Performance Comparison of Two Identification Methods for Analysis of Head Related Impulse Responses

Kenneth John Faller II[1], Armando Barreto[1], Navarun Gupta[2] and Naphtali Rishe[3]

Electrical and Computer Engineering Department[1] and School of Computer Science[3]
Florida International University
Miami, FL 33174 USA

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

***Abstract*-Head-Related Impulse Responses (HRIRs) are used in signal processing to model the synthesis of spatialized audio which is used in a wide variety of applications, from computer games to aids for the vision impaired. They represent the modification to sound due to the listener's torso, shoulders, head and pinnae, or outer ears. As such, HRIRs are somewhat different for each listener and require expensive specialized equipment for their measurement. Therefore, the development of a method to obtain customized HRIRs without 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. During several recent experimental attempts to improve on this previous method, a noticeable increase in percent fit was obtained using the Steiglitz-McBride iterative approximation method. In this paper we report on the comparison between these two methods and the statistically significant advantage found in using the Steiglitz-McBride method for the modeling of most HRIRs.**

I. Introduction

Humans have the remarkable ability to determine the location and distance of a sound source. How we are able to do this has been a topic of research for some time now. Some aspects of this topic are well understood while other aspects still elude researchers. For example, it is known that the time difference between the arrival of a sound to each ear provides a strong cue for the localization of the sound source in azimuth, while elevation is primarily determined by the perceived modification of sound that takes place in the pinnae or outer ear [1]. Many modern technologies benefit from generating synthetic sounds that have a simulated source location. Currently there are two approaches to synthetic spatial audio: multi-channel and two-channel approaches. The multi-channel approach consists of physically positioning speakers around the listener (e.g., Dolby 5.1 array). This is an effective solution but impractical for the majority of applications that utilize spatial audio. The two-channel approach is more practical because it can be implemented using digital signal processing (DSP) techniques and delivered to the user through headphones.

One such technique is the use of Head-Related Impulse Responses (HRIRs). HRIRs capture the location-dependent spectral changes that occur due to environmental (walls, chairs, etc.) and anatomical (torso, head, and outer ears or pinnae) factors [1]. This approach requires the availability of an HRIR for each ear and each position (elevation, azimuth) of the sound source. The sound signal is then convolved with the HRIR for each ear, to create a binaural sound (left channel, right channel), which gives the listener the sensation that the sound source is located at a specific point in space (Fig. 1). This ability to emulate spatial audio with only two channels has broadened its uses in several important areas: human/computer interfaces for workstations and wearable computers, sound output for computer games, aids for the vision impaired, virtual reality systems, "eyes-free" displays for pilots and air-traffic controllers, spatial audio for teleconferencing and shared electronic workspaces, and auditory displays of scientific or business data [1].

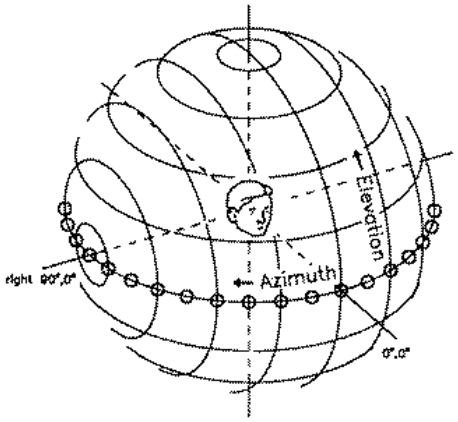

Fig. 1. Diagram of spherical coordinate system [2]

At present, the HRIRs that are used for the synthesis of spatialized audio are either generic or individual. Generic HRIRs are measured using a manikin head (e.g., M.I.T.'s measurements of a KEMAR Dummy-Head Microphone [3]) or using a limited number of subjects to represent the general population (e.g., the CIPIC Database [4]). Individual HRIRs

require the subject to undergo time consuming measurements with specialized equipment. Furthermore, a trained and experienced technician is necessary to operate the equipment. Unfortunately, access to the equipment necessary to measure HRIRs is limited for the general public. As a consequence, many spatialized audio systems rely on generic HRIRs, although these are known to reduce the fidelity of the spatialization and increase phenomena such as front to back reversals [5]. These reversals occur when a sound simulated in the front hemisphere is actually perceived in a symmetrical position of the back hemisphere, or vice versa.

Previous research by our group has sought to create a model to generate customized HRIRs with only a few simple measurements. The basic model that resulted from previous research comprises a single resonance feeding its output to a set of parallel paths, each with a magnification and a delay factor, which could be obtained from measurements of the head and pinnae and the use of Prony's method (Fig. 2) [5][6]. Prony's method is an algorithm for finding the coefficients for an IIR filter with a prescribed time domain impulse response. The algorithm implemented is the method described in reference [7].

During recent experimentation on this topic, Prony's method ("Prony") was substituted by the Steiglitz-McBride iteration method ("STMCB"). The STMCB method is similar to Prony in that it also tries to find an IIR filter with a prescribed time domain impulse response. The only difference is that the STMCB method attempts to minimize the squared error between the impulse response and the input signal. A noticeable improvement was observed after the substitution of Prony with STMCB for HRIR modeling. The algorithm for the STMCB method implemented is the method described in reference [8].

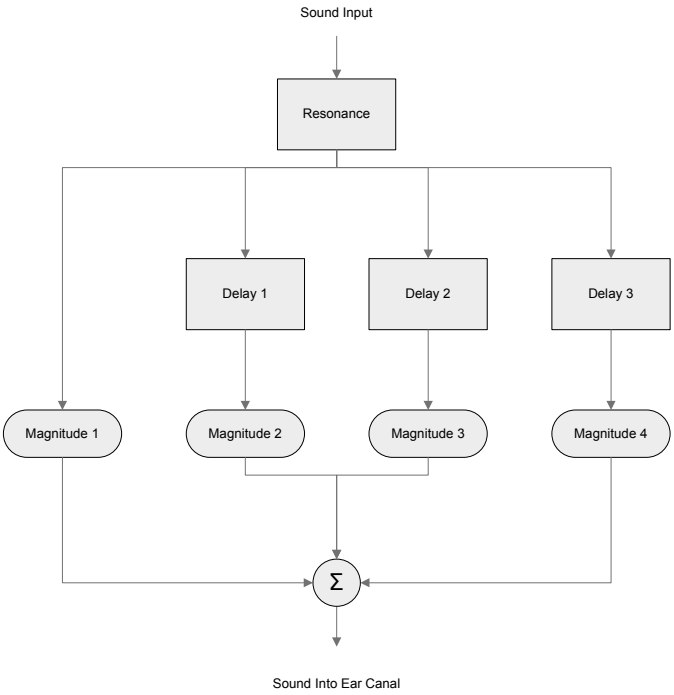

Fig. 2. Block diagram of pinna model

II. Methodology

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

A. Best Fit Iteration Algorithm

The purpose of this experiment is to show that there is a statistically significant modeling improvement when STMCB is used for HRIR analysis instead of Prony. In order to do this, a sample population of HRIRs is necessary. Fortunately the CIPIC database, which is a database that contains HRIRs recorded at 44.1 kHz. from 45 subjects for various azimuths and elevations, is available from [1]. This database contains a large number of HRIRs and is impractical to analyze all azimuths and elevations for both ears. Hence, only HRIRs for the right ear at 0º elevation and 25 different azimuths ranging from -80º to 80º were involved in this comparison.

A Matlab® script was created to iterate through each of the CIPIC HRIRs described above. The script attempts to discover the best fit between a measured HRIR and the HRIR that can be reconstructed by adding the partial 2nd order responses (equivalent to a full path from top to bottom in Fig. 2) extracted from the HRIR using both Prony and STMCB. Both of these methods can estimate a full signal with a smaller segment of the original signal. Furthermore, considering that the original HRIR is believed to consist of a primary resonance and at least two delayed echoes [5], processing the entire HRIR with Prony or STMCB at once would result in a large approximation error sequence, as defined in equation (1). Therefore, data "windows" of increasing sizes have to be tried iteratively, to define each of the 2nd order "echoes" that make up the HRIR, as indicated in Fig. 3. The sizes of the windows to use are determined by iteration, subject to the constraints found in previous work in this area [5]: The first window is at least 5 samples which results in window1 in Fig. 3 starting at 5. Additionally, the windows are not allowed to grow wider than 10 samples.

In this comparison study, the reconstructed HRIRs will only consist of three 2nd order responses that are obtained from Prony or STMCB. These are the "primary" response and two delayed responses, referred to as "echoes." While there may be other late components in the HRIRs, such as the third echo recovered in [5], it is clear that these first three components contain most of the power in the HRIR and were selected as the basis of comparison to keep the number of iterations manageable. Once the primary response and echoes are determined, the reconstructed HRIR is created by adding the extracted responses at the determined delays and comparing the resulting sequence to the original HRIR, in terms of mean square (MS) value:

$$\text{Error} = \text{Original HRIR} - \text{Reconstructed HRIR}, \quad (1)$$

$$\text{Fit} = [1 - \{\text{MS(Error)}/\text{MS(Original HRIR)}\}]. \quad (2)$$

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.

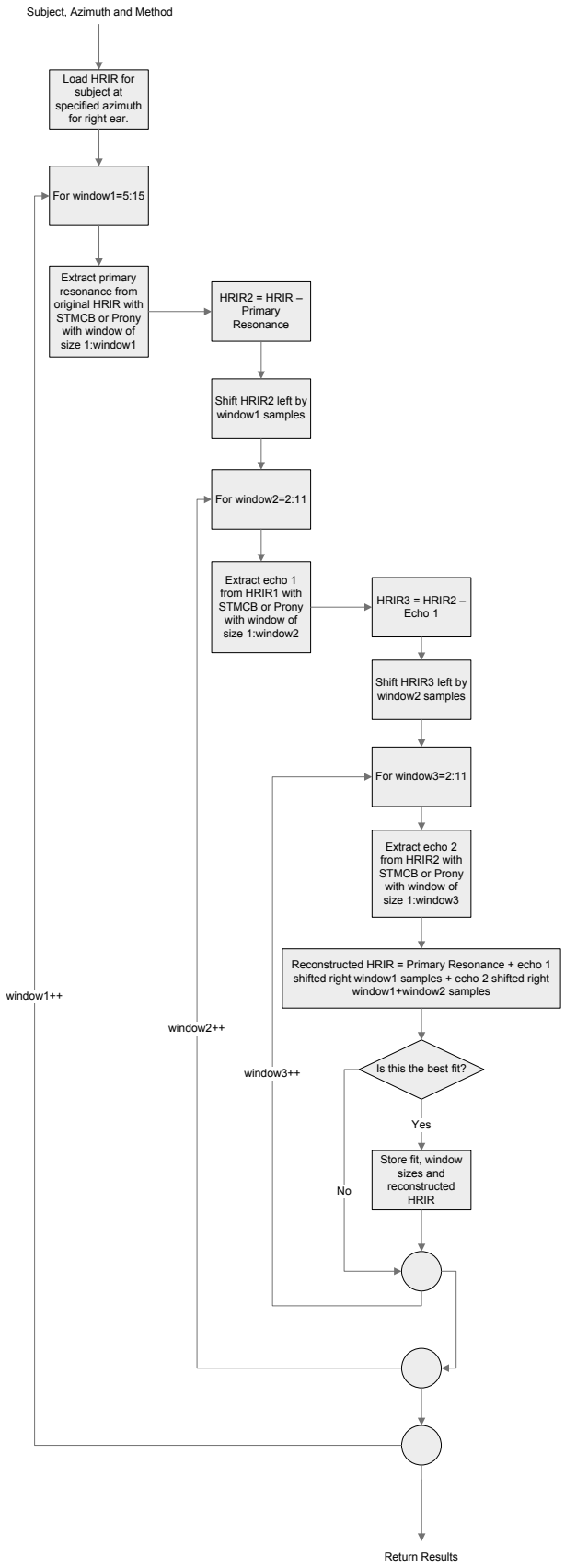

Fig. 3. Flow chart for the iterative process that determines best fit.

B. Statistical Analysis Algorithm

Additional Matlab® scripts were created to statistically analyze the results of the previous section. Matched-*t* tests were utilized in order to determine statistical significance of performance differences observed when the modeling task used Prony or STMCB, for each given source azimuth. The fit obtained through STMCB was subtracted from the fit obtained through Prony, for each azimuth. The 45 differences for one azimuth form a single sample and there were 25 samples (i.e., 25 azimuths) in total.

To assess whether the STMCB significantly improved the fit percentage, the following hypotheses were tested:

$$H_0: \mu = 0. \tag{3}$$

$$H_a: \mu > 0. \tag{4}$$

Here μ is the mean improvement that would be achieved by using STMCB over Prony in the modeling process. The null hypothesis says that no improvement occurs, and H_a says that the fit from STMCB is higher on average.

In this case, the one-sample *t* statistic is:

$$t = \frac{\bar{x} - 0}{s / \sqrt{n}}, \tag{5}$$

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 very small improvement would not be sufficient to claim that STMCB is a 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:

$$\bar{x} \pm t * \frac{s}{\sqrt{n}}. \tag{6}$$

The procedure followed and a complete example implementation is available in [9].

III. Results And Disscusion

The following section will overview and discuss the results obtained. Table 1 displays the mean fits for both STMCB and Prony. The "Gain" column is calculated by subtracting the Prony column from the STMCB column. For example, at azimuth -80º the fit improved from 81.20% (with Prony) to 87.57% (with STMCB), which results in a 6.36% gain.

TABLE 1
MEAN FIT OF PRONY AND STMCB

Azimuth (º)	Prony	STMCB	Gain
-80	81.20%	87.57%	6.36%
-65	75.80%	80.86%	5.05%
-55	70.83%	77.97%	7.14%
-45	69.42%	76.04%	6.61%
-40	68.17%	75.05%	6.88%
-35	70.15%	76.61%	6.45%
-30	68.09%	73.50%	5.41%
-25	68.48%	73.53%	5.05%
-20	69.35%	73.82%	4.46%
-15	66.90%	71.48%	4.58%
-10	65.72%	70.49%	4.77%
-5	61.78%	68.48%	6.70%
0	61.20%	66.52%	5.33%
5	59.98%	65.87%	5.89%
10	58.79%	63.22%	4.43%
15	60.21%	63.49%	3.28%
20	60.07%	62.08%	2.01%
25	60.18%	66.71%	6.53%
30	63.31%	66.96%	3.65%
35	63.04%	72.46%	9.42%
40	68.84%	75.00%	6.15%
45	67.71%	75.92%	8.21%
55	74.76%	82.13%	7.38%
65	77.05%	85.49%	8.44%
80	82.73%	88.66%	5.93%

TABLE 2
RESULTS OF MATCHED *t* PAIR PROCEDURE

Azimuth (º)	Null Hypothesis	p	t
-80	1	9.356E-11	8.445E+00
-65	1	9.716E-03	2.704E+00
-55	1	1.092E-03	3.496E+00
-45	1	3.337E-06	5.319E+00
-40	1	4.020E-09	7.311E+00
-35	1	2.176E-11	8.895E+00
-30	1	7.245E-06	5.086E+00
-25	1	3.127E-06	5.339E+00
-20	1	2.527E-04	3.982E+00
-15	1	2.440E-04	3.993E+00
-10	1	3.970E-05	4.567E+00
-5	1	5.826E-06	5.152E+00
0	1	7.957E-04	3.603E+00
5	1	2.299E-04	4.013E+00
10	0	5.191E-02	1.998E+00
15	0	1.717E-01	1.390E+00
20	0	3.411E-01	9.624E-01
25	1	7.808E-04	3.610E+00
30	1	2.388E-02	2.340E+00
35	1	2.726E-07	6.063E+00
40	1	7.363E-05	4.375E+00
45	1	4.562E-08	6.591E+00
55	1	4.130E-10	7.993E+00
65	1	8.665E-09	7.082E+00
80	1	9.702E-06	4.998E+00

* Degrees of freedom (df) is 44

To investigate the statistical significance of this apparent improvement achieved by using STMCB, the fit values associated with the HRIRs from each of the azimuth values studied were processed with the "ttest" command in Matlab®. This command performs a t-test of the hypothesis that the data submitted to it (in this case, the fit differences between STMCB and Prony) comes from a distribution with a pre-specified mean (in this case 0). 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 (CI1 and CI2) of a 95% confidence interval on the mean [10]. Table 2 summarizes the p-value and t-statistic results, for each population of fit differences, by azimuth. The second column of this table ("Null Hypothesis") displays a flag that summarizes the result of the test, in terms of significance. If the flag is "0", it means that the null hypothesis cannot be rejected in those cases, since the difference is not significant ($p > 0.05$). If the flag is "1", it means that 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.

As seen in Table 2, the improvement in percent fit with the use of STMCB is significant for many of the azimuths studied. In fact there were only 3 azimuths in which that was not the case: 10º, 15º and 20º. For these azimuths the null hypothesis cannot be rejected, which says that no statistically significant improvement in performance has occurred. However, the vast majority of the results support the view that the use of the Steiglitz-McBride approximation methods within the iterative process outlined in Figure 3 results in improved performance, as opposed to the use of the traditional Prony method [10].

From a different point of view, a statistically significant but very small improvement could be insufficient to prefer the use of an iterative method, such as STMCB, over a single-pass method, such as the traditional Prony algorithm. To illuminate this point, Table 3 displays the improvement of fit observed for each studied azimuth in terms not only of the mean improvement, but also indicating its standard deviation, and, most importantly a 95% confidence interval ([CI1, CI2]) for this improvement.

TABLE 3
CONFIDENCE INTERVAL AND STANDARD DEVIATION OF RESULTS

Azimuth (º)	CI 2	CI 1	Mean	SD
-80	4.845%	7.882%	6.363%	5.055E-02
-65	1.287%	8.822%	5.054%	1.254E-01
-55	3.025%	11.259%	7.142%	1.371E-01
-45	4.107%	9.117%	6.612%	8.338E-02
-40	4.984%	8.778%	6.881%	6.314E-02
-35	4.990%	7.914%	6.452%	4.866E-02
-30	3.265%	7.550%	5.408%	7.132E-02
-25	3.142%	6.952%	5.047%	6.341E-02
-20	2.204%	6.721%	4.463%	7.518E-02
-15	2.268%	6.889%	4.578%	7.691E-02
-10	2.662%	6.868%	4.765%	6.999E-02
-5	4.080%	9.323%	6.702%	8.726E-02
0	2.347%	8.303%	5.325%	9.914E-02
5	2.929%	8.841%	5.885%	9.839E-02
10	-0.038%	8.892%	4.427%	1.486E-01
15	-1.476%	8.028%	3.276%	1.582E-01
20	-2.201%	6.225%	2.012%	1.402E-01
25	2.885%	10.180%	6.533%	1.214E-01
30	0.507%	6.797%	3.652%	1.047E-01
35	6.289%	12.552%	9.421%	1.042E-01
40	3.318%	8.988%	6.153%	9.435E-02
45	5.699%	10.720%	8.210%	8.356E-02
55	5.516%	9.236%	7.376%	6.191E-02
65	6.038%	10.842%	8.440%	7.994E-02
80	3.538%	8.320%	5.929%	7.957E-02

In order to verify the validity of the percentages of fit found by the automated script employed for the comparison, a few individual modeling results were inspected. Two of these individual results are used for illustration. Figure 4 shows one original (measured) HRIR sequence (subject 24, 35º azimuth) in the top panel, as well as the reconstructed HRIRs obtained through STMCB (middle panel) and Prony (bottom panel). This figure confirms that the main morphology of the measured HRIR sequence has been preserved when the three 2nd order responses found by either STMCB or Prony were assembled together. This is in agreement with the high numerical values found by our comparison script in this case (approximately 94% for both STMCB and Prony). These results, in turn, confirm that the limitation to the modeling of just two "echoes" was not too restrictive.

In contrast, Figure 5 displays the results of approximating a different measured HRIR (subject 27, 20º azimuth). The original and reconstructed HRIR sequences appear in the same order as for Figure 4: original at the top, STMCB reconstruction in the middle, and Prony reconstruction at the bottom.

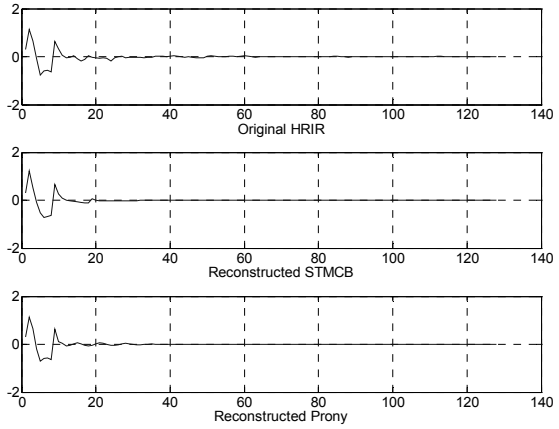

Fig. 4. Plot of the original and reconstructed HRIRs for subject 24 at 35º azimuth.

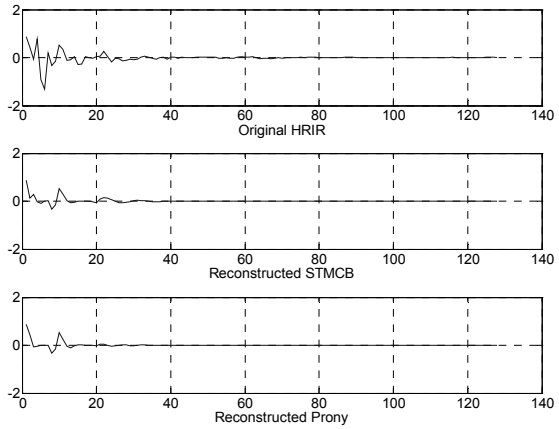

Fig. 5. Plot of the original and reconstructed HRIRs for subject 27 at 20º azimuth.

The fit for this particular case was about 28%, for both methods. As seen in the figure, the reconstructed HRIRs do not resemble the original. 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 this 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 [5][6]. However, further research is needed to ultimately pinpoint the reasons for the degradation of this technique for some azimuth values.

IV. Conclusion

We have implemented a semi-automated comparison of the modeling of measured HRIRs as triads of 2nd order responses. The extraction of these responses was achieved by the Stieglitz-McBride and Prony sequence approximation methods. The fit of reconstructed HRIRs obtained by re-assembling the 2nd order responses extracted to the original measured HRIRs was used as the figure of merit to compare the advantage of using one approximation method over the other. According to the analysis of our results, it has been shown that there is a statistically significant increase in percent fit when STMCB is used rather then Prony for the modeling of most of the HRIRs studied. On the other hand, while the STMCB decomposition of HRIRs at 10º, 15º and 20º had also a better average fit than the corresponding Prony decomposition, the statistical significance of the superiority of STMCB at these three azimuths was not confirmed.

Since STMCB was significantly better than Prony for most of the azimuth angles studied, and it still had a better average fit for the three exception cases, it seems reasonable to recommend the use of STMCB signal approximation methods for HRIR modeling.

Acknowledgment

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] 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.

[3] 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>.

[4] 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.

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

[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/>.