Decomposition and Modeling of Head-Related Impulse Responses for Customized Spatial Audio

KENNETH JOHN FALLER II[1], ARMANDO BARRETO[1], NAVARUN GUPTA[2] and NAPHTALI RISHE[3]

Electrical and Computer Engineering Department[1] and School of Computing and Information Science [3]
Florida International University
Miami, FL 33174
USA
kfall001@fiu.edu http://dsplab.eng.fiu.edu/

Department of Electrical and Computer Engineering[2]
University of Bridgeport
Bridgeport, CT 06604
USA

Abstract: - In order to achieve full binaural sound, audio signals must be synthesized using specialized digital filters characterized by their so-called Head-Related Impulse Responses (HRIRs). They represent the modification to sound due to the listener's torso, shoulders, head and pinnae, or outer ears. Conventional methods of measuring individual HRIRs require the use of expensive and cumbersome equipment to obtain the HRIRs of every prospective user of the spatial audio system. Alternatively, a manikin may be used to create generic HRIRs. Unfortunately, the access to the necessary equipment is very limited and generic HRIRs do not provide as high spatialization fidelity as individually measured HRIRs. Hence, the creation of a method to create customizable HRIRs without the use of the specialized equipment is extremely desirable. In previous research on this topic, Prony's modeling method was used to obtain an appropriate set of time delays and a resonant frequency to approximate measured HRIRs. In an effort to expand upon this research the Prony method was substituted for the Steiglitz-McBride iterative approximation method and a noticeable improvement in the approximation was achieved. This paper shows that the improvement achieved is statistically significant for most HRIRs and, therefore, may be advantageous for HRIR decomposition.

Key-Words: - Head-Related Impulse Responses (HRIR), Prony modeling method, Steiglitz-McBride iterative approximation method, customizable spatial audio.

1 Introduction

Three dimensional (3-D) spatial audio has become increasingly popular in scientific, commercial and entertainment systems [1]. Spatial audio can be used in a variety of applications, from assistance for the visually impaired to enhancement of computer video games. Currently, there are two methods to achieve spatial audio: multi-channel or two-channel. The multi-channel approach requires that speakers be physically positioned around the listener (e.g., Dolby® 5.1 array). This is an effective but expensive solution and is impractical some applications. The two-channel approach is more practical, especially considering that it can be implemented using digital signal processing (DSP) techniques and delivered to a listener through a pair of speakers or headphones. Special filters are required in order to achieve spatial audio in a two-channel system. The special filters are a pair of characteristic transfer functions that mediate between the sound at its point of origin and the left and right eardrums of the listener. The head of the listener has been considered the main source of modification of the original sound; hence the transfer functions have traditionally been referred to as Head-Related Transfer Functions or HRTFs.

HRTFs model the effect of anatomical (torso, head, external ear, etc.) and environmental (walls, floor, etc.) factors which cause modification of a sound as it propagates from its source to each of the listener's eardrums. Therefore, every position and each ear will

have a specific HRTF. Filtering a sound signal with the two HRTFs corresponding to a specific source position results in a binaural sound (left channel, right channel) that, when played to a listener through stereo headphones will cause a perception similar to that of a sound emanating from the source location in question (Fig. 1).

It is assumed that these HRTFs represent a linear and time-invariant (LTI) transformation. This assumption significantly increases the amount of analytical tools available from the realm of dynamic systems analysis that can be utilized for HRTF analysis. For example, since the transfer functions of LTI systems are complex functions, they have a magnitude and a phase associated to them. When the LTI system is considered as excited by combinations of steady-state sinusoidals, the magnitude of the transfer function reflects the "Magnitude Response" of the system, and its phase reflects the "Phase Response" of the system. These two responses indicate how the magnitude and phase of sinusoidal components at different frequencies are changed by the system as a signal is processed by it [2]. This is the core concept of the "Frequency Domain" analysis of LTI systems, and it is extremely useful in the study of HRTFs (Fig. 2).

On the other hand, LTI systems are represented also by their "Impulse Response." Moreover, according to the theory of LTI systems, the output of an LTI system to an arbitrary input can be determined by convolution of the input signal with the impulse response of the system. For similar reasons to the HRTFs, their associated impulse responses have been traditionally referred to as Head-Related Impulse Responses (HRIRs). The HRIR of a system can be determined when a brief signal, or impulse, is presented to the HRTF. In reality, however, it is impractical to drive an acoustic system with a very short pulse with very high amplitude to approximate an impulse. Instead, the practical measurement of HRIRs uses a combination of special signals, called Golay codes, which are used to estimate the HRIR indirectly [2].

Traditionally HRIRs are implemented as finite impulse response (FIR) filters. FIR filters are digital filters implemented through computation which does not contain feedback. The transfer functions of an FIR filter only contains zeros meaning that only the numerator of the transfer function has to be considered (1). The standard number of coefficients used for HRIRs is 512 or 256 coefficients but in some cases, as few as 128 have been used as representative of the spectral cues created by anatomical factors [2]. For our research, all the HRIRs used were reduced to 128 coefficients.

$$B(z) = b(1) + b(2)z^{-1} + \cdots + b(n+1)z^{-n} \quad (1)$$

Currently, there are two main types of HRIRs, according to their creation: generic or individual. Generic HRIRs are obtained from measurements made on a manikin head (e.g., M.I.T.'s measurements of a KEMAR Dummy-Head Microphone [4]) or using a limited number of subjects to represent the general population (e.g., the CIPIC Database [5]). This type of generic HRIRs does not have as high spatialization fidelity as individual HRIRs [2], which require that the actual prospective user of the spatialized audio system undergo time-consuming measurements with specialized and expensive equipment, such as the AuSim HeadZap system which is used at FIU to obtain measured HRTFs. Unfortunately, this severely limits the access to HRIRs obtained this way. As a result of this, a majority of the HRIRs in use are generic.

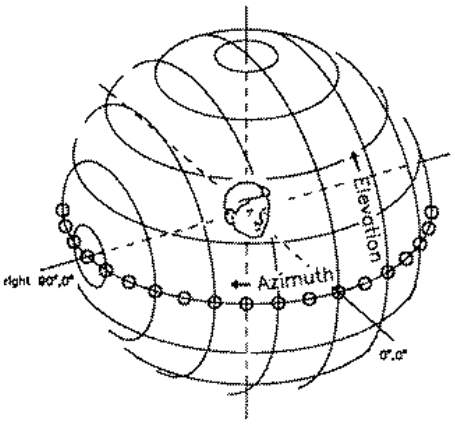

Fig. 1. Diagram of spherical coordinate system [3]

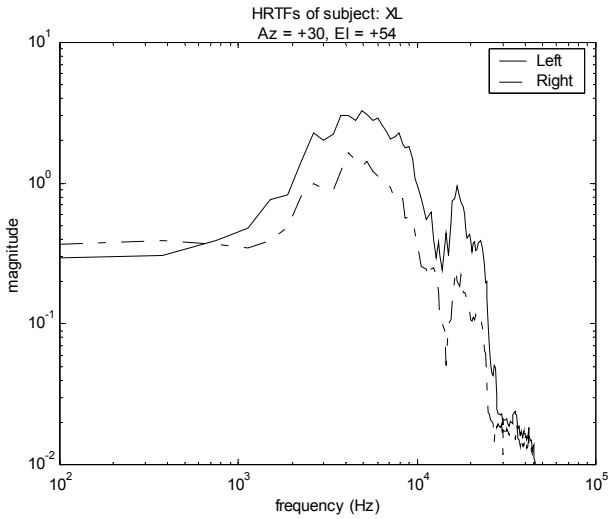

Fig. 2. HRTFs of left (solid) and right (dotted) channels for a subject. The sound source is at azimuth +30º and elevation 54º. The data was recorded at FIU using the HeadZap system [2]

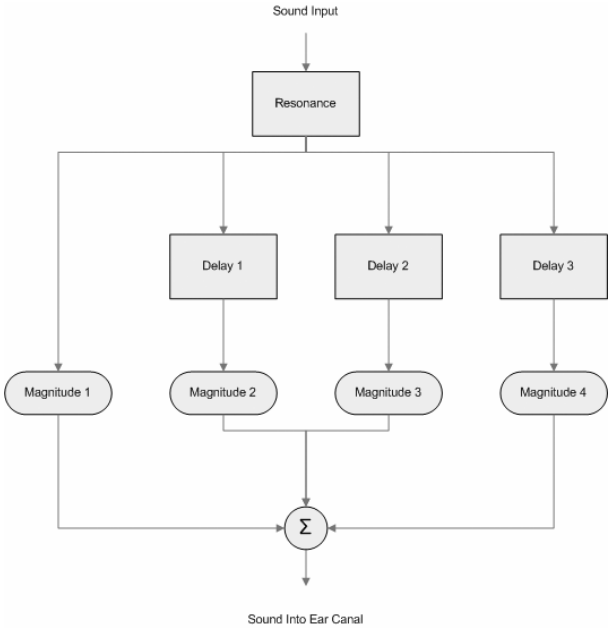

Fig. 3. Block diagram of pinna model [2]

Our group is pursuing the definition of an HRIR model requiring a reduced number of parameters that could be instantiated with values derived from simple anatomical measurements from the prospective listener. This would generate customized HRIRs, as an alternative to generic or individual HRIRs. Customized HRIRs were obtained with a high percentage of fit in [2]. Our pinna model (Fig. 3) consists of a resonance block that feeds into four different magnitude and delay pairs. The outputs of these parallel paths are then re-combined to result in a customized HRIR [2][6]. Our initial goal was to verify that experimentally measured HRIRs could be decomposed to obtain parameters for our model that would result in a reasonably similar "reconstructed" HRIR. The required HRIR decomposition was originally accomplished by successive application of the Prony method of signal approximation [2]. The algorithm for Prony implemented was the method described in reference [7]. Our algorithm for successive deconstruction of an experimentally measured HRIR, such as the one shown in Figure 4, is detailed in [6]. The objective of the process is to obtain scaled and delayed damped sinusoidal components, such as those shown in Figure 5, which specify the associated parameters to instantiate the resonance, magnitudes and delays needed to approximate the original HRIR using our model. The adequacy of the "reconstructed" HRIR that results was evaluated by a measure of "fit" between the original HRIR and the one reconstructed using the model.

Recently, these experiments were repeated but the Prony method was substituted by the Steiglitz-McBride iteration method ("STMCB"). The STMCB method is similar to Prony in that it also tries to find an infinite impulse response (IIR) model for a signal. The STMCB method attempts to minimize the squared error between the impulse response and the signal it approximates in an iterative fashion. A noticeable improvement was observed after the substitution of Prony with STMCB for HRIR modeling. The algorithm implemented for the STMCB method is as described in reference [8].

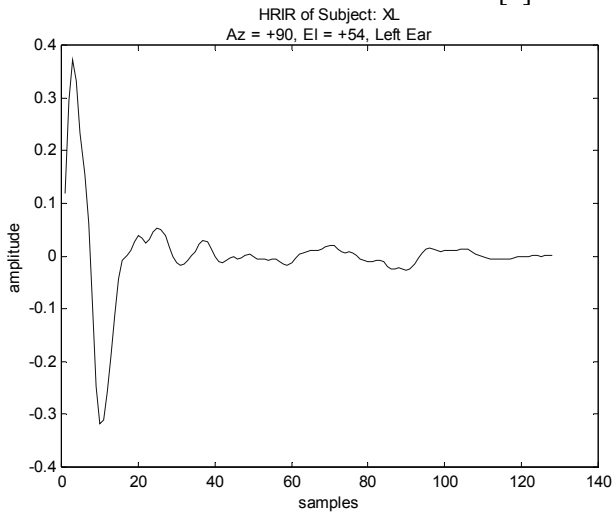

Fig. 4. The HRIR to be deconstructed [2]

2 Methodology

The apparent advantage of using the STMCB method prompted a systematic evaluation of this

performance difference. The following subsections describe the methodology used to compare STMCB and Prony for HRIR modeling.

2.1 Iterative HRIR fitting

In this experiment, statistics will be utilized to rank the performances of the two modeling methods when used for HRIR analysis. The use of a database that is a good representation of the prospective user population is needed to obtain statistically valid results. A subset of the CIPIC HRIR database, which contains 45 subjects recorded at 44.1 kHz with various anatomical properties for multiple azimuths and elevations, was chosen for the comparison [1]. Our analysis focused on HRIRs for the right ear, for sources at ear level (0º elevation), and involved 25 azimuth values, which range from -80º to 80º.

In order to accelerate the process of comparing the two methods, a Matlab® script was created to iterate through the CIPIC database. Each HRIR signal was decomposed into three separate signals: the primary resonance and two resonances with delays which will be referred to as echoes.

This is similar to the research performed in [2] with the exception that the last (third) echo is not considered here. For example, Fig. 5 shows a decomposition of a signal using Prony in [2]. In the experiment reported in this paper everything would be identical, except that signal F4 would not be extracted. It is apparent that the first three signals contain most of the power present in the HRIR. Hence, in an effort to increase the speed of the comparison algorithm only three signals will be extracted. The signals are obtained by passing a small "window" or segment of the original signal to one of the approximation methods (Prony or STMCB) which will return a 2^{nd} order IIR representation that best approximates the window passed. In an additional effort to reduce computation time, the window sizes are restricted to certain ranges determined by previous work in this area [2]: the window used to determine the first echo must start at least 5 samples into the HRIR segment analyzed, which results in window1 in Fig. 6 starting at 5. Additionally, the windows are not allowed to grow wider than 10 samples. Once the three signals are extracted, an HRIR can be created from these components by adding them together at the appropriate delays, and can be compared to the original HRIR in terms of mean square (MS) value:

$$\text{Error} = \text{Original HRIR} - \text{Reconstructed HRIR}, \quad (2)$$

$$\text{Fit} = [1 - \{\text{MS(Error)}/\text{MS(Original HRIR)}\}]. \quad (3)$$

The percentage fit ("fit") between the original HRIR and the reconstructed HRIR was calculated for every subject and every azimuth, and used as the figure of merit to compare the performance of STMCB and Prony for this modeling task.

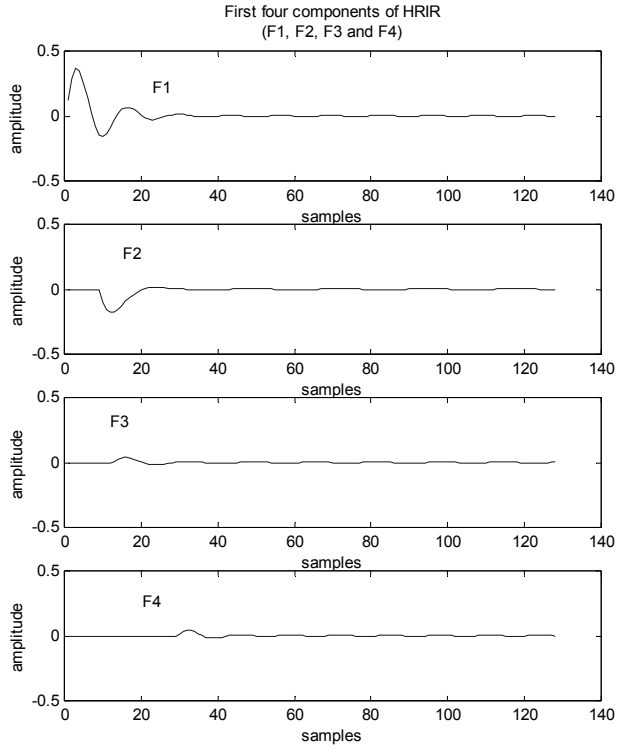

Fig. 5. The four components of the HRIR shown in Figure 4, with their respective delays [2]

2.2 Statistical Significance of the Improvement

Matlab® was utilized again to perform statistical analysis of the results obtained from the algorithm described in the previous section. The fit obtained when STMCB was used was subtracted from the fit obtained when Prony was used, for each azimuth of every subject. The 45 differences for one azimuth constituted a single sample and there were 25 samples (i.e., 25 azimuths) in total.

The following hypotheses were tested to determine if STMCB significantly improved the fit percentage:

$$H_0: \mu = 0. \quad (4)$$

$$H_a: \mu > 0. \quad (5)$$

Here μ is the mean improvement that would be achieved by using STMCB instead of Prony in the modeling process. The null hypothesis says that no

improvement occurs, and H_a says that the fit with STMCB is higher, on average.
In this case, the one-sample *t* statistic is:

$$t = \frac{\bar{x} - 0}{s / \sqrt{n}}, \tag{6}$$

were $\bar{x}$ is the sample mean, *s* is the standard deviation and *n* is the sample size.

The results of the significance test will determine if STMCB outperformed the Prony method for HRIR analysis. Unfortunately, the size of the improvement cannot be determined from these results. A statistically significant but small improvement would not be sufficient to claim that STMCB is a practically superior method. A confidence interval is used to remedy this problem. The confidence interval will display how much STMCB improved over Prony with a margin of error:

$$\bar{x} \pm t * \frac{s}{\sqrt{n}}. \tag{7}$$

The procedure followed in this study is tailored after examples presented in [9].

3 Results and Discussion

This section summarizes and discusses the results obtained. A basic mean of the fits and the gains (fit with STMCB - fit with Prony) achieved for each azimuth was the first calculation obtained. An increase in fit and gain was achieved for every azimuth when the STMCB method was used. Unfortunately, this is not enough to establish a statistically significant difference between the performances of the methods. Hence, a t-test was utilized to compare the methods for each azimuth.

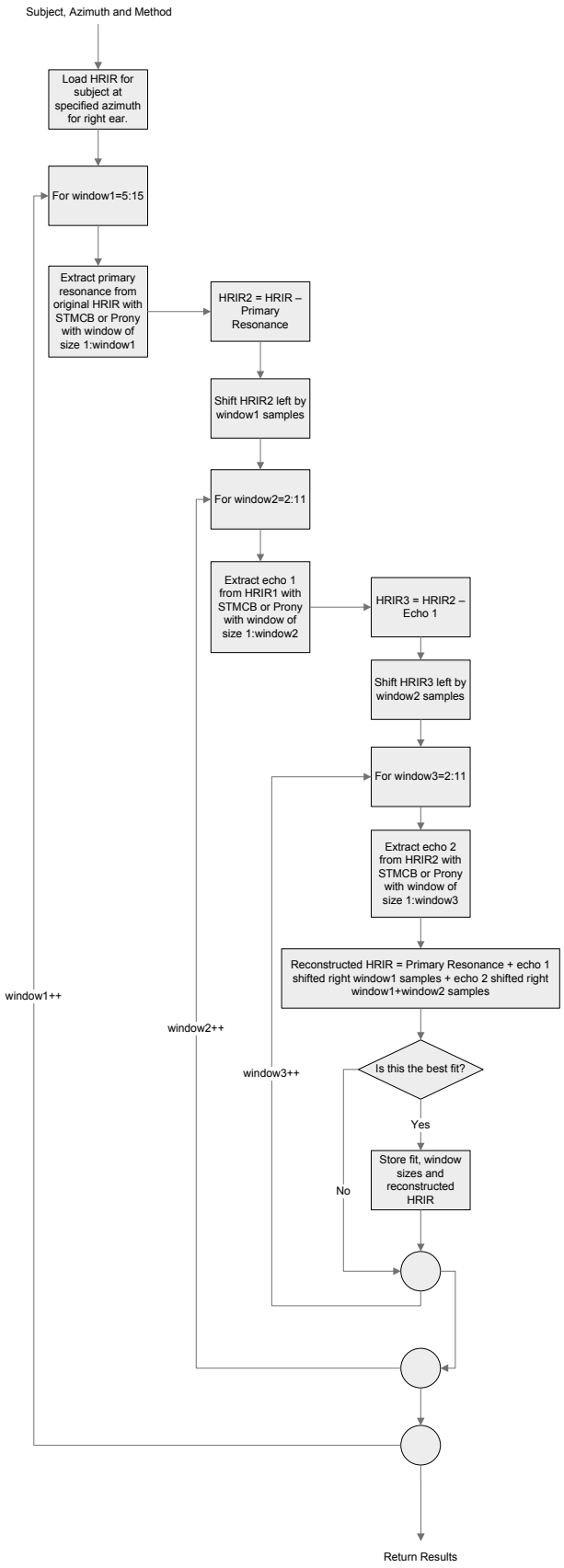

Fig. 6. Flow chart for the iterative process that determines best fit.

The t-test algorithm used was a predefined script in Matlab® called "ttest." This command performs the statistical analysis algorithm described in section 2.2 of this paper. The command provides the values of the t-statistic, as well as the associated p-value, i.e., the probability that the value of the t-statistic is equal to or more extreme than the observed value by chance, under the null hypothesis (mean difference = 0). Additionally, the command provides both limits of a 95% confidence interval on the mean [10]. If the null hypothesis cannot be rejected, then the difference is not significant ($p > 0.05$). Otherwise, the null hypothesis is rejected, with $p < 0.05$, i.e., for these azimuths the use of STMCB resulted in a significant improvement over the use of Prony. With the exception of three azimuths (10º, 15º and 20º) the null hypothesis could be rejected. This means that, for a majority of the angles studied, a significant increase in fit was observed when STMCB was used instead of Prony.

Beyond the global analysis of the modeling results to conclude that the improvement achieved by using STMCB instead of Prony was statistically significant, some specific cases were analyzed in further detail to help in the interpretation of their results. For example, HRIRs that displayed extremely high, low and average fits were examined manually, as displayed in figures 7 and 8 respectively. In these figures, the top plot displays the individual HRIR, the middle plot displays the HRIR reconstructed using STMCB and the bottom displays the HRIR reconstructed using Prony. Figure 7 (subject 24, 35° azimuth) displays reconstructed HRIRs that had a high percentage fit of about 94%. This is consistent because the main morphology is maintained when the HRIR is reconstructed using either Prony or STMCB. These results, in turn, confirm that the limitation to the modeling of just two "echoes" was not too restrictive. Conversely, figure 8 (subject 27, 20° azimuth) shows a plot with relatively low percentage fit of about 28%. As can be seen in this figure, the reconstructed HRIRs do not resemble the original HRIR. It would seem that both methods were able to approximate the second positive "peak" in the HRIR, appearing at a latency of about 12 sampling intervals. On the other hand, it is apparent that both STMCB and Prony minimized the error in the approximation of the first positive peak and the negative peak that immediately follows it by substituting both with a data segment that hovers around zero, which is clearly inappropriate. It is possible that the separation of these two echoes in HRIRs such as these might be very small, particularly considering the limited temporal resolution afforded by the 44.1 kHz sampling rate employed in the development of the CIPIC Database, as compared to the 96 kHz sampling rate used in other previous studies that have attempted this kind of HRIR decomposition [2][6]. However, further research is needed to ultimately pinpoint the reasons for the degradation of this technique for some azimuth values.

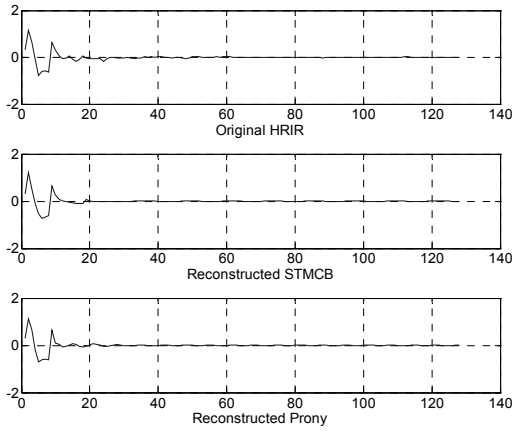

Fig. 7. Plot of the original and reconstructed HRIRs for subject 24 at 35º azimuth.

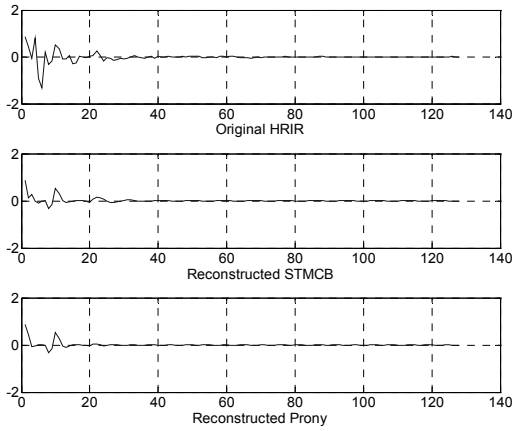

Fig. 8. Plot of the original and reconstructed HRIRs for subject 27 at 20º azimuth.

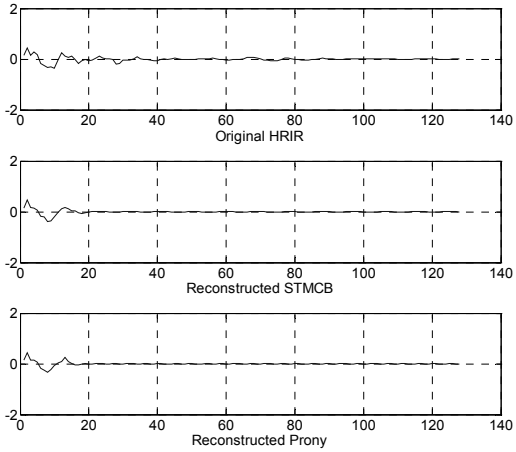

Fig. 9. Plot of the original and reconstructed HRIRs for subject 18 at -5º azimuth.

Figure 9 displays an example in which the average level of fit was achieved (~74% fit). As seen in the plots, the initial, stronger parts of the HRIR are represented extremely well by the reconstruction methods. The latter part of the original HRIR, which has less significant fluctuations, was zeroed out by both methods.

4 Conclusion

There is an interest in the development of methods of audio spatialization that can be easily "customized" to provide each individual listener with maximal imaging fidelity. Our group has pursued this goal through the establishment of a dynamical model capable of generating head-related impulse responses (HRIRs) that can be used for audio spatialization. An emphasis has been placed in reducing the number of parameters required to instantiate this model, and on the fact that these parameters should be, in the long run, easy to define from simple anatomic measurements obtained from each prospective user of the spatial sound system.

With this aim, one of our initial goals was to show that specific sets of parameter values (e.g., resonance characteristics, echo delays and magnitudes) can be found that will result in modeled HRIRs that are a good approximation of measured individual HRIRs. Since we have proposed a method to find the necessary model parameters from a measured individual HRIR, by deconstructing it into multiple second order responses (that appear resized and delayed), it is of interest to identify the modeling approach that achieves the highest efficiency in this process. In this study, the Steiglitz-McBride (STMCB) signal approximation model has been compared to the Prony approximation method, in its use as the core of the iterative process used to deconstruct HRIR sequences. This systematic study showed that, for the HRIRS from the CIPIC database with 0º elevation and azimuths ranging from -80º to 80º, the STMCB method outperformed the Prony method for the vast majority of azimuth values, as judged by the enhanced "fit" (in the least squares sense) between the HRIR created by reconstruction from the three 2nd order partial responses found, and the original, experimentally measured HRIR. There were only three azimuth values (10º, 15º and 20º), out of 25, for which the improved performance of the STMCB approximation did not prove to be statistically significant. It is important to note, however, that for all azimuth values (including 10º, 15º and 20º) the mean value of fit between the reconstructed HRIR and the original HRIR was higher when STMCB was used than when Prony was utilized. Accordingly, it seems reasonable to recommend the use of STMCB signal approximation methods for HRIR modeling, in general.

5 Acknowledgement

This work was partially sponsored by NSF grants IIS-0308155, CNS-0520811, HRD-0317692 and CNS-0426125.

References:

[1] "Spatial Sound." CIPIC Interface Laboratory. University of California, Davis. 23 Aug. 2005 <http://interface.cipic.ucdavis.edu>.

[2] N. Gupta, "Structure–Based Modeling of Head-Related Transfer Functions Towards Interactive Customization of Binaural Sounds Systems." Ph.D. Dissertation, Florida International Univ., 2003.

[3] J. C. Makous, J. C. Middlebrooks, and D. M. Green. "Directional Sensitivity of Sound-Pressure Levels in the Human Ear Canal." Journal of the Acoustical Society of America 86 (1989): 89-108.

[4] W. Gardner, and K. Martin. HRTF Measurements of a KEMAR Dummy-Head Microphone. 18 May 1994. Massachusetts Institute of Technology. 24 Aug. 2005 <http://sound.media.mit.edu/KEMAR.html>.

[5] V. R. Algazi, R. O. Duda, D. M. Thompson and C. Avendano. Workshop on Applications of Signal Processing to Audio and Acoustics, 21-24 Oct. 2001, Audio & Electroacoustics committee of the IEEE Signal Processing Society. New Paltz, NY: IEEE, 2001.

[6] A. Barreto, and N. Gupta, “Dynamic Modeling of the Pinna for Audio Spatialization”, WSEAS Transactions on Acoustics and Music, 1 (1), January 2004, pp. 77 - 82.

[7] T.W. Parks, and C.S. Burrus, Digital Filter Design, John Wiley & Sons, 1987, pp. 226-228.

[8] K. Steiglitz, and L.E. McBride, "A Technique for the Identification of Linear Systems," IEEE Trans. Automatic Control, Vol. AC-10 (1965), pp. 461-464.

[9] D. S. Moore, and G. P. McCabe. Introduction to the Practice of Statistics. 4th ed. New York: W. H. Freeman and Company, 2003. 501-504.

[10] "Ttest - Statistics Toolbox." Matlab Version 7 Release 14. Mathworks Inc. 1 Sept. 2005 <http://www.mathworks.com/>.