HRTF measurement on KEMAR manikin

Mengqiu Zhang, Wen Zhang, Rodney A. Kennedy, and Thushara D. Abhayapala

Applied Signal Processing Group
College of Engineering and Computer Science
The Australian National University, Canberra ACT 0200, Australia

ABSTRACT

Presenting sounds to humans in virtual environments requires convolving the free field signals with a head related transfer function (HRTF), which is a frequency response describing the filtering effects of the pinna, head and torso of a human. Sets of HRTFs are usually measured on the dummy head or human subjects at different directions in an anechoic room. This paper describes the details of an experimental HRTF measurement procedure with emphasis on the design of the test signal and the post-processing to extract HRTFs. We construct a pre-emphasized logarithmic sweep as the excitation signal which separates the nonlinear and time variant distortions from the main desired response. For the received raw data, a series of signal processing techniques are applied to determine the timing offset when the head response begins, to mitigate room reverberation, and to equalize the HRTF measurements. One of the goals of this paper is to provide details of the experimental setup. Also, we intend to publish our processed measurements in the form of a HRTF data base as a resource for the acoustic community.

INTRODUCTION

It is generally accepted that human beings auditory localization is a binaural phenomenon where perceiving a sound spatial position is on the basis of the differences between the sounds at two ears. These auditory differences include the sound pressure levels, the arrival times, and the spectrums of the signal presented at the two ears [Meyer and Neumann, 1972, Rossi, 1998, Cheng and Wakefield, 2001]. The relationship between the location of a sound source in space and the sound wave properties at the two eardrums of a human listener has been represented by the head related transfer function, or HRTF [Cheng and Wakefield, 2001].

Measuring HRTFs using human subjects or dummy heads is a laborious task. The measurement is usually performed in an anechoic room. The basic procedure is to emit a known signal via a loudspeaker placed at a specified azimuth θ, elevation ϕ, and distance r from the subject's head, and record the signal using a microphone placed in or at the entrance of the ear canal of the subject. This is followed by processing of the recorded raw data to obtain the HRTF.

Sets of HRTFs measured at different positions contain all the acoustic cues, which are usually applied to the study of binaural hearing, the sound localization, and the synthesis for virtual auditory reproduction. Considerable effort has been put into setting up HRTF databases. MIT Media Lab set up their HRTF data base in 1995 [Gardner and Martin, 1995]. Maximum length pseudo-random binary sequences were used as the source signals in the experiment. And the measurements consisted of the left and right ear impulse response performed on a KEMAR manikin at 710 different positions sampled at elevations from -40 degrees (40 degrees below the horizontal plane) to +90 degrees (directly overhead). The UC Davis CIPIC Interface Laboratory released their public-domain HRTF database which included a total of 1250 directions and impulse responses were measured for 45 subjects including 43 human subjects [Algazi et al., 2001]. The University of Maryland measured HRTFs at 1132 points for one subject based on the reciprocity principal [Grassi et al., 2003, Zotkin et al., 2006]. There are also measurements from the Listen project, which is a shared-cost RTD project in the Information Society Technologies (IST) Program of the European Commission's Fifth Framework Program. Their database provided 187 points HRIR (Head Related Impulse Response) results.

HRTF experiments are usually performed in an anechoic chamber. But in this paper we produced a feasible experimental setup for a non-anechoic situation, that is, the measurements can be made when there is some reverberation. We call this semi-anechoic. In addition, compared to the existing public domain databases, our intention is to produce a HRTF database on a more densely spaced grid, which poses an advantage for HRTF fitting and interpolating.

In this paper, the details of our experimental HRTF procedure are reviewed with emphasis on the signal processing aspects to process the measurements from a KEMAR manikin.

- Firstly, we depict our experimental facilities and the experimental procedure in Section 2. In Section 3, we elaborate how to design the test signal, which is vital to carry out the reliable and meaningful experiment.
- In Section 4, we provide the circumstantial description on tackling echoes when the measurements are performed in a reverberant environment, and then elaborate on three important steps of how to extract HRTF from raw data.
- Finally, in Section 5, we present our measurement results — as a first phase towards delivering a new public HRTF database.

HRTF MEASUREMENT

There are two methods to measure the HRTFs. One is called the direct HRTF measurement method, the other one is called the reciprocal HRTF measurement method [Zotkin et al., 2006]. The direct method [Gardner and Martin, 1995, Algazi et al., 2001, Teranishi and Shaw, 1968] is to measure the HRTFs by broadcasting test signal at different directions and recording the arrival signal at the entrance of two ears. The reciprocal method [Hunter, 1957] is based on Helmholtz' principal of reciprocity, in which the place of broadcasting and receiving facilities are exchanged. Our experimental setup is built on the direct method.

Spatial Coordinates

Commonly, HRTFs are measured at different azimuths and elevations from a fixed distance. Figure 1 depicts the spherical coordinates system in the measurement setup. We design a rotating semicircular hoop with a diameter of 3 m. A speaker is mounted on the centre of the hoop. And the KEMAR (Knowles Electronics Mannequin for Acoustics Research) manikin head

is centered at the origin of the hoop on a rotatable table. Turning the hoop defines elevation angles which denote up - down direction. While rotating the manikin gets all desired azimuthal angles which represent the left - right direction. The experiment is made in one direction at one time, by rotating the hoop to a desired elevation and then turning the manikin to a particular azimuth. The up - down direction is sampled from -30 degrees to +90 degree in increments of 5 degrees. At each elevation, a full 360 degrees of azimuth is sampled in equal sized increments. The conventional interval is 5 degrees with the elevation from -30 degrees to 75 degrees. Then, it increases to 10 degrees at the elevation of 80 and 85 degrees. And there is just 1 point to be measured at the position of directly overhead. In total we measured HRTFs at a total of 1657 directions around the dummy head.

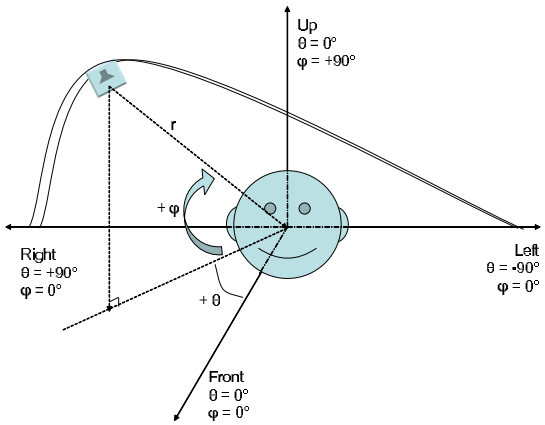

Figure 1: The Spherical Coordinates System in HRTF measurement.

Facilities

In this section, we introduce the main facilities used in our experiment.

Our experimental chamber is 3.2 metres long, 3.2 metres wide, and 2 metres high. The chamber's walls, ceiling, and floor are lined with foam to mitigate the internal reflections and also to isolate outside noise. Because the distance from speaker to ceiling and then to microphone is farther than that from speaker to hoop to microphone, the first incoming reflection would be from the hoop. So we covered the hoop with foam, which is proved to be effective in Section 4.

Our experiment is performed on a KEMAR. The manikin (Type: 45BA, made by G.R.A.S.) is especially built for simulating the diffraction and reflection characteristics of a real human, with two microphones embedded at the eardrum locations to make binaural recordings. In our experiment, it was fitted with a small left pinna (G.R.A.S KB0061) and a large right pinna (G.R.A.S KB0065). The manikin is fixed on a rotatable plate which can be automatically controlled by the computer to precisely set the degree of azimuth.

The microphones (Type: 40AG, made by G.R.A.S.) are both 1/2-inch precision reference microphone for laboratory use. The Type 40AG microphone is a polarized pressure microphone with a large dynamic range (160 dB re. 20 μPa). As a pressure microphone, the Type 40AG measures the sound pressure at the location of its diaphragm. It has a flat pressure-frequency response over its entire working frequency range from 3.15 Hz to 20kHz. The pressure microphone will also include the disturbing effects of its presence in the sound field. These are minimal at low frequencies. But at higher frequencies the effects of reflections and diffractions increase which must be accounted for.

The G.R.A.S. preamplifier (Type: 26AC), which is connected to the microphone, is a small robust unit optimized for acoustic measurements with a very low inherent noise level ($\leq 6\mu$Vrms), and frequency response from below 2 Hz to above 200 kHz. The power supply of the preamplifier is from the G.R.A.S. power module (Type: 12AA). There is a linear-response high-pass filter in each channel of the power module with cut-off frequency at 20 Hz for reducing unwanted low-frequency signals.

The equipment for D/A and A/D conversion is a National Instruments USB-6221. It is a USB high-performance M Series multifunction data acquisition (DAQ) module optimized for superior accuracy at fast sampling rates. There are 16 analog inputs (16 bit) with sampling rate up to 250 kb/s, and two outputs (16 bit) with sampling rate up to 833 kb/s. It also shows perfect I/O performance. Its output impedance is just 0.2 Ω which prevent it from lowering the magnitude of the output signal when connecting with loudspeaker. And its input impedance is higher than 10 $G\Omega$ in parallel, which is vital to avoid influencing the signals from microphones.

The loudspeaker we used in the experiment to broadcast the test signal is a CREATIVE LX270 satellite loudspeaker with frequency response from 40 Hz to 20 kHz.

Experimental Procedure

The testing procedure is controlled by a customized MATLAB program running in a DELL desktop computer. The program is responsible for the test signal production, measurement facilities control, data acquisition and signal processing. The computer is placed outside the chamber to eliminate any unwanted noise induced by the computer hardware and its operator. Then a USB cable through the wall of the chamber is used to connect the computer and the NI USB-6221 data acquisition. One output of the NI device is connected with the amplifier and then to the speaker, which allow the computer to send the test signal to speaker. And two inputs of the NI device are linked with dual channel power module, which enables the computer to receive the signals from the microphones.

Figure 2 shows our HRTF experimental procedure. After setting the desired position, MATLAB generates the test signal, which will be elaborated in next section. Firstly, the digital signal is converted to analog by the NI USB-6221 D/A component at the input sample rate of 44.1 kHz. Then the output unit of NI device passes data to the amplifier which drives the loudspeaker. Finally the loudspeaker broadcasts the test signal. For the receiver part, two microphones fitted at the entrance of KEMAR's ears capture the arrival signal from the speaker. Each microphone is closely connected with a preamplifier. After amplified by preamplifiers and high pass filtered by power module separately, the binaural signals are send to two input channels of the NI data acquisition. Then data acquisition records the two channel data at the sampling rate of 44.153 kHz and sent to computer in which MATLAB program proceeds a series of signal processing to extract HRTFs from recorded raw data. Eventually MATLAB can present the HRTFs in different formats through its graphical capability.

DESIGN OF TEST SIGNAL

Over the past century a number of different signals as exitation [Muller and Massarani, 2001] have been developed to measure the transfer functions, such as stepped sine, impulses, maximum-length sequences, periodic signals of length 2^N, and nonperiodic sweeps. Among those signals, sweep, especially

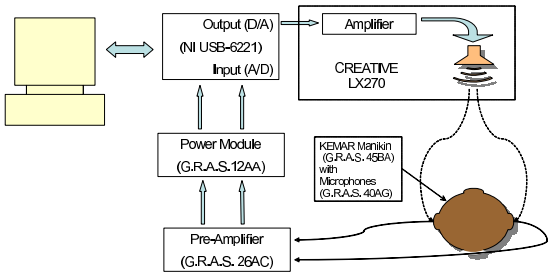

Figure 2: HRTF measurement Procedure

the exponential sweep, is the better choice than others due to the following reasons. Firstly, using a sweep can isolate all harmonic distortion products from the acquired IR, practically leaving only background noise as the limitation for the achievable SNR. The sweep can thus be fed to the loudspeaker with considerably more power without introducing artifacts in the acquired impulse response. Secondly, sweep-based measurements are also considerably less vulnerable to the deleterious effects of time variance. Moreover, the exponential sweep has an exponentially increasing sweep rate which gives itself an advantage of providing more energy at low frequency signals versus the traditional linear sweep.

Profile of the Test Signal

In our experiment, the test signal is generated by MATLAB. The frequency of the chirp varies exponentially as a function of time from 300 Hz to 20 kHz

$$f(t) = f_0 \times \beta^t, \tag{1}$$

where f_0 is the starting frequency 300 Hz (at $t = 0$), and β is the rate of exponential increase in frequency. The corresponding time-domain function for a sinusoidal exponential chirp is

$$x(t) = \sin(2\pi \int_0^t f(\tau) d\tau). \tag{2}$$

Then, applying (1) to (2), we have

$$x(t) = \sin(2\pi f_0 \int_0^t \beta^\tau d\tau) = \sin(\frac{2\pi f_0}{\ln\beta}(\beta^t - 1)). \tag{3}$$

The sweep time of a single chirp is 4.2ms depending on the distance between speaker and the head of manikin. In our experiment it is 1.5 m. Then the time for the chirp traveling from speaker to microphone is about 4.4 ms. So, we choose the chirp duration 4.2 ms to avoid an overlap between the reflections caused by our desired dummy head and that produced by surrounding walls or other auxiliary testing equipment. Thus, there are 185 samples when the chirp is converted to digital signal at sampling rate of 44.1 kHz.

Pre-emphasis

In almost any acoustical measurement, it is not advisable to use an excitation signal with white spectral content which tends to lead to a poor SNR at low frequency [Muller and Massarani, 2001]. Two factors account for this matter. One is the ambient noise which mainly lies in low frequency region. Thereupon strengthening the lower frequency signals is essential for getting an approving SNR. The other is the decrease of sensitivity of the acoustical equipment to lower frequency sound. This means that human beings hearing or acoustical instruments are far more likely to be damaged by treble than bass with the same energy. Hence, to get an accredited testing result without causing any damage, a strong emphasis on lower frequency signal is required [Muller and Massarani, 2001].

Although the exponential chirp has the unique property of self emphasis at lower frequencies, it is not enough for establishing a reliable measurement. We approach this matter by passing the original signal through a particular filter, which can enhance the magnitude of the frequencies of interest and leave other components untouched. There are two main kinds of digital filter: finite impulse response (FIR) filter and infinite impulse response (IIR) filter. We prefer FIR filter due to its few advantages over IIR filter. The first is its property of being able to realize a linear phase response, which means that there is no phase distortion and the input signal will be delayed by a constant time when it is transmitted to the output. Then no significant changes in shape will be observed on the output signal [Shenoi, 2006]. This is the main reason why we choose FIR filter. The second is that FIR filters are inherently stable because all the poles are located at the origin and thus are located within the unit circle [Shenoi, 2006]. The third advantage is that the unit impulse response of a FIR filter is a finite long sequence [Shenoi, 2006]. It is reasonable to assume in most practical cases that the value of the samples of signals of interest becomes almost negligible after a finite time.

The key points for realizing such a filter by MATLAB is to properly determine the cut-off frequency and the magnitude response. The range of human hearing is generally considered to be 10 Hz to 20 kHz, but it is far more sensitive to sounds between 1 kHz and 4 kHz. For example, listeners can detect sounds as low as 0 dB SPL (sound power level) at 3 kHz, but require 40 dB SPL at 100 Hz (an amplitude increase of 100) [Smith, 1997]. So, we decide to emphasise the signals with frequency under 1 kHz and leave the signals from 1 kHz to 20kHz unchanged. Then the emphasizing pass band is [0 0.045], the unalterable pass band is [0.06 0.91], and the stop ban is [0.92 1] (frequencies here are relative or normalized to frequencies at the sampling rate of 44.1 kHz). Three different gains are given to the signals in the emphasized pass band. The lower the frequency of the signal is, the higher the gain is given to the signal. The other design factor is the order of the filter. It is chosen on a trial-and-error basis until the specifications are satisfied [Cheng and Wakefield, 2001]. In this case, the order should be 128 as the best tradeoff between specifications and the large amount of computation. After determining these parameters, we use the MATLAB function of `Firls` to realize it.

Figure 3 shows the filter's magnitude response and phase property. As shown, there is more than 10 dB gain in the region of lower frequency from 0 to 1 kHz while that of frequency greater than 1 kHz gets no gain. The phase in pass band is linear which enables the signal unchanged in shape after emphasized. Figure 4 demonstrates the comparison between original chirp and the pre-emphasized one. As shown in Figure 4 (c), the initial phase of the filtered signal remains the same as the original signal but the amplitude in the lower frequency is around 4 times larger.

The Integrated Test Signal

Averaging is an effective way to improve the signal to noise ratio (SNR). Therefore, 16 pre-emphasized chirps are broadcasted. There is a 80 ms break between chirps (3528 zero samples at the same sampling rate of 44.1 kHz). The length of break is determined experimentally. From results of experiments, we determined that reverberant echoes die out after 80 ms from the time that the signal was sent out. So, one test session consists of 16 exponential upsweep chirps with 80 ms break between two. The whole test signal lasts 1347.2 ms, thereby having 59408 samples at the sampling rate of 44.1 kHz. The total of 16 emphasized chirps and breaks are then integrated into one test signal as shown in Figure 5. By us-

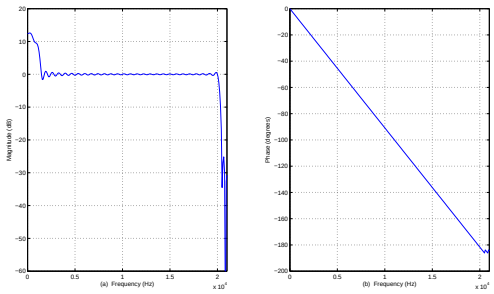

Figure 3: Pre-emphasis Filter (a) Frequency Response (b) Phase Property

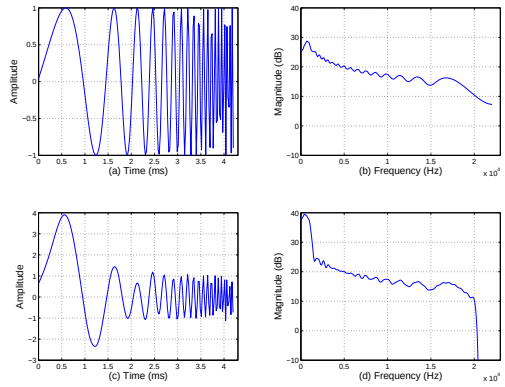

Figure 4: Test Signal (a) Original Chirp in Time Domain (b) Spectrum of the Original Chirp (c) Pre-emphasized Chirp in Time Domain (d) Spectrum of the Pre-emphasized Chirp

ing such assembled signal, data acquisition output and input channels are to be triggered once in one measurement. This is beneficial to avoiding data missing and asynchronism caused by starting up and closing down the device many times.

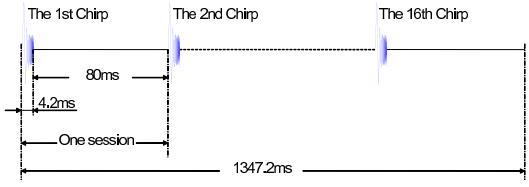

Figure 5: The Test Signal

POST-PROCESSING TO EXTRACT HRTFS FROM RAW DATA

The impulse response of the whole system consisting of the data acquisition, the amplifier, the speaker, the microphone, the pre-amplifier, and power module. The entire received data of each channel measured at one direction are 59408 samples corresponding to the broadcasted test signal. But we discard the first received chirp because there might be some missing data at the beginning due to the delay of A/D component. Then, a series of signal processing done on the following 15 received chirps in computer by MATLAB are of great importance in extracting HRTFs from the raw data. Figure 6 shows the post-signal processing procedure. The processing is meaningful only on the premise of successful alignment.

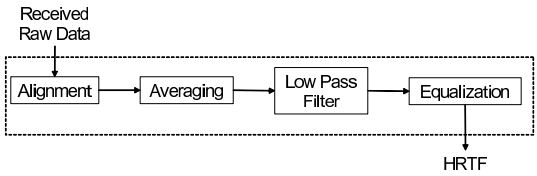

Figure 6: Post-Signal Processing Procedure

Alignment and Averaging

As discussed in Section 3, the integrated sequence length of the test signal is 59408. We broadcast it once at one designed position. However 2×59472 samples are actually captured by both left and right ears as recorded by the two of input channels of the NI data acquisition. That is because the sampling of NI data acquisition input channels is a little bit higher than that of output channels by about 0.12%. Then, each received sequence can be divided evenly into 16 sections, where every section has 3717 samples instead of the 3713 emitted samples. In order to avoid missing useful information caused by the unpredictable delay of input channels, we discard the first 3600 samples. From the 3601^{th} sample, we redivide the following 55755 samples into 15 sections. Thus, the new 15 sections must contain corresponding 15 impulse responses each. The next crucial question left is in where those responses are exactly located. And the response time may vary among 15 sections due to the effects of the testing system or the transmission path. From Figure 7, we can see that the received signals have different response delay. If we do not align them first, the averaging result will be meaningless. Therefore, alignment is required at the very beginning in the signal processing to synchronize their steps without introducing further noises when doing averaging.

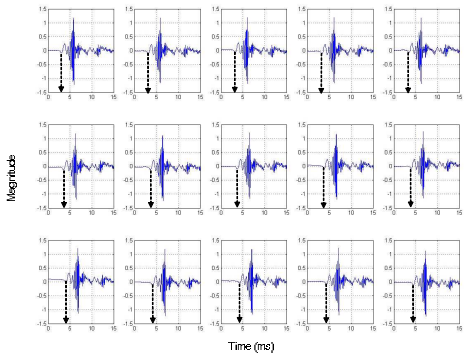

Figure 7: Received Signals with Different Response Delay

Because of the transmission delay and the system delay, the first sample corresponding to the start point of test signal will emerge at the point of N, as shown in Figure 8. Calculating the correlation C between the test signal s and the received signal r through,

$$C_j = E(s - Es)(r_j - Er_j), \tag{4}$$

where j is from m to $m+199$ (See Figure 8), we can determine the start point of N. This cross correlation must be calculated 15 times to find out every first response sample relevant to each incidence.

Before we fix the end point of the impulse response, another important issue should be clarified. The question is if the re-

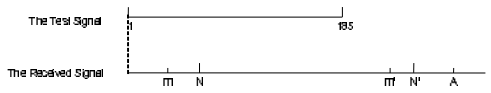

Figure 8: Alignment Procedure (N — The Start Point of the Received signal's First Sample; N′ — N′ = N + 184; A — The Start Point of the Reverberation Signal's First Sample; m — The Start Point for Calculating the Correlation to Confirm N; m′ — The Start Point for Calculating the Correlation to Confirm A)

ceived signal of our great interest is contaminated by the room reverberations. We can determine the time of the room reflections by doing the experiment of system response (the same procedure as HRTF measurement without dummy head) because there are no useful reflections to capture in this circumstance. The test signal is 185 samples long. So, the system impulse response should be as the same length as the test signal. That is the end point A should be at the position of N'. We can also calculate the correlation C' between the data began from the m'^{th} sample and also the transmitted chirp by

$$C_i^{'} = E(s - Es)(r_i - Er_i), \tag{5}$$

where i is from m' to $m' + 30$, to determine the position of A. Then we get the start point of the first unwanted reverberation, thereby confirming the time of room reverberation.

If A is less than N', few reflections samples will be involved in the system response and of course in the HRTF sought. So the position of A should be later than N' as possible. Before we covered the hoop carrying the speaker, A is in the position of $N + 178$, which is less than N'. After we lined the hoop, A goes to $N + 195$, 10 samples later than N'. 10 samples mean that the first sample of reflection came to microphones is 0.23ms later than the whole broadcasted signal reached the microphones. And the distance for the sound traveling within the time of 0.23ms is 7.8cm. Because the time of lower frequency components reaching the dummy head is much earlier than that of higher frequency components, the contribution to HRTF at the end of broadcasted signal is mainly from higher frequency components, which is the filtering effect by the pinnae. Comparing with the dimension of pinnae, 7.8cm is longer than the diffraction path. Then we won't miss any useful information when discarding samples of the room reflection.

And now we can make sure where the end point of HRTF is. Generally, the reverberation times at the highest frequencies are usually much shorter than that encountered at low frequencies. Therefore, it manifests that capturing can be stopped almost immediately after excitation signal is swept through [Muller and Massarani, 2001]. As discussed above, the interval between the test signal and the room reverberations is long enough for microphones to capture all cues of HRTF. So we discard the samples after $N + 194$ for the purpose of accurately fixing the length of the data carried full HRTF information without any unwanted reverberation. Therefore the total of HRTF data is 194 samples in our experiment.

As shown in Figure 9, 15 signals were synchronized to the same step. Then averaging can be done after aligning 15 measurements. Alignment and averaging are both done in time domain.

Low Pass Filtering

As mentioned in Section 2, the power module placed just ahead of NI data acquisition is a linear-response high-pass filter with

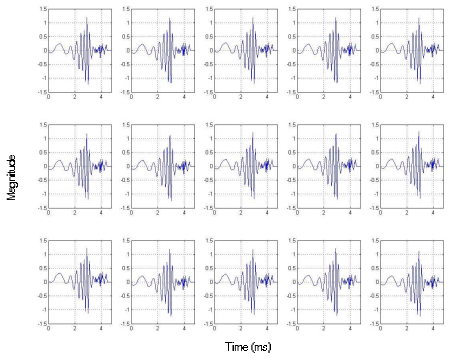

Figure 9: Signals after Alignment. Related to Figure 7, same signals time aligned and truncated.

the cut-off frequency of 20 Hz. It aims to mitigate most of low frequency noises, such as noises caused by air movement.

For noises with frequency above 20 kHz, we design a low-pass filter with cut-off frequency of 20.5 kHz. We also choose to use the FIR filter on account of its advantages discussed in previous section of the design of pre-emphasis filter. But the implementation is different. In this situation, our approach lies in designing by window function which requires us to set desired pass band ripple and stop band attenuation first. The maximum of 3% pass band ripple and the minimum of 40 dB stop band attenuation are deemed to be acceptable in our experiment. Then we select the common used kaiser window to realize the low-pass filter. What we should do next is to estimate the order of the filter by using MATLAB function of `Kaiserord(f,a,dev,fs)`. After the order of the filter has been obtained, the following step is to find the value of the unit impulse response of the filter by using MATLAB function of `fir1(n,Wn,window)`.

Figure 10 describes the filtering properties of the designed filter and draws a comparison between filter′s input and output signals. Figure 10 (a) shows that its pass band ripple and stop band attenuation meet our requirements very well, and its transition band is even less than 1 kHz. As shown in Figure 10 (c) and (d), the signals with frequency under 20 kHz are identical while the spectrum of others is smoothed, which indicates that the filter is effective. At this stage, we get the raw HRTF but it is colored by the undesired frequency response of the measurement system. The equalization must be executed to remove the embedded effects and we consider this issue next.

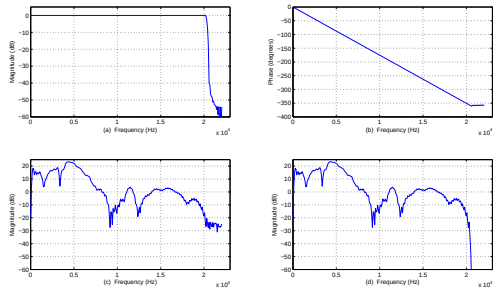

Figure 10: Low Pass Filtering (LPF) (a) Magnitude Response of LPF (b) Phase Property of LPF (c) Signal before Filtering (d) Signal after Filtered

Equalization

No matter what testing system is used to measure the HRTFs, equalization must be performed on the received data to remove the effects due to the measurement apparatus, such as speaker and microphones. There are two approaches to tackle this problem. One is equalization with respect to a reference direction (typically to be 0° azimuth and 0° elevation), where each HRTF measured at directions other than reference direction is divided by the reference HRTF [Jot et al., 1995]. This method is no doubt effective on eliminating all effects induced by transducer and receiver components. Unfortunately it also removes some effects of our great interest caused by dummy head. The other method is to measure the transfer functions of the measurement apparatus with precision sound calibration equipment, and the inverses of these functions can be used to equalize the raw HRTF result [Cheng and Wakefield, 2001]. But not every laboratory is fully equipped to do such calibrations on every device used in the experiment.

Our approach is to measure the system response by moving away the manikin and leaving the microphones in the same position as that in the manikin. The same signal processing procedures are also done on the received data. Then the measured transfer function just includes frequency responses of the data acquisition, the amplifier, the speaker, the microphone, the pre-amplifier, and power module. What we should do next is divid the raw HRTF by the system response.

Mathematically, we can describe this procedure as follows. Let $x(n)$ be the known test signal to be broadcasted at the appointed direction. Let $s(n)$ be the system transfer function. Let $h_{l,\theta,\phi}(n)$ and $h_{r,\theta,\phi}(n)$ be the unknown left and right ear transfer function of interest, respectively. Let $\hat{h}_{l,\theta,\phi}(n)$ and $\hat{h}_{r,\theta,\phi}(n)$ be the measured left and right ear raw HRTFs, respectively. Finally, let $\hat{s}(n)$ be the measured system response. Then when we perform the experiment on the manikin, in time domain we have

$$\hat{h}_{l,\theta,\phi}(n) = x(n) * s(n) * h_{l,\theta,\phi}(n) \tag{6}$$

or, equivalently, we rewrite it in frequency domain as

$$\hat{H}_{l,\theta,\phi}(k) = X(k) \cdot S(k) \cdot H_{l,\theta,\phi}(k). \tag{7}$$

When we perform the experiment without manikin, we have

$$\hat{s}(n) = x(n) * s(n) \tag{8}$$

and

$$\hat{S}(k) = X(k) \cdot S(k). \tag{9}$$

Obviously, the following computation shows how we get the final HRTFs,

$$H_{l,\theta,\phi}(k) = \frac{\hat{H}_{l,\theta,\phi}(k)}{\hat{S}(k)} = \frac{X(k) \cdot S(k) \cdot H_{l,\theta,\phi}(k)}{X(k) \cdot S(k)}. \tag{10}$$

Figure 11 presents the HRTFs of left ear HRTF at the direction of 0° azimuth and 0° elevation before and after equalization: (a) is the measured HRTF, $\hat{H}_{l,\theta,\phi}(k)$, (b) is the measured system response, $\hat{S}(k)$, (c) is the left ear HRTF at the direction of $\theta = 0°, \phi = 0°$.

RESULTS

We measured the HRTF at 1657 directions on a KEMAR manikin as discussed in Section 2. So, there are 2×1657 pairs of HRTFs responses in our database. We will present the structure of HRTFs in the frequency and time domains.

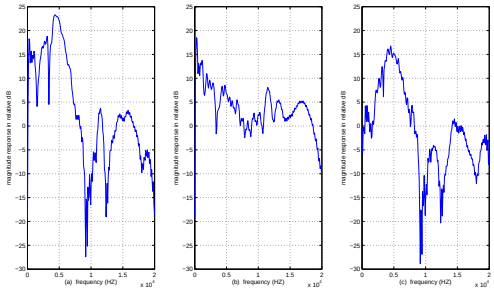

Figure 11: Equalization (left ear HRTF measured at direction of $\theta = 0°, \phi = 0°$)(a) Rough HRTF (b) System Frequency Response (c) Final HRTF after Equalization

Frequency Domain Representation of HRTFs

The main spectral characteristics of HRTF, such as notches and peaks, are easily presented in frequency domain. Figure 12 depicts the magnitude response of HRTF measured at the direction of $\theta = 0°$ and $\phi = 0°$ as a function of frequency. There are three spectral notches at higher frequencies. The deepest notch appears at the frequency of 9 kHz. The second deep one presents at about 12.5 kHz, and the third at 18 kHz. The position of peak is located at frequency 4.5 kHz. These features matched well the results presented in [Teranishi and Shaw, 1968] and [Lopez-Poveda and Meddis, 1996]

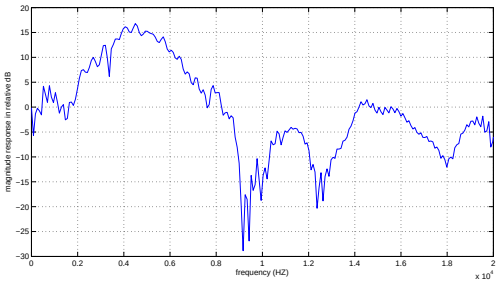

Figure 12: Main Spectral Characteristics of HRTF

Figure 13 shows frequency domain representations of left ear and right ear HRTFs measured in the median and horizontal plane, respectively. As shown in Figure 13, HRTFs show less fluctuation at the elevations above 30°. The peaks at 4.5 kHz appear almost the same position as denoted by the dots in the figure. The positions of the notch at 9 kHz, shown by stars, change more significantly from 8.5 kHz to 10.5 kHz. Donuts show the position of notch at higher frequencies, left ear of which move upward in frequency as elevation decreases from 10 ° to -30 ° while those of the right ear move downward.

Diffraction effects can be easily observed in Figure 14 and Figure 15, where measured HRTFs are drawn as a function of azimuth angle. Note that the contralateral HRTFs (which correspond to the azimuths +90° and -90° for the left ear and right ear, respectively), show more rippling than the ipsilateral ones. The ripples are obvious especially at a higher frequencies. And the rippling phenomena exist in the HRTFs measured at elevations 60° to 120° for left ear and elevations $-60°$ to $-120°$ for right ear. These results agree with the fact that most energies in the higher frequency from the sound source are shadowed by the head [Cheng and Wakefield, 2001].

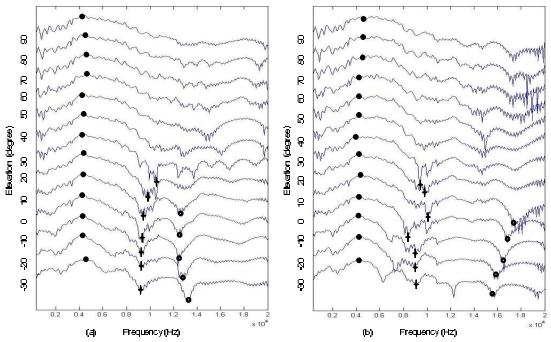

Figure 13: Frequency-Domain Comparision of HRTFs Measured at 13 Elevation Angles in Median Plane (Azimuth = 0°) (a) Left Ear Measured HRTFs (b) Right Ear Measured HRTFs

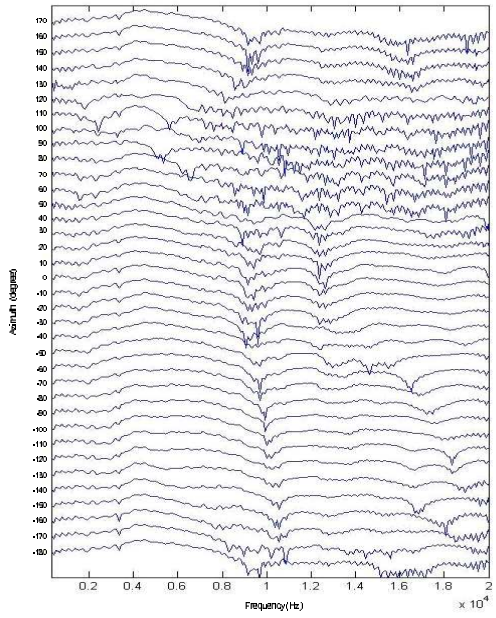

Figure 14: Frequency Domain Representation of Left Ear HRTFs

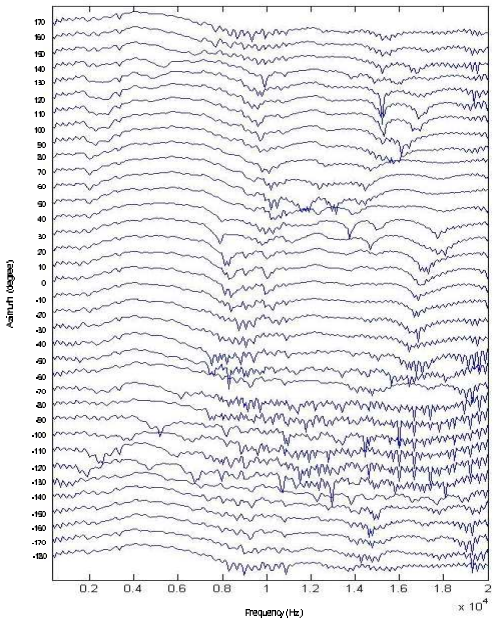

Figure 15: Frequency Domain Representation of Right Ear HRTFs

Time Domain Representation of HRTFs

The time domain representation of HRTFs is usually called head-related impulse response (HRIRs). They contain the same spatial cues as HRTFs, but present them in a different form.

Figure 16 and 17 show the time domain comparision of HRIRs measured at 36 azimuth angles, from -180° to 180° in 10° increments, in the horizontal plane. Diffraction effects can also be clearly noticed in these figures. Figure 16 plots 72 left ear HRIRs. The position of peak is located in negative azimuthal area, where the speaker is placed. The valley is of course in the positive azimuthal part, where most energy of stimuli is shadowed by head when speaker is turned to the right ear side. For the right ear HRIRs, as shown in Figure 17, the positions of peak and valley are placed in the opposite way to that of left ear.

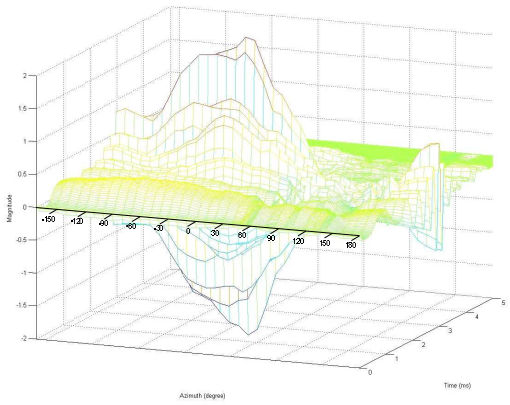

Figure 16: Time-Domain Representation of Left Ear HRIRs Measured at 36 Azimuth Angles in Horizontal Plane (Elevation = 0°)

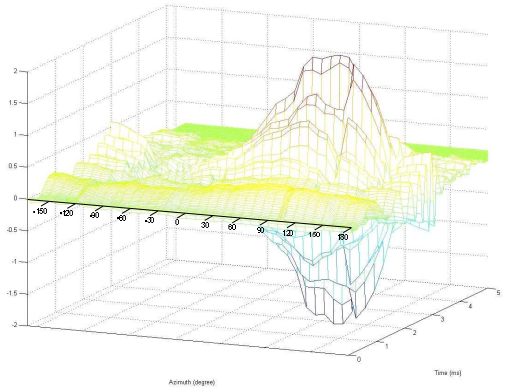

Figure 17: Time-Domain Representation of Right Ear HRIRs Measured at 36 Azimuth Angles in Horizontal Plane (Elevation = 0°)

Individual Differences of HRTFs

It is well known that HRTFs vary from person to person. We did not measure HRTFs on human subjects yet. But our trials on KEMAR manikin, which is adapted with large and small shape pinna respectively, also reveal this fact in some way. Figure 18 shows that although measured at the same direction of $\theta = 0°$ and $\phi = 0°$, HRTFs are different, even between left and right ear of same size.

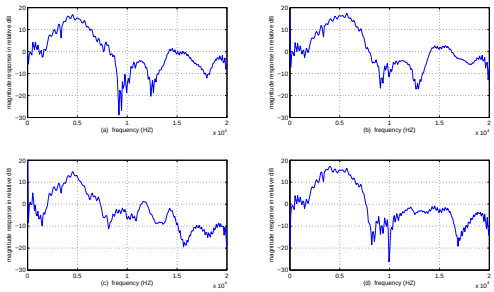

Figure 18: Differences of HRTFs (measured at direction of $\theta = 0°, \phi = 0°$)(a) Left Small Ear HRTF (b) Right Small Ear HRTF (c) Left Large Ear HRTF (c) Right Large Ear HRTF

CONCLUSION

In this paper, we introduced our HRTF measurement procedure. We focused on the design of the test signal and the post-processing procedures to extract the HRTFs. The measurements and the procedure can be executed in a semi-anechoic sound chamber, which has some reverberation. This makes the setup much more cost effective and the results fully proved feasibility.

Future research will focus on experiments on human subjects to get individual HRTF data.

REFERENCES

V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano. The cipic hrtf database. In *Proc.IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics*, pages 99 – 102, New Paltz, NY, USA, October 2001.

C. I. Cheng and G. H. Wakefield. Introduction to head-related transfer functions (hrtfs): Reprentations of hrtfs in time, frequency, and space. *J. Acoust. Soc. Am*, 49(4):231–249, April 2001.

W. G. Gardner and K. D. Martin. HRTF measurements of a KEMAR. *J. Acoust. Soc. Am*, 97(6):3907–3908, Jun. 1995.

E. Grassi, J. Tulsi, and S. Shamma. Measurement of head-related transfer functions based on the empirical transfer function estimate. In *Proc. International Conference on Auditory Display*, pages 119–122, Boston, MA, USA, July 2003.

J. L. Hunter. *Acoustics*. Englewood Cliffs, N.J., 1957. ISBN b1038197.

J. Jot, V. Larcher, and O. Warusfel. Digital signal processing issues in the context of binaural and transaural stereophony. *J. Audio Eng. Soc.*, page 3980, February 1995.

E. A. Lopez-Poveda and R. Meddis. A physical model of sound diffraction and reflections in the human concha. *J. Acoust. Soc. Am*, 100(5):3248–3259, November 1996.

E. Meyer and E. Neumann. *Physical and Applied Acoustics: An Introduction*. Academic Press, New York, 1972. ISBN 0124931502. Chapter 7. Physiological and Psychological Acoustics.

A. Muller and P. Massarani. Transfer-function measurement with sweeps. *J. Audio Eng. Soc.*, 49(6):443–471, June 2001.

M. Rossi. *Acoustics and Electroacoustics*. Artech House, Norwood, MA, 1998. ISBN 0890062552. Chapter 10. Humans and Sound.

B. A. Shenoi. *Introduction to digital signal processing and filter design*. Wiley-Interscience, Hoboken, New Jersey, 2006. ISBN 0471464821.

S. W. Smith. *The scientist and engineer's guide to digital signal processing*. California Technical Publishing, San Diego, CA, 1st ed edition, 1997. ISBN 0966017633. Chapter 22. Audio Processing.

R. Teranishi and E. A. G. Shaw. External-ear acoustic models with simple geometry. *J. Acoust. Soc. Am*, 44(1):257–263, 1968. Received in November 1967.

D. N. Zotkin, R. Duraiswami, E. Grassi, and N. A. Gumerov. Fast head-related transfer function measurement via reciprocity. *J. Acoust. Soc. Am*, 120(4):2202–2215, October 2006.