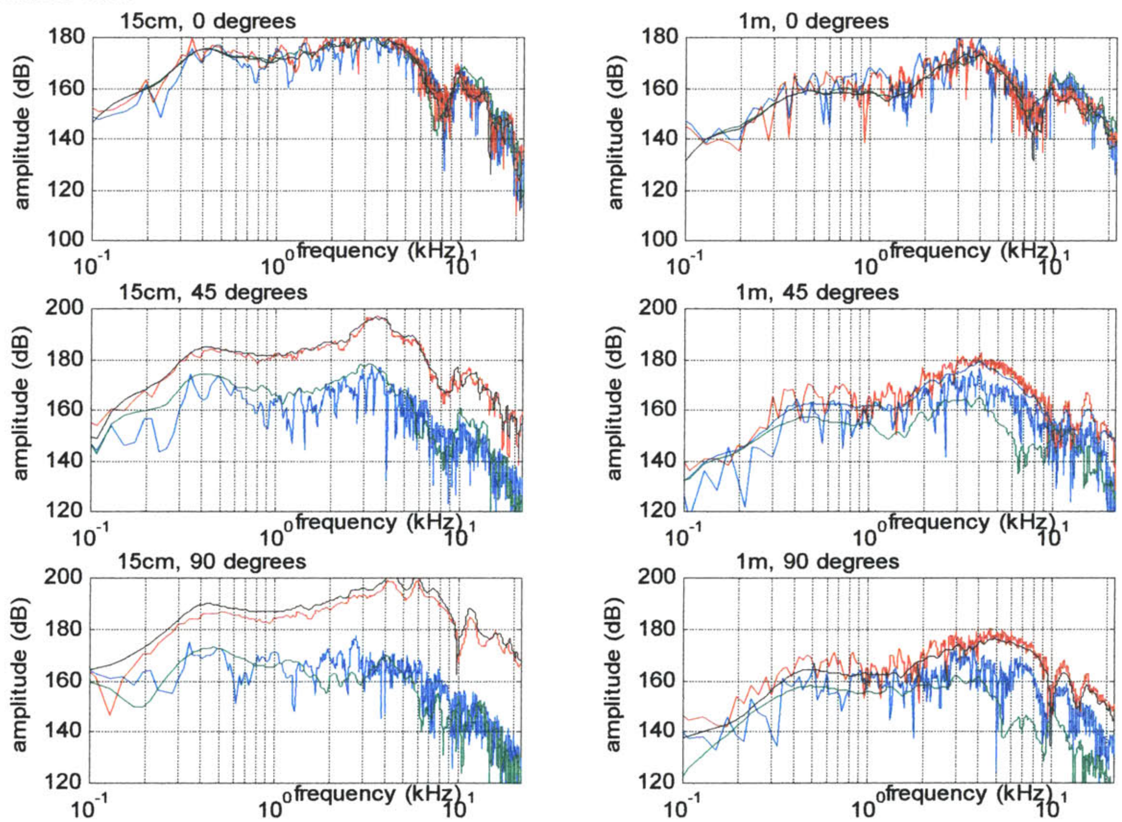


Figure 24: Magnitude response for KEMAR in the corner position for six sound source positions. Red is right ear echoic, black is right ear anechoic, blue is left ear echoic, green is left ear anechoic.

3.5.3.2 Interaural Level Differences

At zero degrees, the ILD should be near or slightly less than zero (to the extent that the left wall consistently biases the ILD). The average ILD for the corner position is expected to be a little less negative than the left-ear-to-wall position because the wall behind KEMAR will contribute equal reverberation to each ear. As sound source angle and distance increase, the ILD is expected to decrease (compared to the anechoic ILD). The direct sound at the left ear decreases as angle increases, but the interaural difference in reverberation increases as the sound source angle and distance increase.

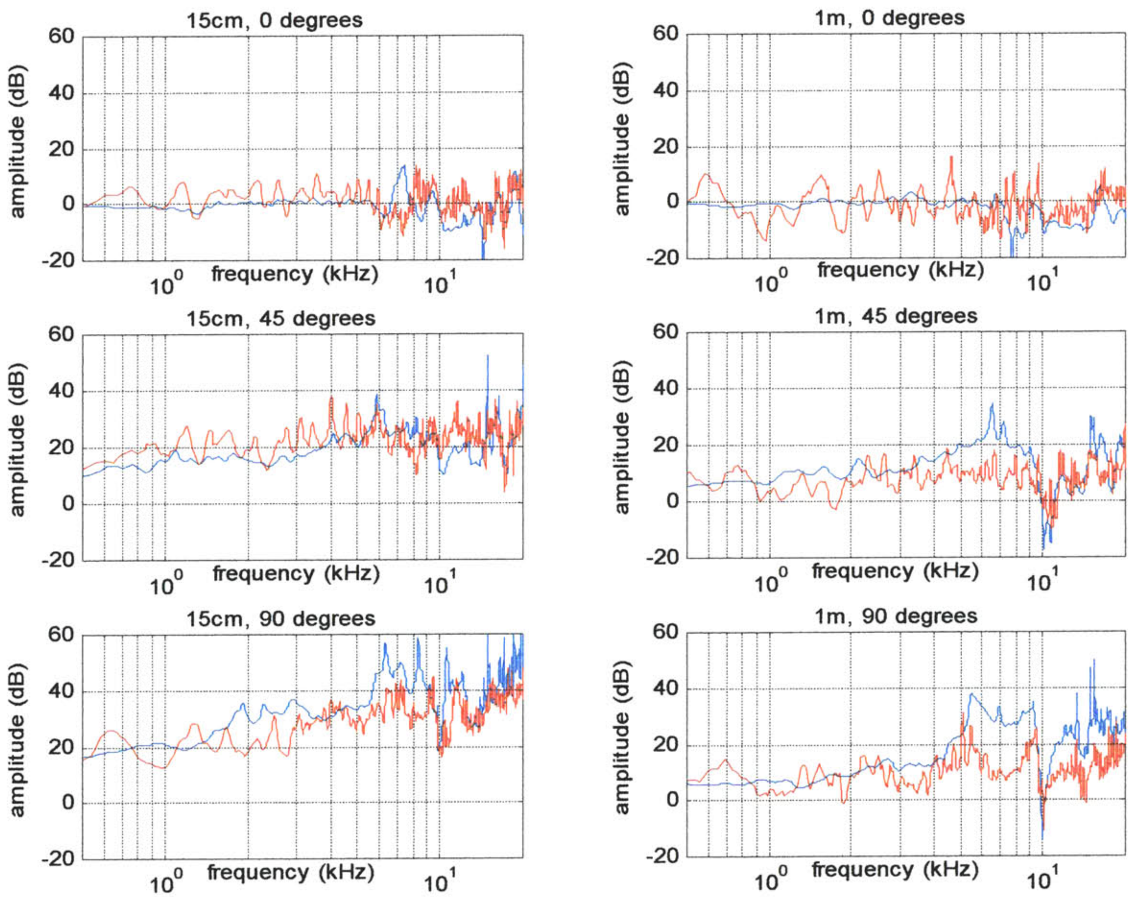


Figure 25: Anechoic ILD with KEMAR in the center of the room (red) and smoothed echoic ILD with KEMAR's left ear and back to walls for 6 positions.

The general trends in the ILD measurements shown in *Figure 25* are as expected. The differences between the anechoic ILD and the echoic corner-position ILD increase with increasing angle, distance, and frequency. The corner position ILD is larger than expected at zero degrees. This could be explained by comb-filtering, which can account for the cancellation that appears to be occurring. Overall, the ILD appears to average just above zero at 15 cm and around zero for 1 m. The measurement at 15 cm is most likely higher than expected because of the extreme sensitivity to sound source placement due to sensitivity to the relative timing of the direct and reverberant bursts, which affects whether direct and reflected energy cancels or adds when the left ear is very close to the wall.

3.5.3.3 Interaural Time and Phase Differences

The ITD for KEMAR in the corner of the room for 6 sound source positions is shown in *Figure 26*. The reverberant case is shown in red and anechoic (center-of-room) case is shown in blue. The reverberant ITD varies more from the anechoic ITD than for any of the room positions considered in previous sections due to the greater reverberant energy in this position. In particular, there is no consistency in the ITD cues at different frequencies. As sound source angle and distance increase, the distortion of the ITD by the reverberation increases. The discontinuity in interaural phase seen as sound source angle increases is consistent with comb-filtering.

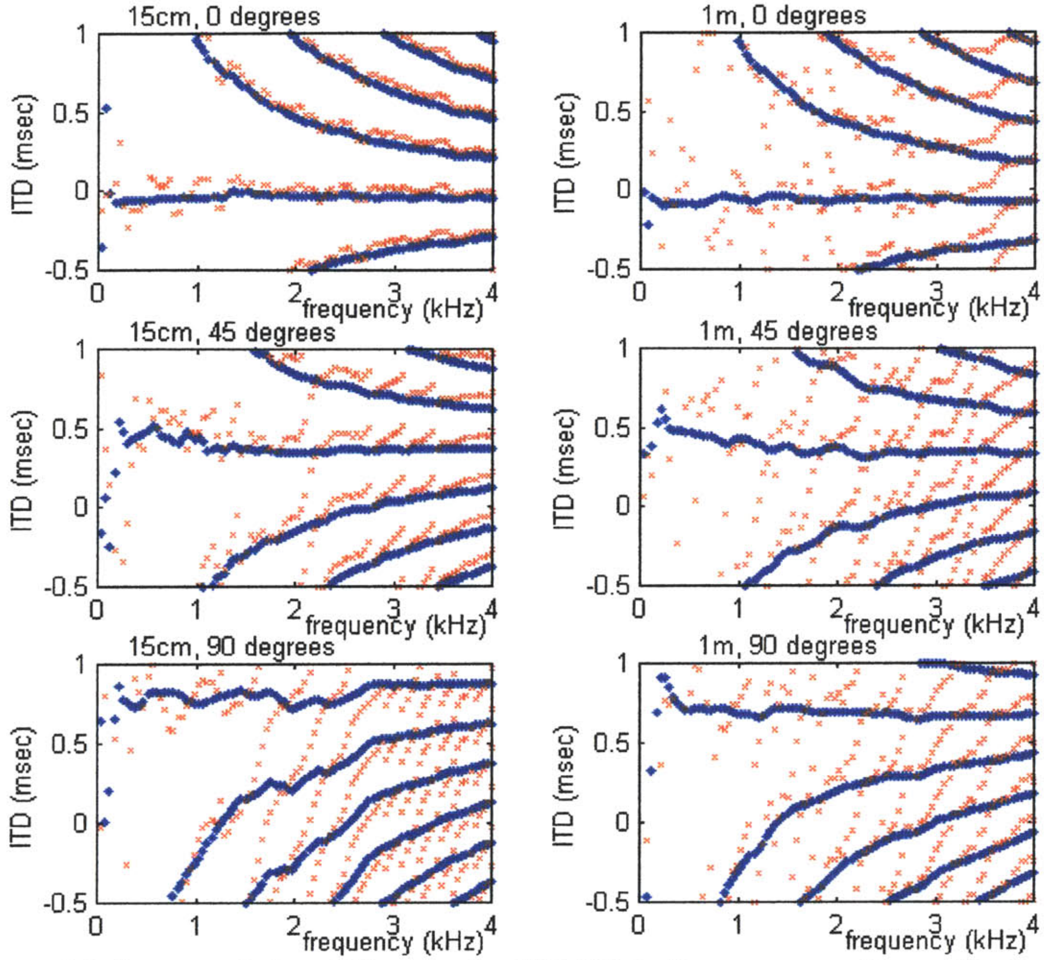


Figure 26: Interaural time difference for KEMAR in the corner position with six sound source positions. The pseudo-anechoic case is shown in blue and the reverberant case is shown in red.

3.5.4 Summary of Room Position Effects

The effects of room position on basic acoustic cues are summarized in this section. The RMS difference between cue value for the reverberant and corresponding anechoic conditions is used to summarize the magnitudes of the effects of reverberation on the acoustic cues.

3.5.4.1 Magnitude Responses

Figures 27a and *27b* show the RMS difference in the magnitude responses for the right and left ears respectively. The RMS difference depends on both room position and

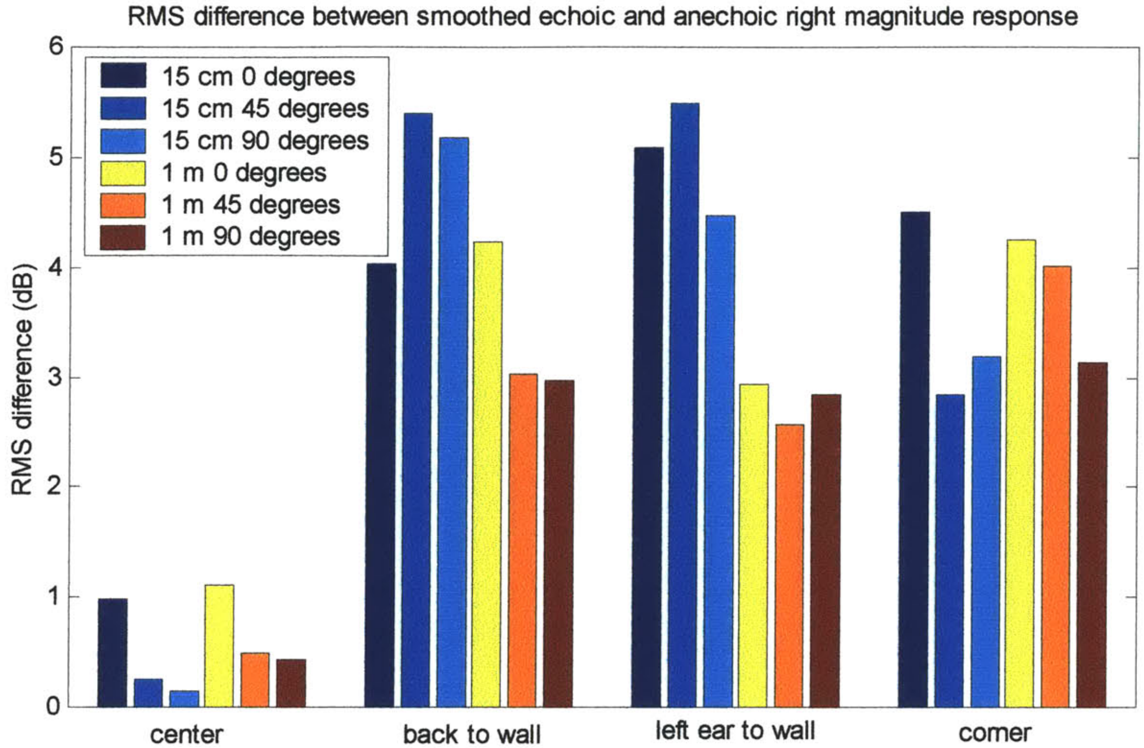
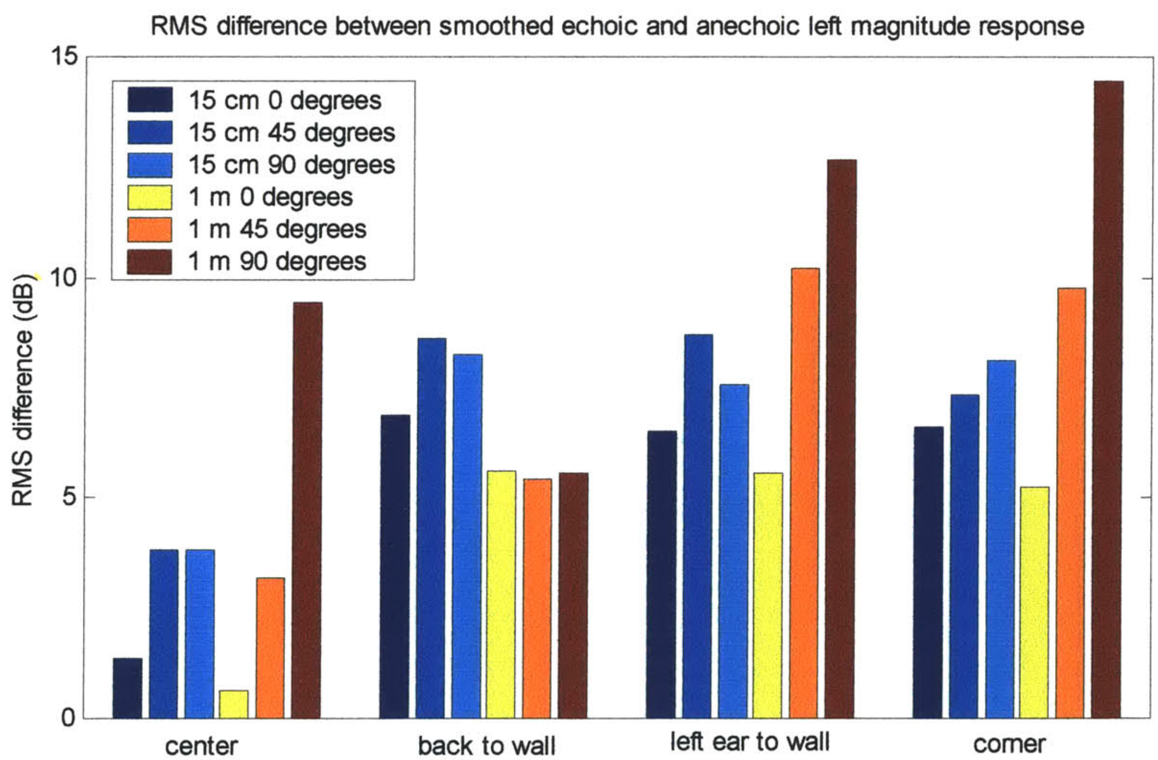


Figure 27a: RMS of smoothed echoic right magnitude response from anechoic right magnitude response for 4 KEMAR positions in the room and 6 sound source positions.



27b: Same as 27a, but for left magnitude response source position. The only consistent effect of room position is that the RMS difference is least when the source is in the room center. In general, the effect of the reverberation is much larger at the left ear than that at the right ear (as expected given that the right ear generally receives more direct sound energy).

3.5.4.2 Interaural Level Differences

The RMS difference between the echoic and pseudo-anechoic ILDs was compared for all four room positions and all six sound source positions. The results are summarized in *Figure 28*. As expected, the RMS difference is generally smallest when KEMAR is located in the center of the room (where the reverberation is weakest).

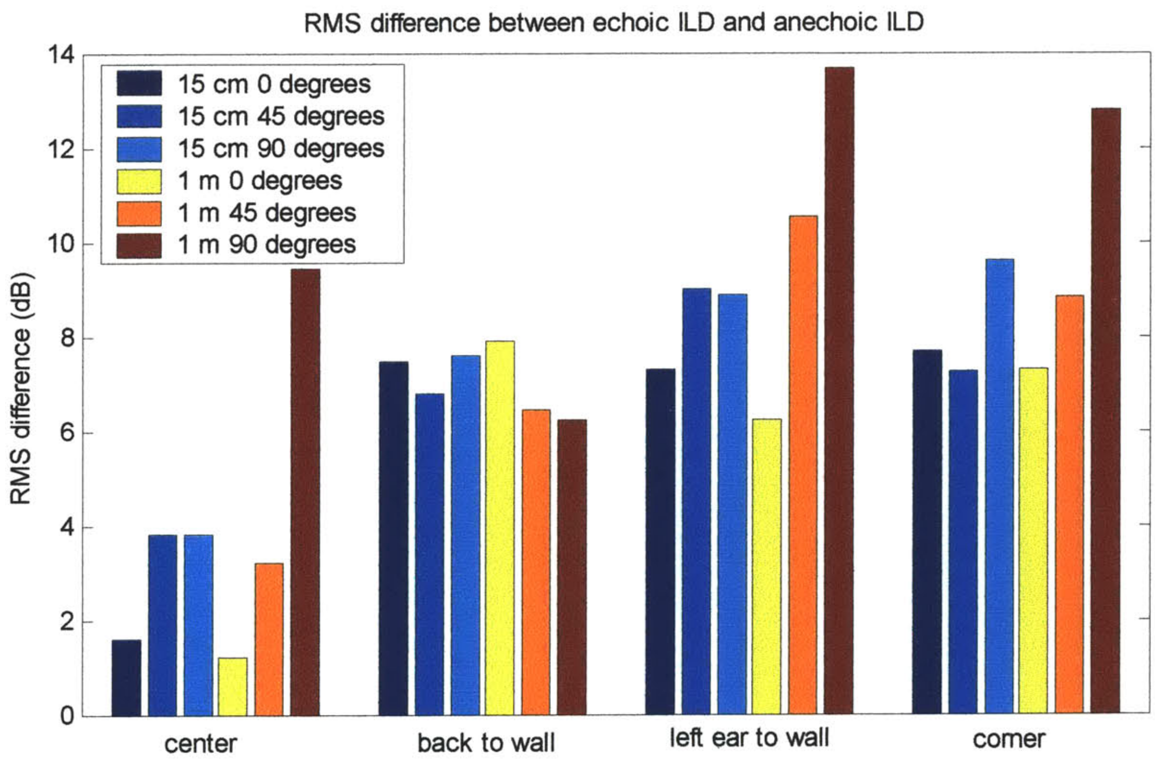


Figure 28: RMS difference of echoic ILD at four KEMAR positions in the room and 6 sound source positions from the anechoic ILD in the middle of the room at the same sound source positions.

When KEMAR is in the back-to-wall position, the RMS difference is generally greater than that when he is at the center position, and less than when he is in the other two positions. Having the wall behind KEMAR affects both ears approximately equally, which leads to small variations in the ILDs. In contrast, for the corner and left-ear-to-wall positions, the walls produce strong early reflections, which cancel and reinforce the direct sound in a frequency-dependent manner. As a result, the RMS difference for these room positions is generally greater than for other room positions. When KEMAR is positioned with his back to the wall, the reverberation generally is larger than the reverberation reaching the ears when KEMAR is in the center of the room. Thus, the RMS difference is larger for the back-to-wall position than for the center position for most sound source positions. The exception to this is for angles near 90 degrees, where the difference in reverberation reaching the left ear versus the right ear is substantially greater for the center position.

The largest RMS difference occurs for the left-ear-to-wall position. In this position, the reverberant energy reaching the left and right ears differs drastically. Also, the greater reverberant energy adds to the weaker left ear signal, causing large effects in the left ear total signal level. For the left-ear-to-wall position, the center position, and the corner position, there is a strong correlation between source angle, source distance, and the RMS difference. Most of this difference occurs at frequencies above 5 kHz.

3.5.4.3 Interaural Time and Phase Differences

Changing the position of the subject in the room is not expected have any consistent effect on ITD across all frequencies though it is expected to add frequency-dependent variations in the ITD. Changing the location of the subject does not change the distance between the subject's ears or their location with respect to one another, so there is not expected to be any significant change in overall ITD. However, since the reverberation will cause frequency-dependent interactions between direct and reverberant responses, the reverberation should cause variations in the ITD as a function of frequency.

The results summarized in *Figure 29* show that ITD (as computed by cross correlating the left and right ear signals) is roughly invariant with the subject's position in the room, but is dependent on sound source position. The small fluctuations seen in ITD across subject positions in *Figure 29* are within the range seen across subjects in *Figure 7*. These fluctuations are probably caused by slight differences in KEMAR's position or the sound source position in different trials. As seen in *Figure 7*, the ITD at 90 degrees is about twice that at 45 degrees, and the maximum ITD calculated is slightly less than 1 msec. The ITD at zero degrees is again slightly negative. Because KEMAR's head is.

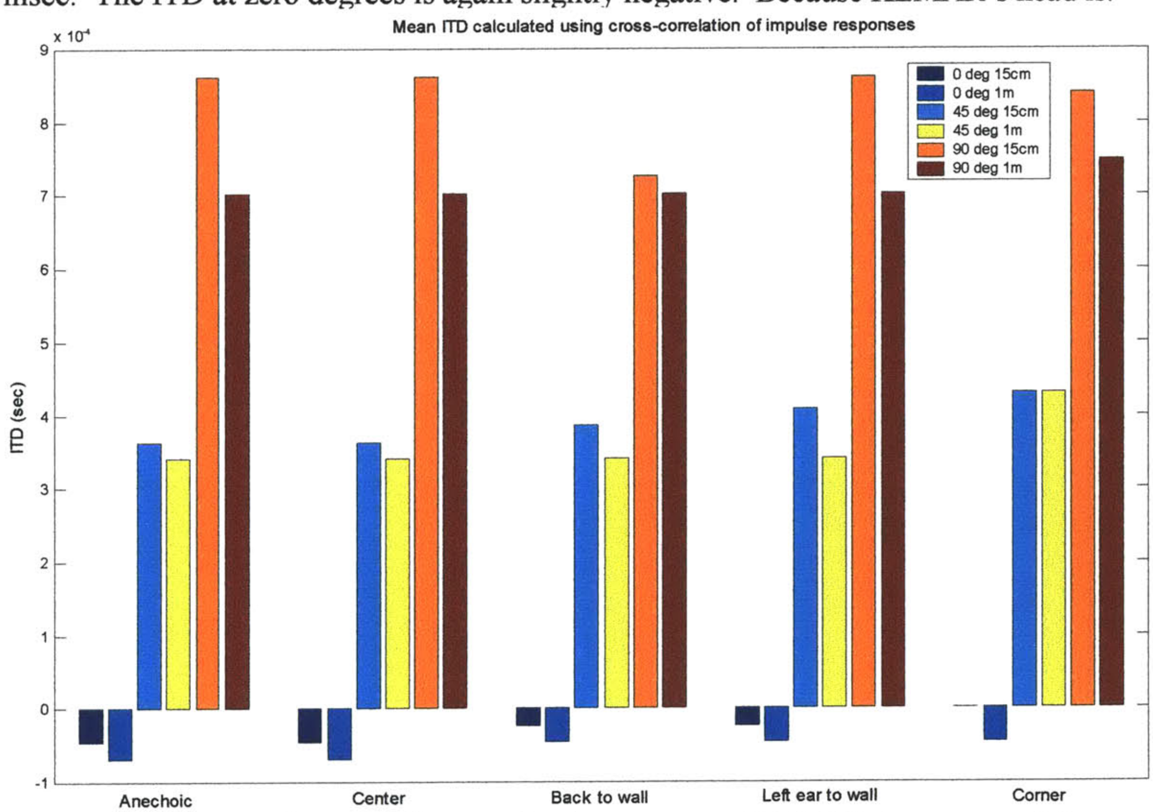


Figure 29: Interaural time difference for 4 positions of KEMAR in room and 6 sound source positions.

symmetrical, the negative ILD indicates that the speaker was again placed slightly left of center in front of KEMAR

Figure 30 shows the average ITD calculated from the ITD/IPD plots for the anechoic condition and the 4 echoic room positions. These results are consistent with those found by cross-correlating the left and right ear signals shown in *Figure 29*. This confirms that the average ITD depends on sound source position and not reverberation.

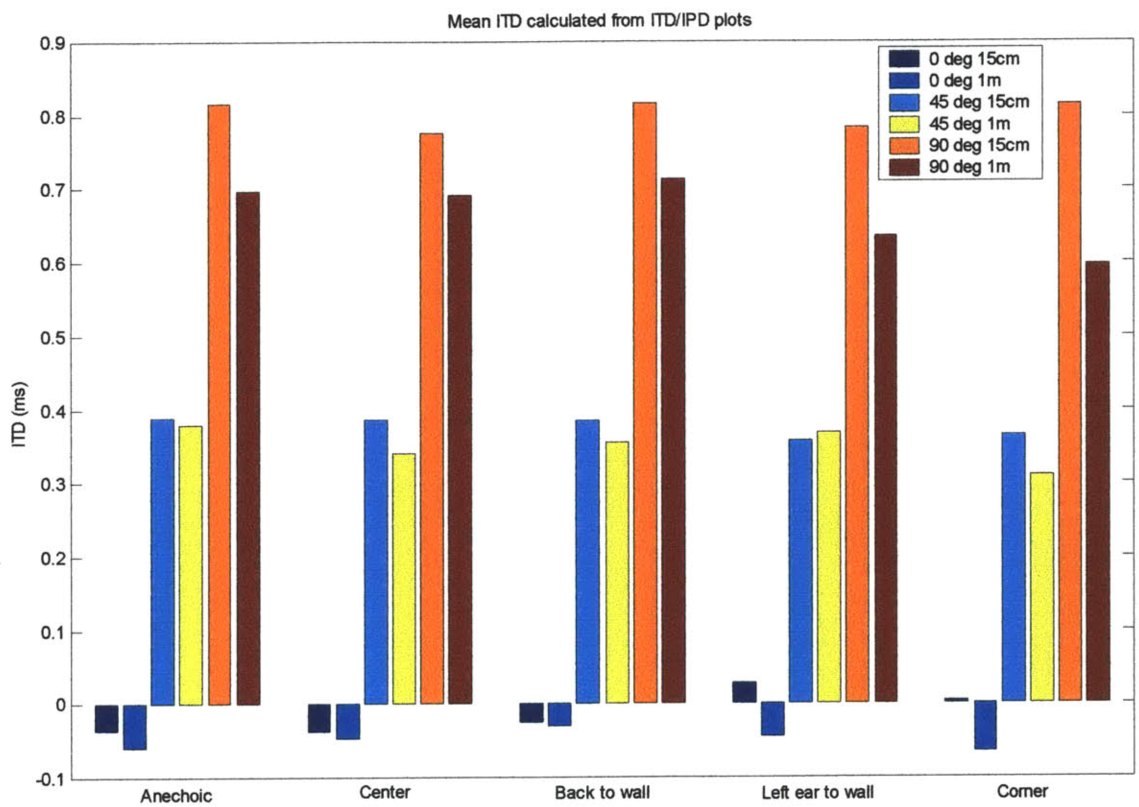


Figure 30: Average interaural time difference calculated from ITD/IPD plots for anechoic case and 4 KEMAR room positions with 6 sound source position.

The variance of the ITD around its mean was calculated from the ITD/IPD plots and is shown in *Figure 31*. This quantifies the observation made in previous sections that the variance increases with reverberation. *Figure 31* shows that variance increases significantly as reverberation increases. In the anechoic condition, the variance is negligible measuring less than 0.01ms. In the echoic center condition, the variance increases to up to 0.07 ms. In the near-wall room positions, the variance increases dramatically to as much as 0.35 ms. In general, variance increases as sound source distance increases (for most positions), and as angle increases.

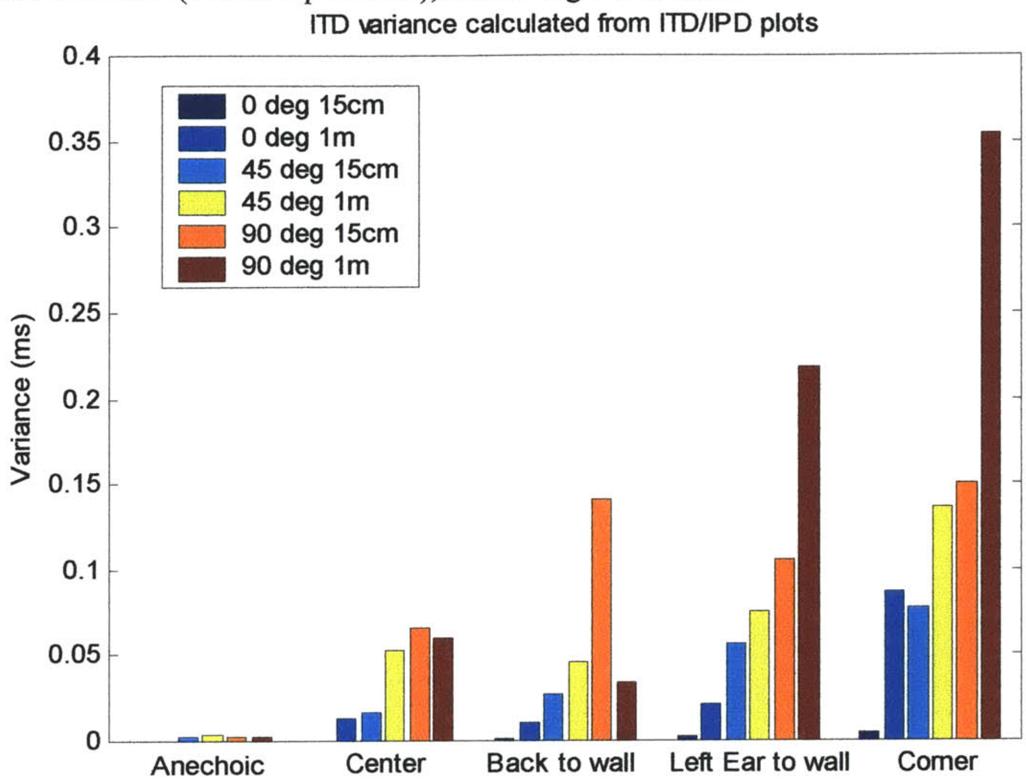


Figure 31: ITD variance calculated from ITD/IPD plots for anechoic condition and 4 KEMAR room positions with 6 sound source positions.

4.0 Conclusions

The results found in this study are qualitatively consistent with acoustic theory. The effect of reverberation on HRTFs increases as the sound source moves from 0 to 90 degrees and from 15 cm to 1 m. Reverberation has the largest effect on the HRTF at the ear furthest from the sound source, where the direct energy is smallest. Reverberation adds frequency-to-frequency variation in the magnitude response, "fills in" notches in the magnitude response, and for conditions near walls where there is a strong initial reflection, it causes comb-filtering, sometimes even leading to an overall decrease in the energy at the near-reflection ear. In contrast, near the center of the room, essentially random fluctuations are seen in the magnitude response due to the reverberation. In general, the reverberation tends to decrease the ILD magnitude as the source moves laterally or its distance increases. The ILD magnitude is smallest when the subject is positioned such that the ear with the least direct sound energy receives the most reverberant energy. These ILD decreases are caused primarily by changes in the signal level at the farther ear, where the direct sound energy is lowest. For near-wall conditions, comb-filtering causes fluctuations around the mean anechoic ILD due to frequency-dependent interactions at the ear near the wall reflections. Reverberation does not cause a consistent bias in ITD values across frequency, but does result in increased variability in the ITD. For center-room positions, reverberation has a minor effect on ITD, adding only very small fluctuations. For near wall conditions, where comb-filtering occurs, there is large distortion in the ITD across frequency, even though broadband correlation is unaffected. These "distortions" may contribute to difficulties in localizing sources in the reverberant space.

Overall, reverberation in which there is no single reflection of a level near that of the direct sound causes minor distortions of interaural cues, consistent with simply adding variability and random noise and filling in spectral notches of the spectral features. From these distortions, some degradation of left/right and up/down perception is expected for sources near the center of the room. Reverberation off a nearby a wall (causing a reflection of amplitude near time close to that of the direct sound) causes non-random, frequency-dependent changes in binaural cues (due to comb-filtering) and distortion of near-reflection-ear magnitude spectrum. Due to these distortions, degradation of left/right and huge degradation of up/down perception is expected for localization performance when a subject is near a wall. Past studies have shown that localization with reverberation improves with practice, implying that the acoustic cue distortion may be overcome with training [7, 8].

References

[1] Brungart, D. S., W. M. Rabinowitz. "Auditory localization of nearby sources. Head-related transfer functions." *Journal of the Acoustical Society of America* 106 (1999): 1465-1479.

[2] Brungart, D. S.. "Auditory Parallax Effects in the HRTF for Nearby Sources." In *The IEEE Workshop on Applications of Signal Processing to Audio and Acoustics: 1999 Proceedings* held in New Paltz, New York 17-20 October 1999.

[3] Brungart, D. S.. "Preliminary Model of Auditory Distance Perception for Nearby Sources."

[4] Butler, R. A., R. A. Humanski. "Localization of Sound in the vertical plane with and without high-frequency spectral cues." *Perception and Psychophysics* 51 (1992): 182-186.

[5] Hofman, P. M., J. G. A. Van Reswick, A. J. Van Opstal. "Relearning sound localization with new ears." *Nature Neuroscience* 1 (September 1998): 417-420.

[6] Santarelli, S.. "Auditory Localization of Nearby Sources in Anechoic and Reverberant Environments." (2000)

[7] Mershon, D. H., W. L. Ballenger, A. D. Little, P. L. McMurtry, J. L. Buchanan. "Effects of Room Reflectance and Background Noise on Perceived Auditory Distance." *Perception* 18 (1989): 403-416.

[8] Shinn-Cunningham, B.. "Learning Reverberation: Consideration for Spatial Auditory Displays." Proceedings of ICAD (2000)

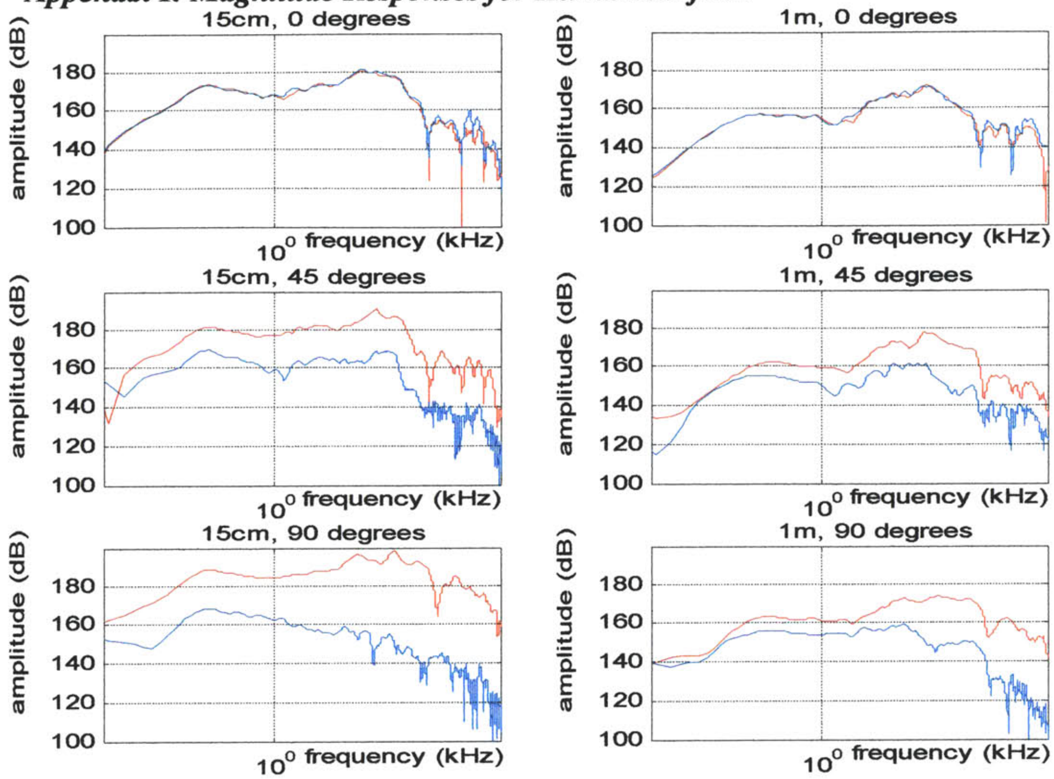
[9] Zahorik P.. "Limitations in Using Golay Codes for Head-Related Transfer Function Measurement." *Journal of the Acoustical Society of America* 107 (2000): 1793-1796.

[10] Schroeder, M.R.. "Integrated-Impulse Method Measuring Sound Decay Without Using Impulses." *Journal of the Acoustical Society of America* 66 (1979): 497-500.

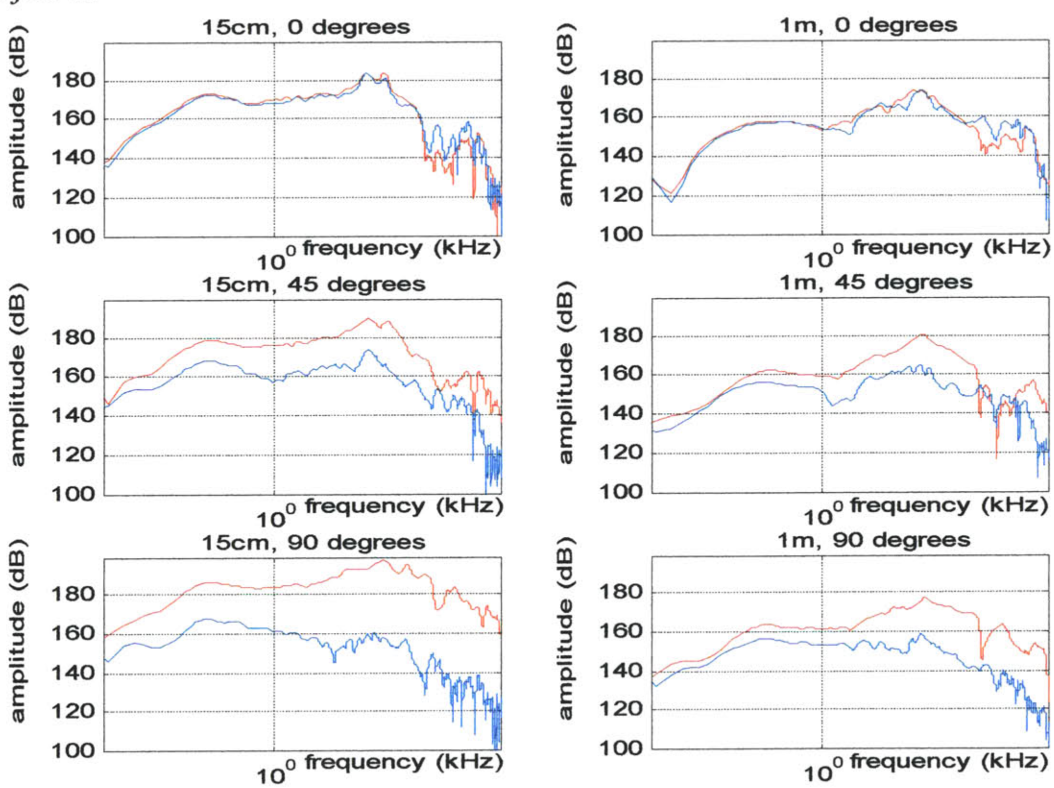
[11] Wightman, F. L., D. J. Kistler. "Headphone Simulation of Free-Field Listening. II: Psychophysical Validation." *Journal of the Acoustical Society of America* 85 (1989): 868-878.

[12] Shinn-Cunningham, B., S. Santarelli, N. Kopco. "Tori of Confusion: Binaural Localization Cues within Reach of a Listener." *Journal of the Acoustical Society of America* 107 (2000): 1627-1636.

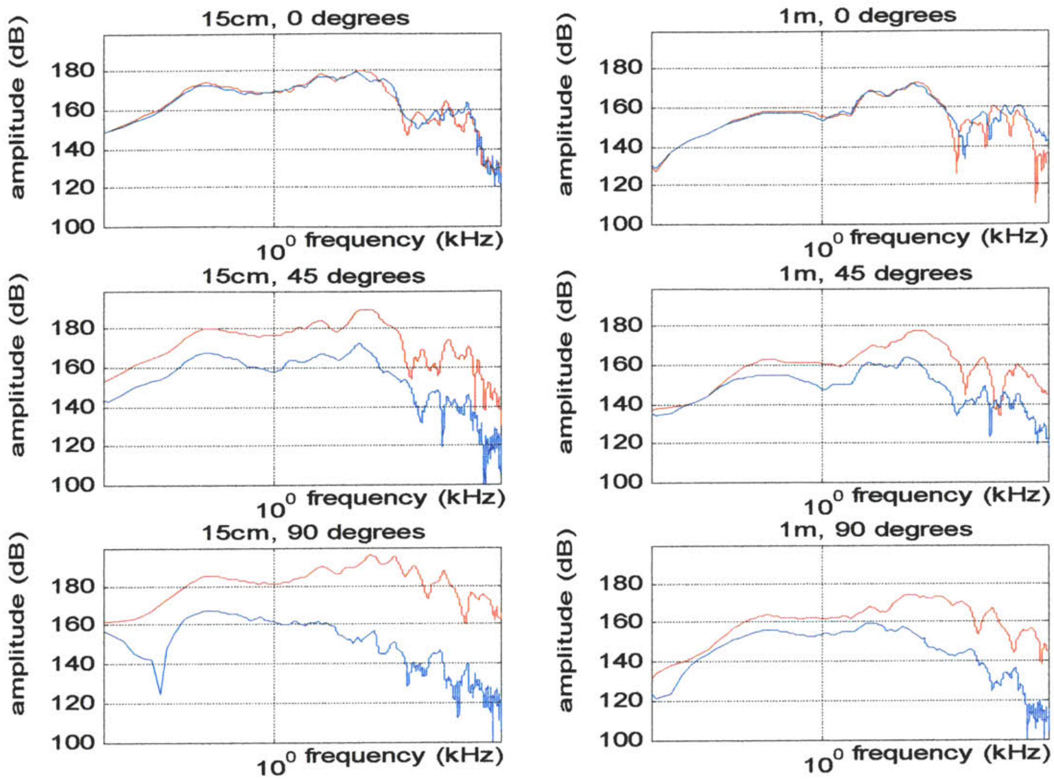
Appendix I: Magnitude Responses for Human subjects



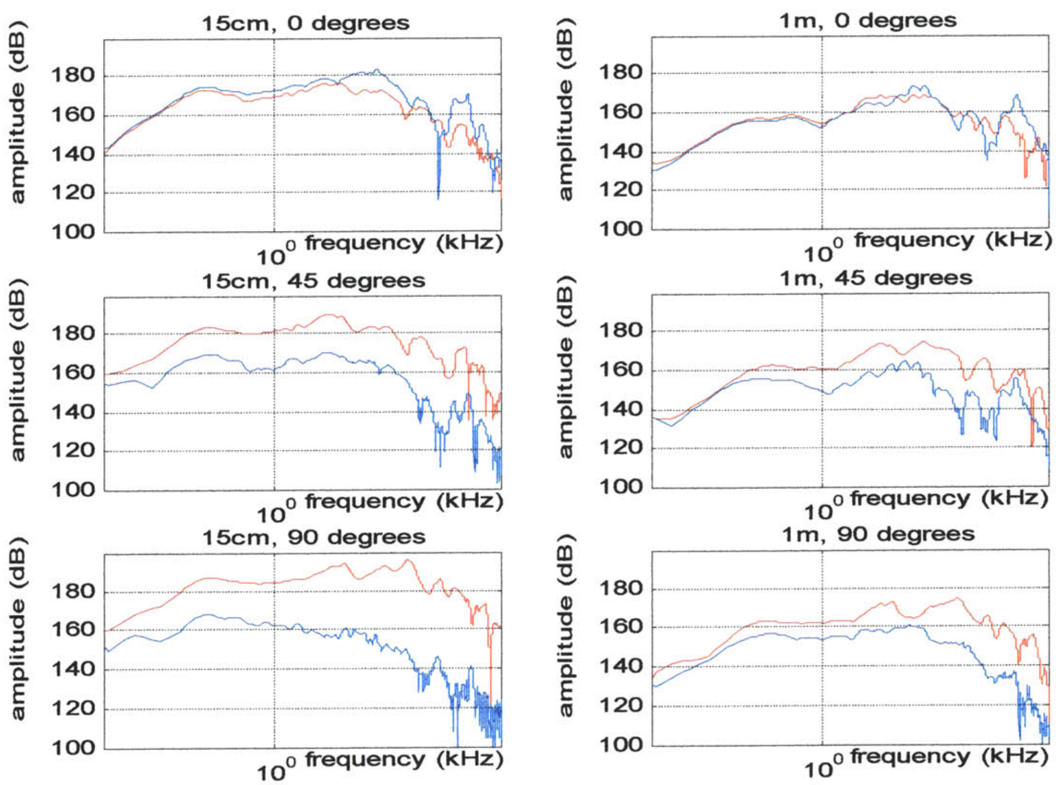
Subject AJ



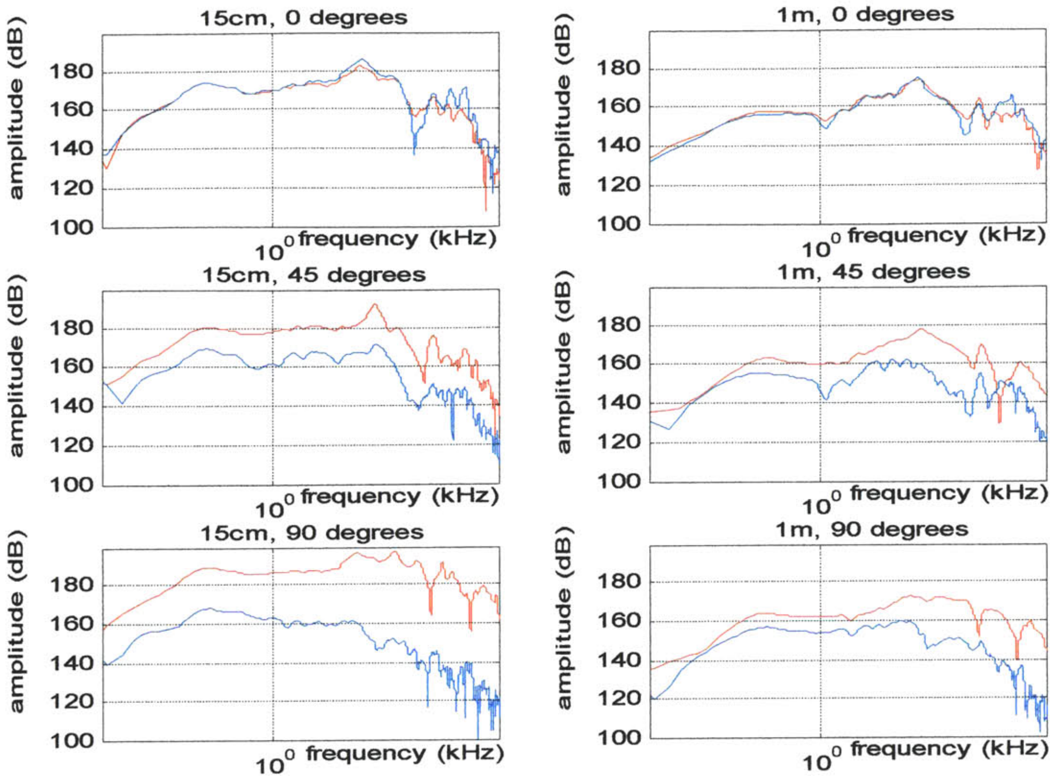
Subject AL



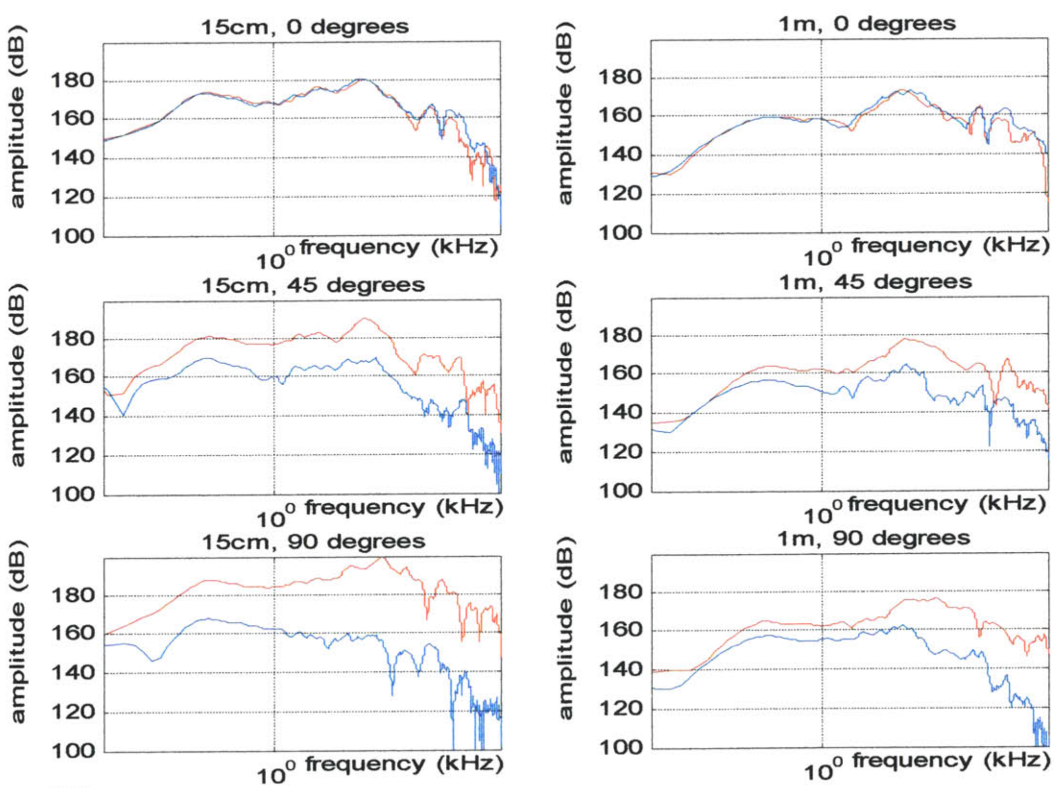
Subject AP



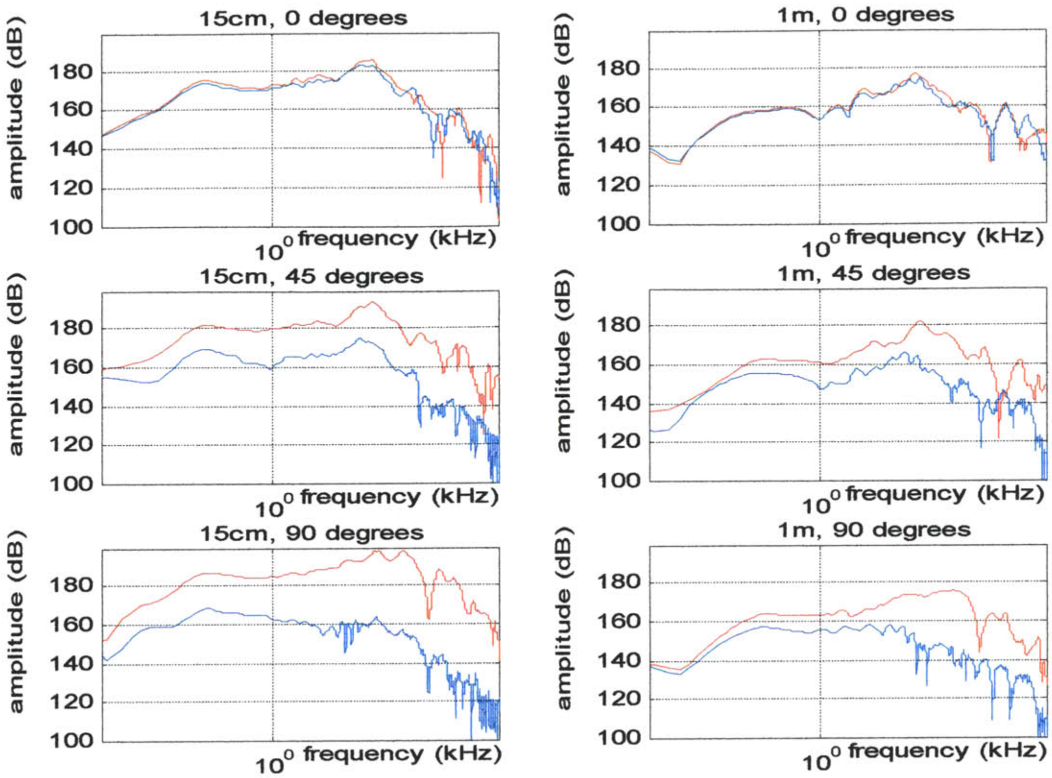
Subject BJC



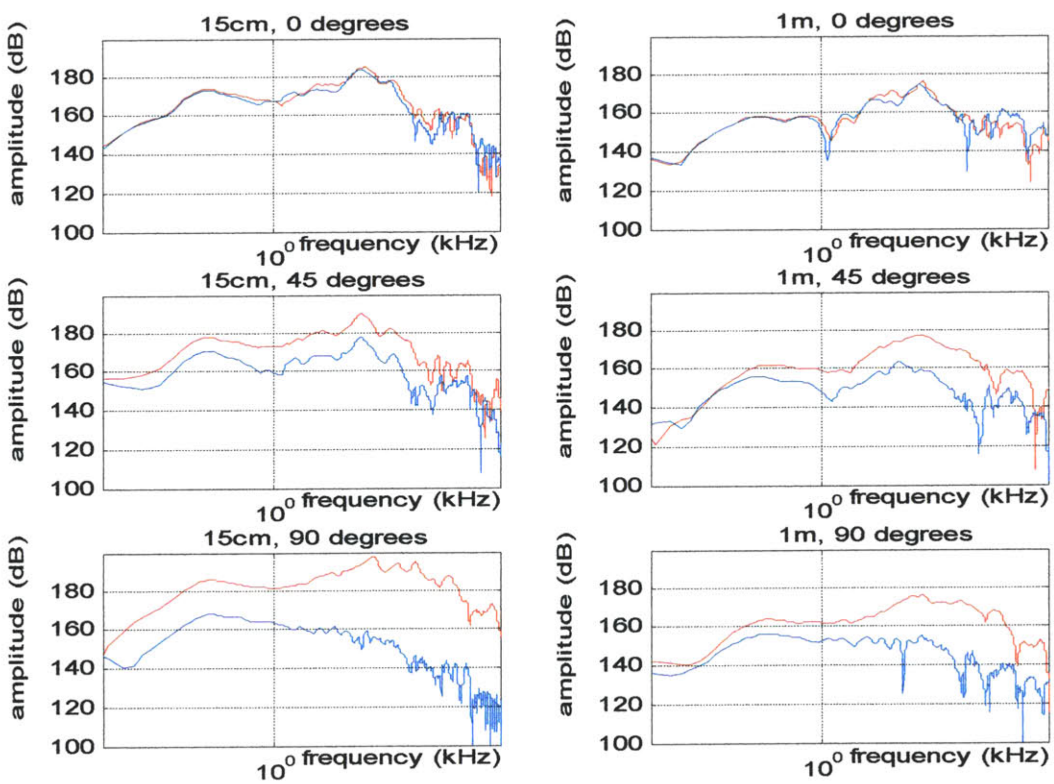
Subject DR



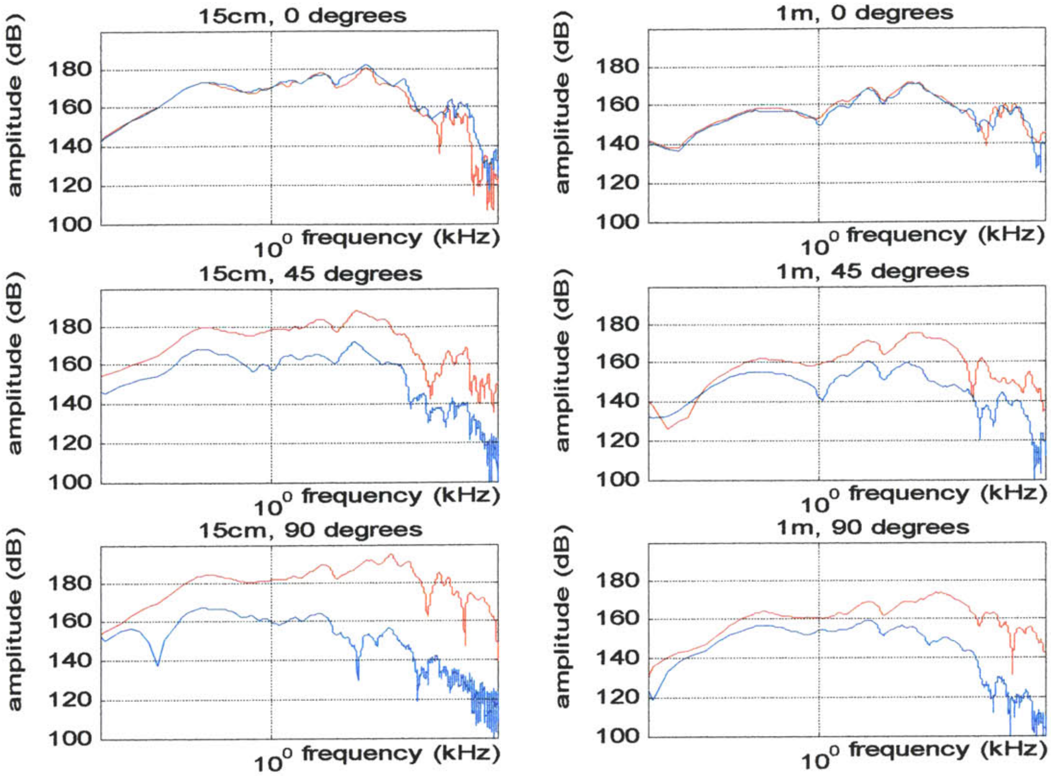
Subject EK



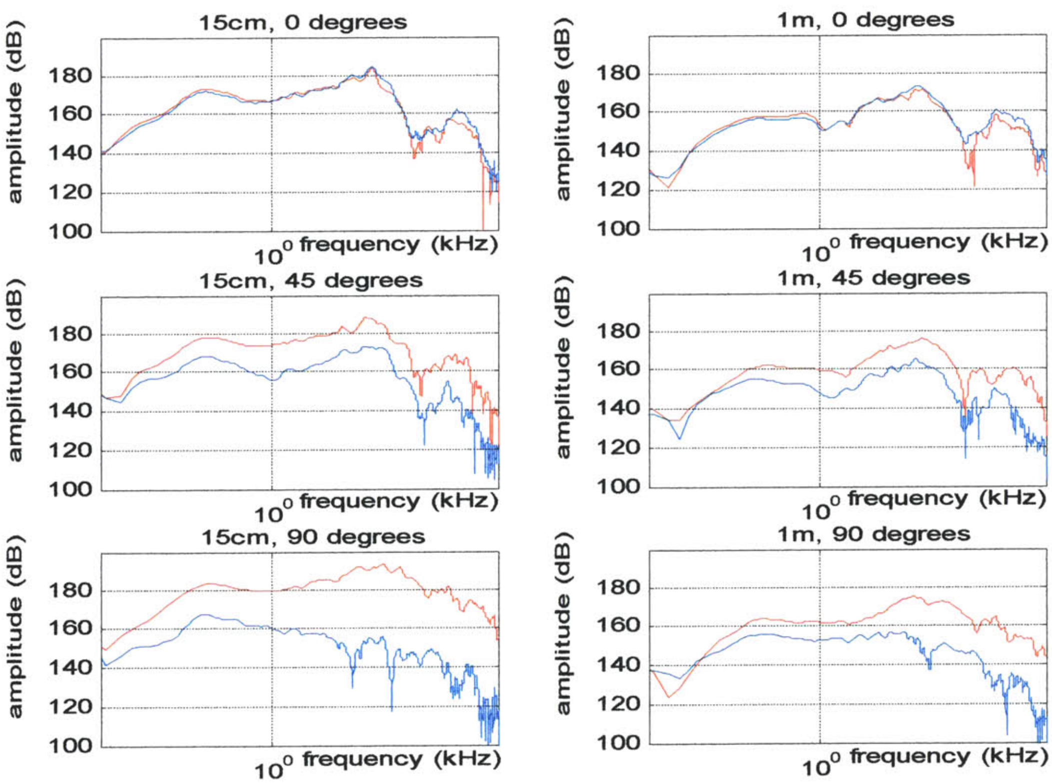
Subject JJS



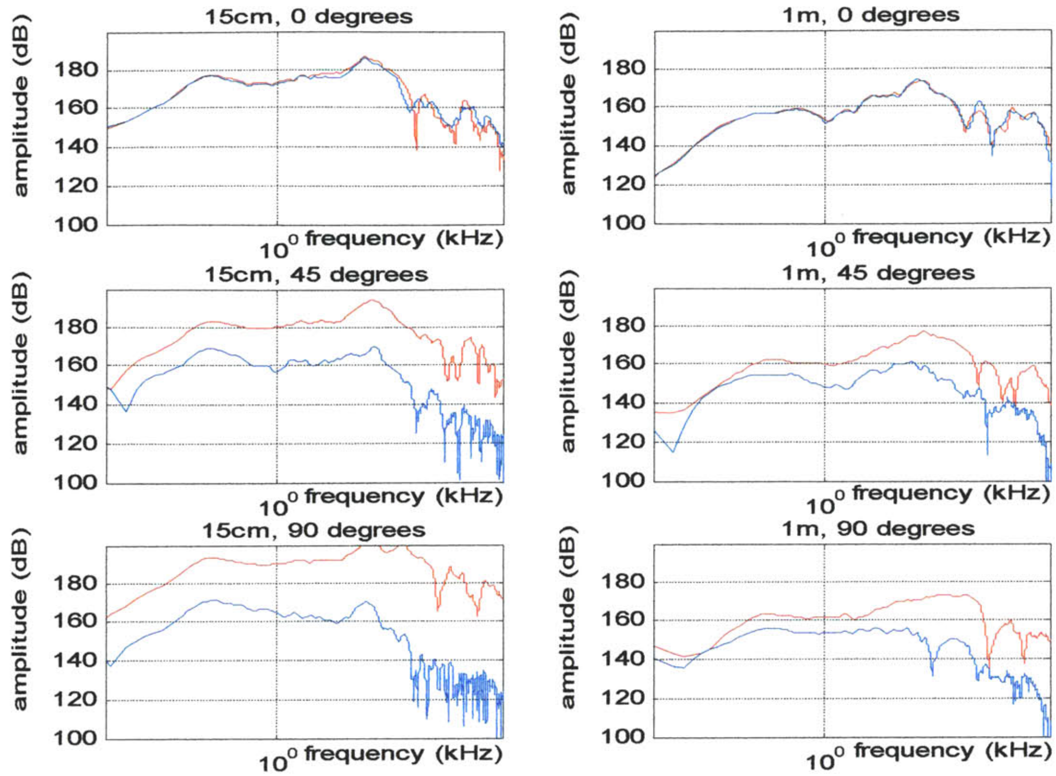
Subject LG



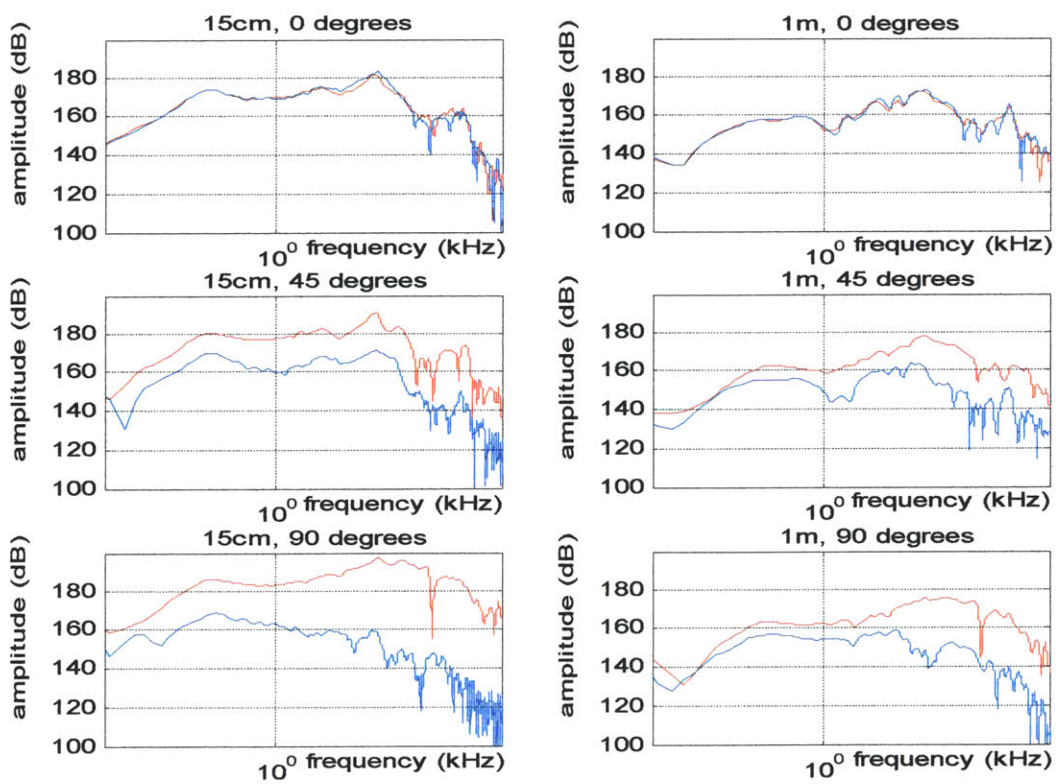
Subject MA



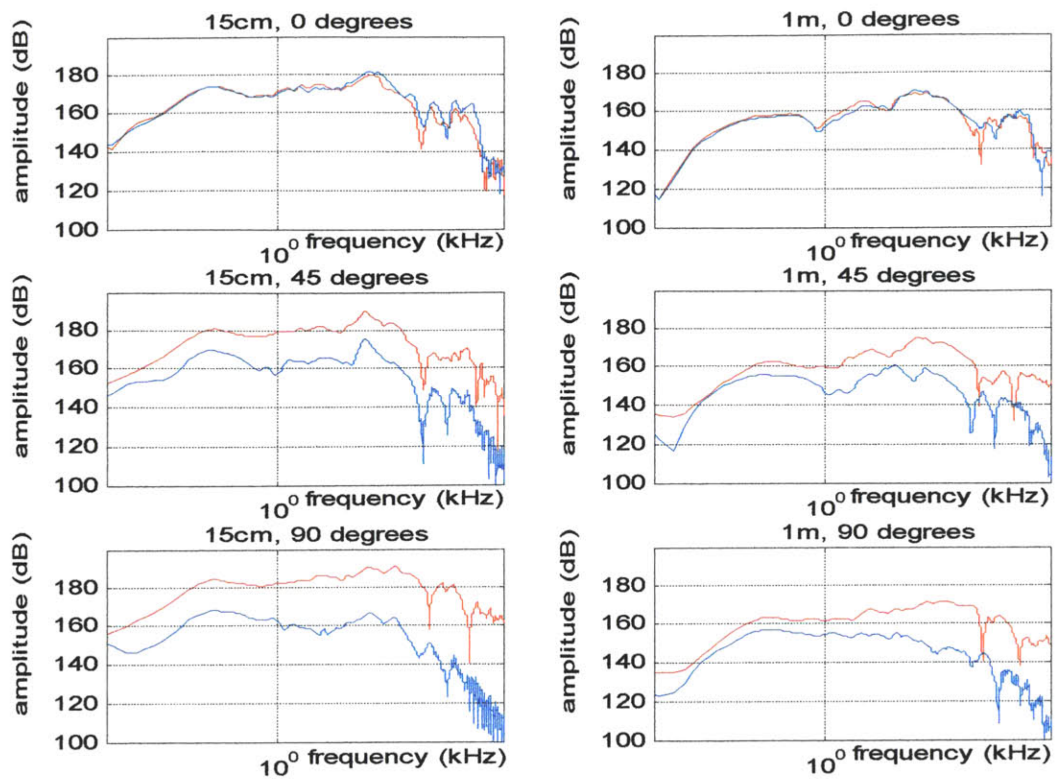
Subject NB



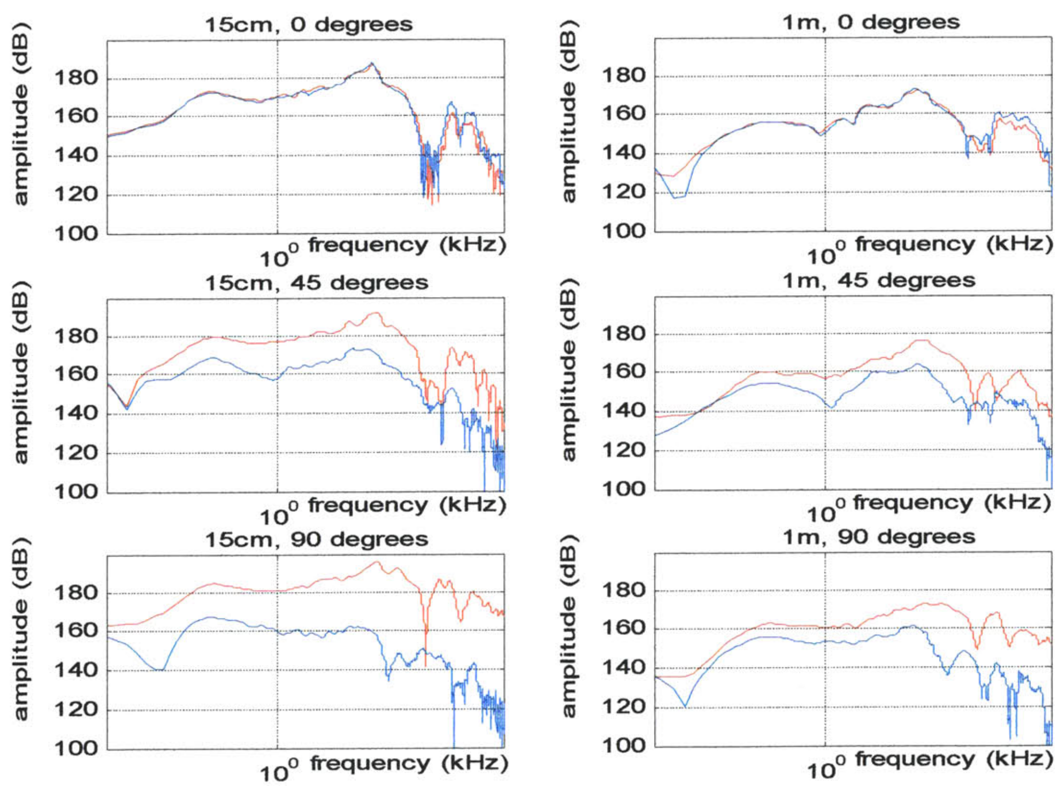
Subject NK



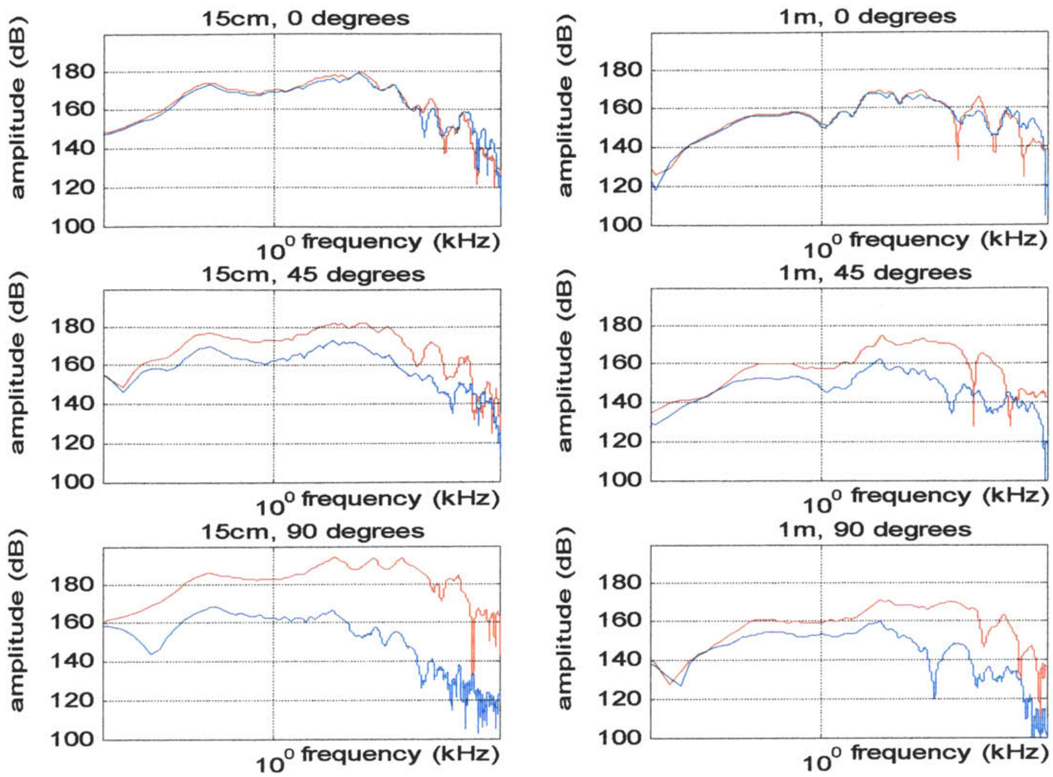
Subject OP



Subject SAH



Subject VB


15cm, 0 degrees
1m, 0 degrees
15cm, 45 degrees
1m, 45 degrees
15cm, 90 degrees
1m, 90 degrees
amplitude (dB)
frequency (kHz)
100
120
140
160
180
10^0

Subject YZ