The Manila Journal of Science ***8***:1 (2012) pp. 19-31

Evaluation of Sonic Fidelity of Full and Approximated HRTFs With Reverberation From the Sonic Fidelity of Loudspeakers

Elson Chiu, Archie Cullano, Garyl Gordiel, Edward Kung, and Clement Ong*
Computer Technology Department, College of Computer Studies, De La Salle University
2401 Taft Avenue, Manila, Philippines
*Email: clem.ong@delasalle.ph

The goal of a music playback system is to reproduce as closely as possible the sound experience of live music. Despite the excellent frequency response afforded by smaller, lighter transducers, headphones produce an unnatural "in the head" sound experience that many acute listeners find distracting.

Head-related transfer functions (HRTFs) are an individualized summarization of the direction-dependent acoustic filtering a free-field sound undergoes due to a person's head, torso, and pinna varying as a function of source position and having large intersubject variation. The common acoustical pole and zero (CAPZ) model requires far fewer variable parameters to represent HRTFs. In this study, different approximations of the CAPZ model are processed and evaluated in their ability to emulate the external sound field of loudspeakers while headphones are worn.

The subjective results show that there was no audible drop in quality when HRTFs were incorporated to the sound and that no approximation was singled out as having the best sound quality, but it was observed that the amount of balance between the poles and zeros had an audible effect to the listeners.

1. INTRODUCTION

Earphones work in a way wherein each ear can only hear the sound coming from its own earpiece. This means that there is no natural way for the sound that is produced by the left earpiece to be heard by or to reach the right ear and vice versa. This creates an unnatural experience for the listener since the sounds are perceived to be coming from inside the head (Wang, Yin, & Chen, 2008). This has a completely different experience compared to listening to a standard stereo system using loudspeakers. This kind of system would be able to reproduce the closest, if not exactly like, the original fidelity of the music as it was recorded.

When using loudspeakers, the sound experiences attenuation, reflection, diffraction, etc., from the outer environment before arriving at the ears (Moorer, 2009). Interaural time differences (ITDs) and Interaural level differences (ILDs) are important parameters for the perception of sounds originating from the horizontal plane. ITDs are described to be the time difference in the arrival times of a sound's wave front at the left and right ears. ILDs are the difference in amplitude generated in the left and right ears by a sound. A sound is perceived to be closer to the ear at which the first wave

front with the greater amplitude arrives (Cheng & Wakefield, 2001). ITDs and ILDs do not describe a unique spatial location; therefore, the ability to localize in the median plane is attributed to a monaural hearing mechanism that relies on the spectral coloration of a sound produced by the torso, head, and external ear or pinna.

The unnaturalness of the sound that is produced by the earphones can be removed with the use of head-related transfer functions (HRTFs) and incorporating reverberation to the sound file. HRTFs are typically stored as impulse responses called head-related impulse responses (HRIRs). There is a unique set of HRIRs for each azimuth and elevation for the left and the right ears. In this case, the HRTFs to be used are measured by using human test subjects. The reverberation recreates the sound that bounces from the walls and the floors, called the early reflection, and the sound that bounces several times across the room, called the late reverberation. Reverberation needs to be incorporated to the sound since HRTFs only give the spatial perception of where the sound comes from; it does not include the sound propagation characteristics, sound reflections in the room, etc. This is because most, if not all, HRTFs are measured in anechoic chambers.

This study discusses the loudspeaker sound field emulation in earphones using HRTFs. This is a system that gives a sound an enhanced "out-of-head" listening experience using a pair of earphones.

2. LOUDSPEAKER SOUND FIELD EMULATION IN EARPHONES USING HRTFs

The system produces a sound file that contains spatial information given by the HRTFs and information on the room's acoustical environment given by the early and late reverberation. The HRTFs used in the system were taken from the Listen HRTF database website (Warusfel, 2003). The site used human test subjects to measure the HRTFs and features the measurements of 51 individuals. There are different HRTFs for each level of azimuth and elevation represented as HRIRs. The elevations measured for ranges from -45° to 90° with 15° increments. For the -45° to 45° degree elevation, there are 24 azimuth positions ranging from 0° to 345° with 15° increments. The 60°, 75°, and 90° elevations have 12 (30° degree increments), 6 (60° degree increments) and 1 azimuth position, respectively. Each HRIR is sampled at 44.1 KHz and quantized to 24 bits. In this database, there are a total of 187 HRIRs per person, with each WAV file storing an HRIR pair corresponding to the person's left and right ears. An HRTF represented in this way (i.e., as a long FIR filter derived from the HRIR) is an all-zero model and requires a large number of coefficients to be able to realize the strong peaks and dips in response – each HRIR in the Listen database is in fact 8,192 samples large.

HRTFs are individualized and different for each ear and are dependent on the source position of the sound, leading to a theoretically infinite number of HRTFs per individual. It has been found, however, that it is possible to reduce the complexity of representation by approximating the HRTFs using pole-zero modeling, and reduce this even further by use of the common acoustical pole and zero (CAPZ) model (Haneda, Makino, Kaneda, & Kitawaki, 1999): in essence, the peaks and dips associated with the ear pinnae's resonant structure and other nonchanging characteristics correspond to the poles of the filter (and thus, the poles are constant), while the zeros relate to the difference in position of the sound source versus the listener. The reduction in filter complexity is realized by the nonchanging poles being constant for any direction, making the coefficients of multiple, discrete sound sources reduced by having to add only the zeros of each sound direction – see Haneda, et al. (1999) for a numerical comparison of an all-zero model versus CAPZ.

HRTFs approximated with the use of CAPZ trade reduced complexity for deviation in frequency and phase response (just as any

approximation does), and thus, it is important to determine the how one affects the other and determine any optimal point between filter complexity and sound quality. The target simulated environment is an acoustic space (room) with a pair of loudspeakers placed at a 30° angle to the listener, at a distance between 1 to 2 m.

The emulation of the room's acoustical environment is comprised of four phases: preprocessing, early reflection, late reverberation, and decorrelation. The parameters used for the early reflection and late reverberation are actual measurements of a room where the subjective evaluation of the sound is also held.

2.1 HRTF Module

The HRTF module incorporates necessary spatial characteristics to the original input sound in order to simulate the position of two sound sources that are 30° to the left and the right of the listener, as shown in Figure 1. To test the system, full HRTFs as well as the CAPZ approximation of these HRTFs were used. In this module, the elevation of the simulated sound sources was varied. The elevations used where 0°, 15°, 30°, and 45°.

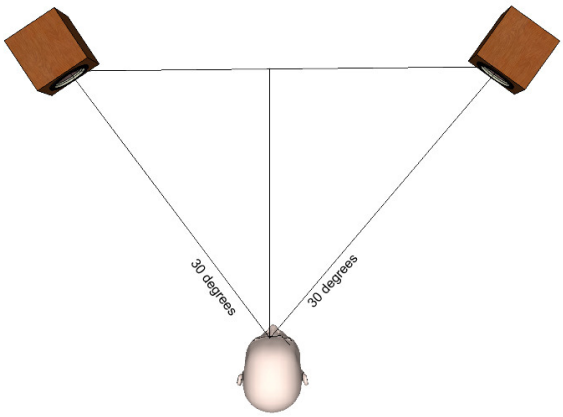

Figure 1. Emulated Setup of Azimuth

Figure 2 illustrates where a sound source with a 30° elevation would be. For the full HRTF manipulation, the left channel of the original sound file was convolved with the direct sound of the left HRIR and the crosstalk of the right HRIR. The right channel, on the other hand, was convolved with the direct sound of the right HRIR and the crosstalk of the left HRIR. After this, ITD was incorporated to both of the crosstalk channels to simulate the late arrival time of the crosstalk sound compared to the arrival time of the direct sound. The new left channel produced for the output is a combination of the left direct sound and right crosstalk, while the new right channel is a combination of the right direct sound and the left crosstalk. The CAPZ manipulation has the same basic concept, but instead of convolving the HRIRs to the corresponding channel, the system filters the channels using the equivalent CAPZ approximation of the HRTFs used in the full HRTF manipulation. The CAPZ approximation of the full HRTFs used was first computed, and once computations were accomplished, the resulting poles and zeros were then used to create a filter. A detailed explanation on how to compute for the CAPZ values can be found in Haneda, et al. (1999).

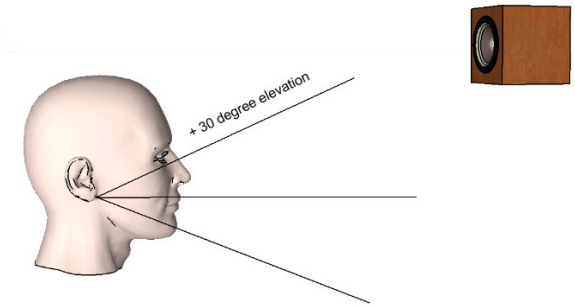

Figure 2. Emulated Setup of Elevation

2.2 Reverberation Module

The original raw stereo sound file is processed with reverberation in parallel with HRTF. The reverberation module incorporates the early reflection of and the late reverberation to produce a simulated sound reflection of a room as shown in Figure 3.

The raw sound file, which is in stereo format, is passed through a filter that incorporates a roll-off to emulate the generally lower reflection coefficient of the room surfaces

and materials as frequency increases. The stereophonic signal is converted to a monophonic source prior to being passed to the reverberation algorithm, to emulate the mixing effect of multiple reflections. The sound would then be incorporated with early reflection and late reverberation to include the emulated room characteristics. Finally, the sound is decorrelated to produce a wide and diffused reverberation image.

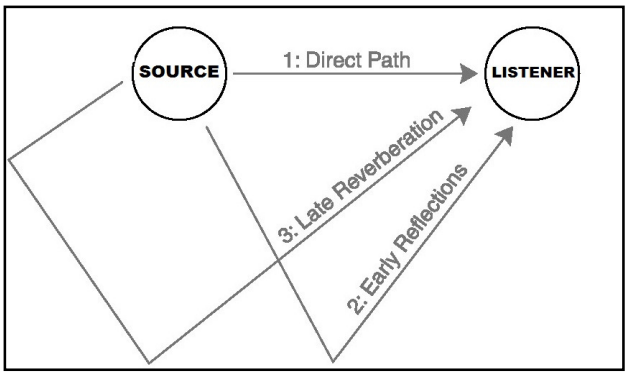

Figure 3. How a Monophonic Sound Travels in a Room and is Perceived by the Listener

2.2.1 Early Reflection

The reflected sounds from walls and floors of a room would require a room impulse response (RIR) filter that will simulate the early reflection by emulating a virtual room modeled from an actual room. The RIR would require the size of the room, the average of reflective coefficient in six frequencies, and the positions of the source and listener. The reflective coefficient or the amount of sound that is reflected by the room is inversely proportional to absorption coefficients.

2.2.2 Late Reverberation

The late reverberation was simulated using an infinite impulse response (IIR) filter where the original signal is processed using six parallel comb filters followed by an all-pass filter, as suggested by Giesbrecht, McFarland & Perry (2009) and Wang, et al. (2008) and described by Moorer (2009), with each comb unit having a different feedback gain computed using coprimes, to avoid flutter echo. The reverberation time or RT_{60} determines the upper value of the feedback gain of any of the comb filters, to mimic how sound decays in the room. The RT_{60} depends on the materials inside the room because everything absorbs sound; typically, the higher the frequencies, i.e., above 4 KHz, the greater the absorption of sound energy.

There are different ways to determine RT_{60}: the Sabine Formula requires the room volume and the (sound) absorption area as well as estimates of the absorption coefficients of the room materials, such as the walls and furniture (see, for example, Sengpiel [n.d.] for an online version). Alternatively the room in question can be measured directly, using specialist equipment or, as in the case of this study, using computer-based general audio test systems with RT_{60} measurement capabilities such as PRAXIS (Waslo, 2009).

2.3 Mixing and Multiplexing

After the reverberation and HRTF modules, the mixer module convolves the new left and right channels from the HRTF module with the reverberation for the left and right ears, respectively. These two signals will then be passed to the multiplexing module, where the left and right audio signals from the mixer are interleaved for wave file encoding. The WAV file is used for the evaluation.

3. IMPLEMENTATION

The raw WAV sound file is a digital-perfect CD extract. The rip is a 25-s clip of one of the songs from the "Best Audiophile Voices" audio CD. This clip is passed to the HRTF module and the reverberation module. The HRTF module uses the HRIR from the Listen database. The output sound of the module is called SFAlpha for referencing.

The processed sound of the HRTF module is approximated with the use of the CAPZ model to reduce the number of coefficients to represent the HRTFs. This is called SFBeta for referencing. To define the combination of poles and zeros that are used for testing, the approximated sounds' frequency responses are

manually inspected. An example of a combination of pole and zero's frequency response is shown in Figure 4. There are a total of five combinations of poles and zeros that are used: 20 poles (P)-40 zeros (Q), 20P-230Q, 30P-50Q, 50P-100Q, and 70P-180Q. Each combination is subjectively evaluated by rating its clarity, brightness, nearness, spaciousness, and its sound quality on a scale of 0–100.

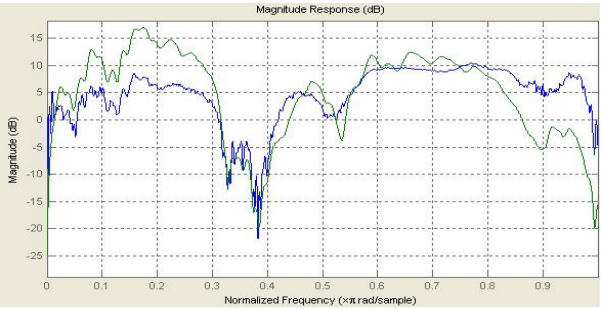

Figure 4. Frequency Response of a 70P – 180Q CAPZ Approximation (Green) Plotted Against Full HRTF Frequency Response (Blue)

The reverberation module handles the emulation of the room's acoustical environment that affects the sound's characteristics. There are four phases that the reverberation module follows: preprocessing, early reflection, late reverberation, and decorrelation. The preprocessing phase converts the stereophonic original WAV sound file to a monophonic sound which then passes through a low pass filter. The early reflection phase is recreated by the RIR, modeled from the room shown in Figure 5 and 6 with dimensions of 5.47×3.98×2.78 m. The reverberation module can handle varying reflective coefficients versus frequency. The LPF response was based on the room's measured RT_{60} for various frequency bands. In general the filter cutoff was near 6 KHz with the response 40 dB down by 12.5 KHz. The RIR function used in the system is based on McGovern's (2004) room impulse response generator. The preprocessed sound and RIR are convolved using Perry's high speed convolution (Giesbrecht et al., 2009). The late reverberation phase uses six parallel comb filters cascaded to one all pass filter. The six comb pass filters have coprime values, and the all-pass filter has a 6-ms delay. The RT_{60} characteristic of the room to be emulated by the reverb was measured with the use of PRAXIS, using a calibrated condenser microphone and a high-performance sound card. The reverb sound is then decorrelated to produce a stereophonic reverberation field. The decorrelation is based on the mono to stereo upmixing using decorrelation (Lundkvist & Oman, 2009).

Figure 5. Back Portion of the Sound Test Room

Figure 6. Front Portion of the Sound Test Room

Three qualitative tests were run to provide a detailed evaluation of the HRTF-modified music. There were a total of 10 listeners present in each of the tests and were same throughout the whole test process. The listener profiles are differentiated by their self-impression of listening acuity, where only one participant out of the 10 felt he or she was not an acute listener, and by the average number of

listening hours on earphones per day; the distribution is shown in Table 1.

Table 1. ***Average Daily Listening Time of Respondent***

Daily Hours of Listening	Number of Respondents
1	2
2	3
3	1
4	2
5	1
6	1

The qualitative tests are administered through a series of listening sessions and used to determine the following: the optimal sampling rate and bit depth of the sound, sound quality of the different CAPZ approximations, and the audibility and fidelity of the externalization of the sound.

The quantitative testing comprises of the time and space complexity of the algorithms. The time complexity is the time it takes for the algorithm to finish its processing, and the space complexity is the measure on how much processing power and memory consumption the algorithm uses.

There are five stimuli used in the testing of the optimal sampling rate and bit depth. Multi-stimulus test with hidden reference and anchor (MUSHRA) was the test method. One of the stimuli is the raw sound file following the Redbook CD standard: sampled at 44.1 KHz, 16-bit depth (44/16). The other four stimuli are SFAlphas with 44/16, 44/24, 96/16 and 96/24 characteristics. Listeners are asked to rate the five stimuli from 0 to 100 according to fidelity (higher is better). They do not have prior knowledge as to which or what sound they are listening to so as to lessen the bias present in their ratings. The results are averaged to see which stimuli would have the highest rating according to its quality. The combination of sampling rate and bit depth that are evaluated as the best is used as the characteristics of the sound for the CAPZ performance tests.

CAPZ is an approximation of the full HRTF. For this study the HRTFs for elevations of 0°, 15°, 30°, and 45° were used as reference. Five different SFBetas for each elevation was generated: 20P-40Q, 20P-230Q, 30P-50Q, 50P-100Q, and 70P-180Q. As in the previous set of testing, MUSHRA was also used here – five approximations and a reference HRTF (SFAlpha) at four different elevations give a total of 24 stimuli. The stimuli are evaluated through survey, using a scale of 0–100; the higher the value, the better for each characteristic. The survey included the following qualifiers, with accompanying descriptors:

1. Clarity – music is clear, distinct and pure (versus diffused, blurred, thick)
2. Fullness – music is full or robust (versus thin)
3. Spaciousness – music sounds open, large (versus closed)
4. Nearness – vocalist or instruments sound near or up close to you (versus sounding distant)
5. Fidelity – the reproduction sounds similar to the original.

The CAPZ approximation rated as best is used as a reference for sound externalization performance.

For determining the quality of the externalization of the sound, the CAPZ-modified music through headphones is compared to a pair of loudspeakers playing the unprocessed Redbook-standard music clip. For this evaluation, a modified ABC test method is used. The listeners are not blindfolded and are not asked to find the hidden reference from test music clips. Instead, the listeners rate the similarity of test music clips “B” and “C” to the sound produced by the loudspeakers “A”. The SFBeta (CAPZ approximation) with the best performance is represented as “B” and the

SFAlpha (full-range HRTF processed sound), is represented as "C". The SFAlpha and the SFBeta are mixed with early reverberation based on varying distances mimicking loudspeaker to listener distance. The distances used for the reverberation are 1, 1.5, and 2 m; figure 7 shows the test setup. The test subject is asked to determine the similarity between music clips 'B' and 'C' from music clip 'A'. There are a total of 12 stimuli, stemming from varying listening distance (1m, 1.5m, and 2m) and four elevations (0°, 15°, 30°, and 45°). The test subject is asked to rate sound quality and its spaciousness on a scale 0–100. This is to determine if applying reverberation to the sound would enhance the effect of externalization. After, the test subject is asked to listen to the sound produced by the loudspeaker and listen to music clips "B" and "C" again but rating it according to its similarity with respect to "A" from a scale of 1.0–5.0. The scale is based on a five grade impairment scale as recommended by the International Telecommunication Union (ITU), with 5.0 as excellent, 4.0 to 4.9 as good, 3.0 to 3.9 as fair, 2.0 to 2.9 as poor, or 1.0 to 1.9 as bad.

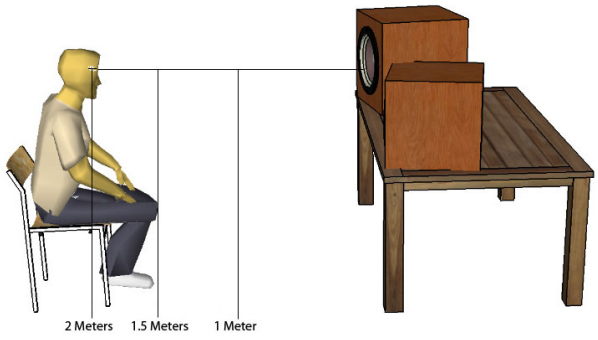

Figure 7. Illustration of the Test Setup

The time complexity was measured by the total run times of the CAPZ approximations and the reverb processes. In quantifying the space complexity, the task manager of the computer was used to check the memory consumption of the whole MATLAB program prior to and during the processing of the approximation. The differences between these memory values represent the memory requirements of the actual processes.

4. RESULTS AND ANALYSIS

The ratings of each listener for the four different stimuli were subtracted from the rating of the original sound file. This is to see the difference in quality that is present with the sounds. If the result was a negative number, it meant that the quality of that sound was poorer compared to the original sound. After the computations for each of the test subject's ratings, the results for each test sound file of each test subject were averaged to see the overall difference of that sound from the original sound. Figure 8 shows the overall rating of the varying SFAlphas. Given the scoring scale of 0-100, the small averaged difference values reflect the fact that any perceived changes from the reference sound are small.

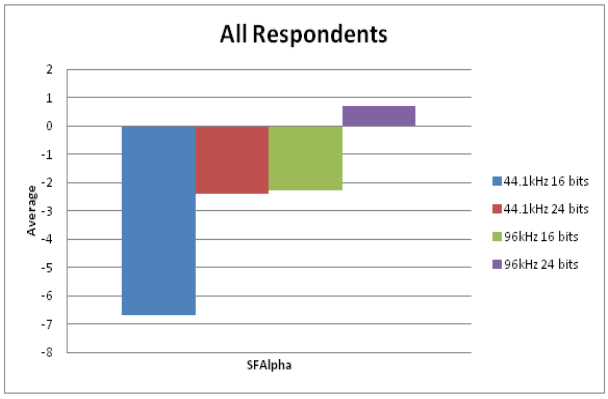

Figure 8. SFAlpha Overall Rating Against the Original Sound

Based on the results shown in Figure 8, the SFAlpha with a sampling rate of 96 KHz and a bit depth of 24 bits were rated as the best in terms of sound quality. Listeners rated the SFAlpha with a sampling rate of 44.1 KHz and a bit depth of 16 bits as the worst. The results with the sampling rate of 44.1 KHz and a bit depth of 24 bits and the sampling rate of 96 KHz and a bit depth of 16 bits were perceived to have almost no audible difference with one another, having an averaged difference of only 0.1. In this first test, it was also observed that the listener's familiarity with sound produced by earphones have an effect to their ratings. Figures 9 and 10 show the rating of test subjects

who listened for more than 4 hours on an average per day and those who listened for less, respectively.

Figure 9 reflects how listeners also rated the SFAlpha with a sampling rate of 96 KHz and a bit depth of 24 bits as the best, but its overall sound quality is still less compared to the original WAV sound. The clips that are upsampled to 96 KHz are also rated better compared to the sound with a sampling rate of 44.1 KHz.

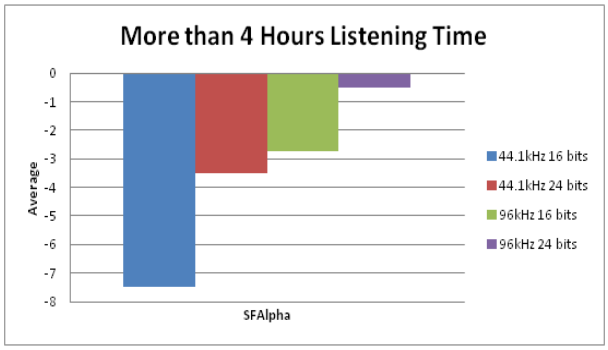

Figure 9. Rating of the Test Subjects With 4 or More Hours of Listening Time

Figure 10 shows that listeners who practiced less than an average of four hours of daily listening still rated the SFAlpha with a sampling rate of 96 KHz and a bit depth of 24 bits as best, while sounding even better than the original music.

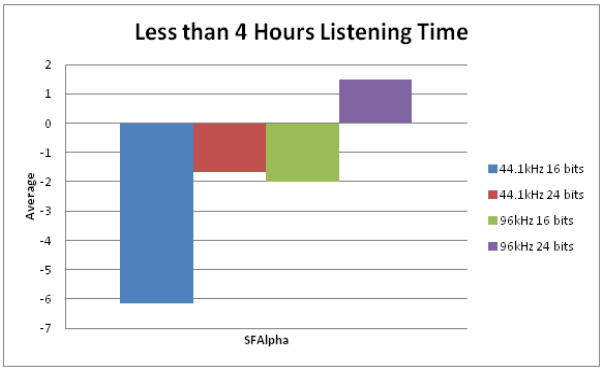

Figure 10. Rating of the Test Subjects With Less Than 4 Hours of Listening Time

The different approximations were evaluated separately by their elevation to determine if this characteristic would result in audible quality differences. Figure 11 shows the subjective results of the SFBeta with a 0° elevation. The SFBeta with an approximation of 30P-50Q was rated overall as the best.

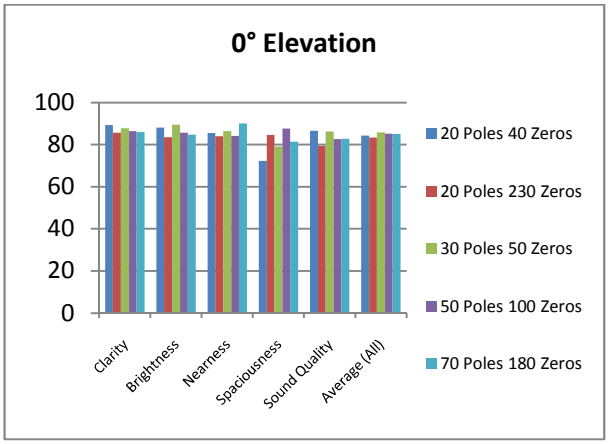

Figure 11. Subjective Evaluation of Different CAPZ Approximation With 0° Elevation

The 70P-180Q approximation had the highest rating in terms of nearness, which means that the approximation was interpreted by the listener as having the least externalized sound among the other approximations. The approximation with 20 poles and 230 zeros was rated as having lowest fidelity, and it seems that increasing the CAPZ filter's zeros without any change in poles would have an audible degradation according to the listeners.

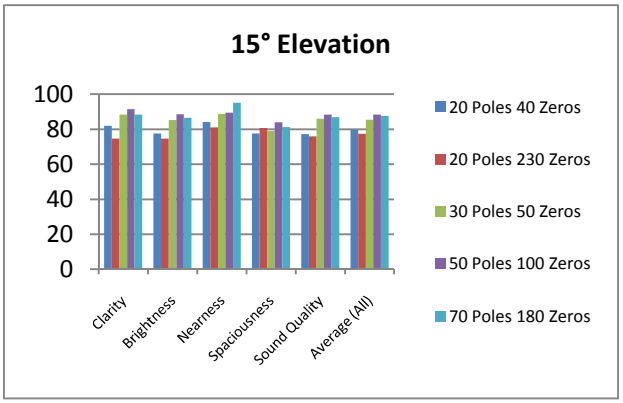

Figure 12. Subjective Evaluation of Different CAPZ Approximation With 15° Elevation

Figure 12 shows the subjective results of the SFBeta with a 15° elevation. The SFBeta with an approximation of 50 poles and 100 zeros was rated overall as the best but the 70P-180Q approximation had almost the same overall results. On the basis of sound quality, the 50P-

100Q approximation was rated best while the 30P-50Q and the 70P-180Q approximations had almost the same results. The approximation with 20 poles and 230 zeros was rated as the worst.

Figure 13 shows the subjective results of the SFBeta with a 30° elevation. The SFBeta with an approximation of 50 poles and 100 zeros was rated overall as the best while the 30P-50Q approximation had almost the same results. The 50P-100Q approximation was rated as the most spacious and at the same time the nearest sound to the listener. On the basis of sound quality only, the 30P-50Q and 50P-100Q approximation were rated as having the best sound quality. The approximation with 20 poles and 230 zeros was again rated as the worst.

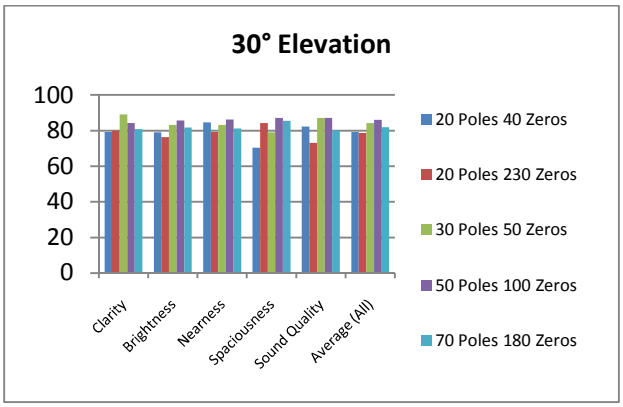

Figure 13. Subjective Evaluation of Different CAPZ Approximation With 30° Elevation

Figure 14 shows the subjective results of the SFBeta with a 45° elevation. The SFBeta with an approximation of 30 poles and 50 zeros was rated overall as the best while the 70P-180Q was rated similarly. The 70P-180Q approximation was rated as the most spacious and the 50P- 100Q approximation was judged as having the nearest sound according to the listener. On the basis of sound quality, the 30P-50Q approximation was rated as the best sound quality. The approximation with 20 poles and 40 zeros was rated as the worst, with the 20 poles and 230 zeros approximation rated similarly.

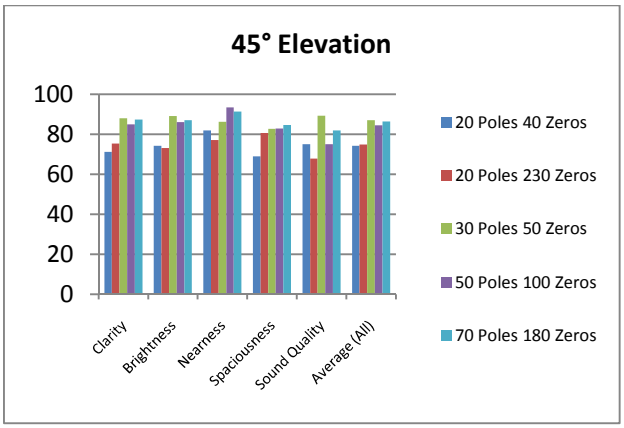

Figure 14. Subjective Evaluation of Different CAPZ Approximation With 45° Elevation

Figure 15 shows overall subjective evaluation of the different CAPZ approximations. Based on the results given for each of the CAPZ approximations, the 30P-50Q approximation generally had the best overall sound, and was thus chosen as the reference for the next test. The 20P-40Q approximation had a consistent degradation with elevation changes, while the CAPZ approximation of 20P-230Q had the overall worst sound quality with varying elevation.

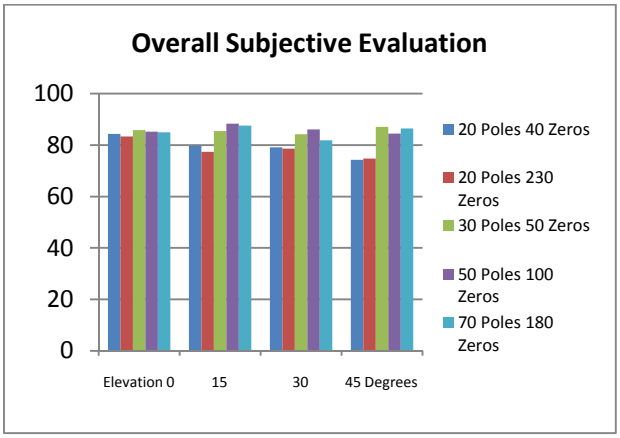

Figure 15. Overall Subjective Evaluation of the CAPZ Approximation

The results of the SFAlpha and the SFBeta ABC testing were evaluated separately according to their elevation and distances, to determine if there were noticeable relations between the overall sound quality with elevation and distance.

Figures 16 and 17 show the results of the SFAlpha and the SFBeta with an elevation of 0°. The SFAlpha with a reverberation parameter that was measured at 2 m was considered as the best in terms of its overall sound quality. Having an emulated distance of at least 2 m from the sound source improved the overall listening experience of the test subject when using earphones as its medium of sound reproduction. The test subject's ratings with music clips with a smaller emulated distance were not perceived to have good sound quality. The listeners were also able to perceive a more spacious room when the emulated distance between listening position and the sound source was increased. The approximation with a reverberation parameter that was measured at 1.5 m was considered as the best in terms of sound quality and a distance of 2 m from the loudspeakers was rated as the most spacious. The listeners had difficulty perceiving the spaciousness when the emulated listening distance was 1.5 m or less.

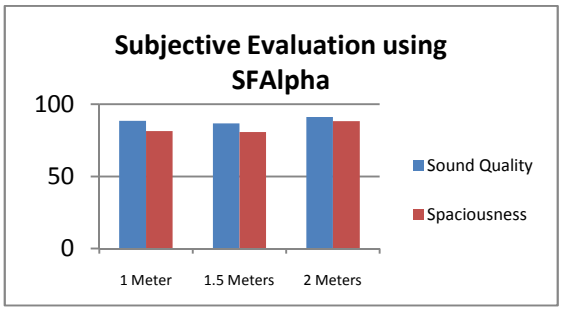

Figure 16. Full-range HRTF Sound With 0° Elevation

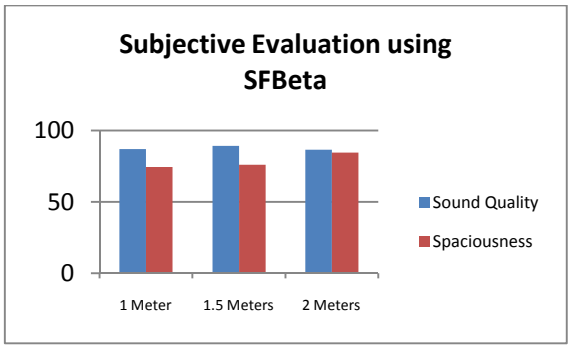

Figure 17. CAPZ Approximated Sound With 0° Elevation

Figures 18 and 19 show the result of the SFAlpha and the SFBeta listening tests with varying elevations and distances. The test music clips had emulated distances of 1, 1.5, and 2 m, had 45°, 30°, and 15° elevations respectively. When elevation was incorporated into the sound files the overall sound quality of the test music clips changed. The SFBeta clips with an elevation of 30° and a distance of 1.5 m, as well as and an elevation of 15° and a distance of 2 m were perceived by listeners as having the best overall sound quality. Listeners also commented that there was a perceptible change in the sound when emulated distance was changed from 1.5 to 1 m.

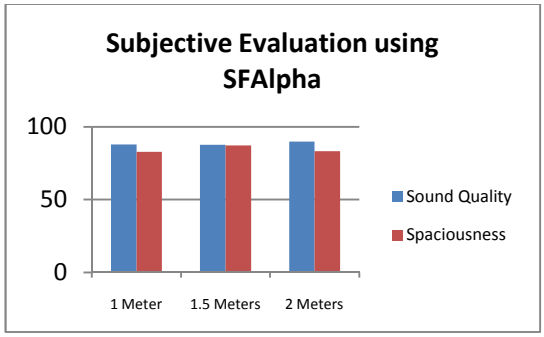

Figure 18. Full-range HRTF Sound Evaluation With Varying Elevation and Distance

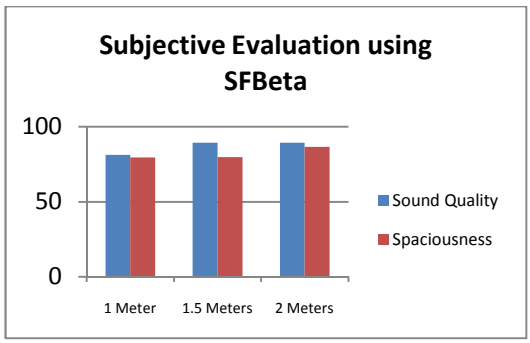

Figure 19. CAPZ Approximated Sound Evaluation in Varying Elevation and Distances

For externalization testing, the results for the SFAlpha with varying elevation and distances are shown in Figure 20.

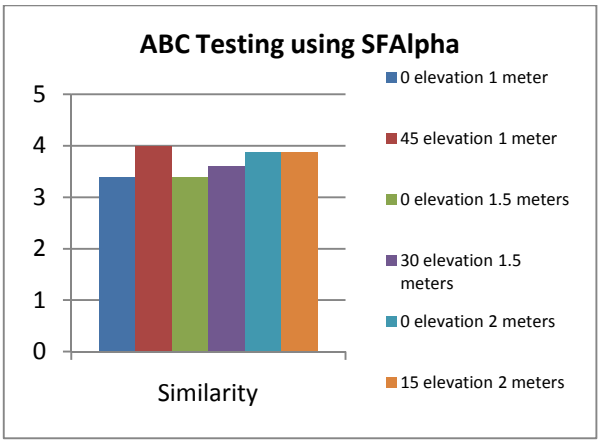

Figure 20. Similarity of Full-Range HRTFs to the Original Sound

The scale used in the externalization tests of the SFAlpha and the SFBeta ranges from 5.0, the highest grade, which would mean complete or very similar emulation of loudspeaker externalization, down to 1.0, with no hint of externalization. All the ratings were in the range of 3.4–4.0 which means that reproductions were perceived to have good similarity between the test music clips and the loudspeaker. Judging by the results, using the full-range HRTFs to implement externalization to the sound, a simulated distance of 1 m and an elevation of 45° from the loudspeaker is optimal. With an emulated distance of 2 m, 0° and 15° elevations had little, if any, perceived differences between each other, thus resulting in similar sound quality ratings.

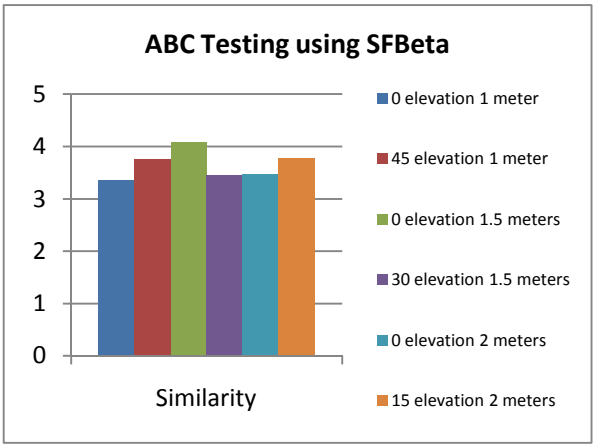

Figure 21. Similarity of CAPZ Approximation to the Original Sound

The results for the SFBeta with varying elevation and distances are shown in Figure 21. The ABC test results range from 3.3–4.1, which mean the CAPZ approximations were also perceived as having good similarity with the sound produced by a pair of loudspeakers. The SFBeta with an elevation of 0° and a distance of 1.5 m was rated as the closest to the sound produced by a pair of loudspeakers. Comparing results with SFAlpha, most of the ratings were similar, with the exceptions of 45° elevation at 1 m, and 0° elevation and a distance of 1.5 m.

Time and space complexity results were based on this study's software implementation, which used the MATLAB development environment on a Windows-based PC. The input sound file was a 25-s music clip, upsampled to 96/24. Table 2 shows the time complexity of the processing of the CAPZ approximation. As expected, an increase of poles and zeros resulted in more time required to execute. There was also an increase in memory usage with greater numbers of approximating poles and zeros.

Table 2. ***Time and Space Complexity of Three Different CAPZ Approximations***

Number of Poles (P) and Zeros (Q)	20P–40Q	30P – 50Q	20P – 230Q
Processing Time (ms)	151.64	186.02	387.72
Memory (KB)	10,904	13,956	16,192

Based on the values in the table, the processing time and memory consumption appear to be proportional to the poles and zeros. In particular, the processing time scales very linearly if an overhead of ~75 ms is applied, resulting in computation load of ~1.3 ms per coefficient.

In this study, the full HRTFs were applied as a zero-order representation, i.e. as an equivalent FIR filter based on the HRTF's impulse response. In order to fairly compare the complexity of such against CAPZ, Haneda, et

al. (1999) suggests that equivalent pole-zero filters, if used to represent a multisound source placed horizontally every 10° around the listener, would require around 32% more coefficients. It can be expected then that the space and time requirements would increase by roughly same percentage if the space-time relationship is truly linear.

The time complexity for the emulation of reverberation is shown in Table 3. Based on the results, the late reverberation phase takes a significant time: 110 to 113 s to finish.

Table 3. ***Time Complexity for Reverberation (in Seconds)***

Distance	1 m	1.5 m	2 m
Preprocessing	0.14	0.15	0.15
Room Impulse Response	0.12	0.14	0.13
Convolution	2.26	5.99	4.43
Early Delay	4.63	4.65	4.66
Comb and All Pass Filter	113.53	112.41	110.67
Total	**120.68**	**123.32**	**120.04**

5. CONCLUSION

It can be seen that the time it takes to process a CAPZ approximation is not long relative to the real-time playback of the sound clip of 25 s. The largest approximation used in the test, 20P-230Q, took slightly under 0.4 s while the approximation that garnered the best results in the listening tests, 30P-50Q, took about half the time, being just under 0.2 s. The memory space taken up during processing for the largest approximation took up 16,192KB while the approximation with the best results took up 13,956KB. Due to the time it takes for the reverberation process, it is impractical to implement the system in real time without any algorithm enhancements, at least while the system is on MATLAB.

The results for the CAPZ approximated HRTFs with added reverberation show that it is able to produce sounds with an average loudspeaker externalization similarity rating of 3.6 out of 5.0. The results for the full-range HRTFs with added reverberation show that it is rated with an average of 3.66 out of 5.00, performing only slightly better than CAPZ approximated HRTFs in terms of externalization. This indicates that the system is able to produce a sound with satisfactory externalization for the listeners. It can also be seen from the results that the CAPZ approximation does not lag far behind the full range of HRTFs in terms of sound quality and spaciousness.

6. RECOMMENDATIONS

Individuals have their own unique HRTF characteristics that might affect their subjective evaluation of music reproduction. Having a different HRTF used in the processing of the sound would be perceived by the listener as either good or bad depending on the closeness of the HRTF characteristics to the listener's own HRTF characteristics. The implementation of a high-definition sound localization system can be achieved using a set of HRTFs of the listener, but HRTFs that are degraded or not identical to a listener's own HRTFs can cause front-back confusion and inaccurate localization of sound (Watanabe, Ozawa, Iwaya, Suzuki, & Aso, 2007). HRTF measurements are not the only one that are unique for each person; the ITDs and ILDs are also unique. These ITDs and ILDs are well-known localization cues, and it has been shown that both ITDs and ILDs are important parameters for the perception of sounds originating from the horizontal plane (Cheng & Wakefield, 2001). Researching the different effects of the ITDs and the ILDs may be another area of development for the improvement of the overall externalization and localization of the output.

The hybrid reverberation algorithm was used to produce an accurate impulse response of the actual room in order to model a virtual room. It was observed that it took a significant

amount of processing time; despite this it was still used, since the system's objective was to produce the best quality output sound. Further reduction of the execution time of the reverberation algorithm while maintaining its accuracy would be a worthwhile research goal.

Acknowledgement

We would like to thank Dr. Joel Ilao and Mr. Macario Cordel II for their help in some digital signal processing concepts and MATLAB functions. We would also like to thank Mr. Karlo Campos and Mr. Christian Echavez for their helpful suggestions on testing methods.

References

Cheng, C., & Wakefield, G. (2001). Introduction to head-related transfer functions (HRTF's) representations of HRTF's in time, frequency, and space. *Journal of the Audio Engineering Society, 49*(4), 231-249.

Giesbrecht, H., McFarland, W., & Perry, T. (2009). Algorithmic *reverberation:* Combining Moorer's *reverberator with simulated room* IR *reflection modeling*. Retrieved from http://web.uvic.ca/~timperry /Elec407-HybridAlgorithmicReverb/ index.html

Haneda, Y., Makino, S., Kaneda, Y., & Kitawaki, N. (1999*).* Common-acoustical-pole and zero modeling of head-related transfer functions. *IEEE Trans. on Speech and Audio Processing, 7*(2), 188-195.

Lundkvist, A., & Oman, P. (2009 March 26). *Mono to stereo upmixing using decorrelation.* Retrieved from http://www.student.ltu.se/~petohm-5/S7006E/Project/

McGovern, Stephen G. (2004). A model for room acoustics. Retrieved from http://www.sgm-audio.com/research/rir /rir.html

Moorer, J. (2009). About this reverberation business. *Computer Music Journal, 3*(2)*,* 13-28.

Sengpiel, E. (n.d.) *Calculation of the reverberation time*. Retrieved from http://www.sengpielaudio.com/calculator-RT60.htm

Wang, L., Yin, F., & Chen, Z. (2008). An "out of head" sound field enhancement system for headphone. In Proceedings of the 2008 International Conference on Neural Networks and Signal Processing (pp. 517-521). Zhenjiang, China: IEEE.

Warusfel, O. (2003). Listen HRTF *database*. Retrieved from http://www.recherche. ircam.fr/equipes/sales/listen/index.html

Waslo, B. (2009). PRAXIS, by Liberty Instruments, Inc. (Version 2.52) [Software and hardware]. Available from http://www.libinst.com/PRAXIS.html

Watanabe, K., Ozawa, K., Iwaya, Y., Suzuki, Y., & Aso, K. (2007). Estimation of interaural level difference based on anthropometry and its effect on sound localization. *Journal of the Acoustical Society of America, 122*(5), 2832-2841.