RENDERING BINAURAL ROOM IMPULSE RESPONSES FROM SPHERICAL MICROPHONE ARRAY RECORDINGS USING TIMBRE CORRECTION

Jonathan Sheaffer sheaffer@ee.bgu.ac.il

Shahar Villeval shaharv@ee.bgu.ac.il

Boaz Rafaely br@ee.bgu.ac.il

Acoustics Lab,
Dept. of Electrical and Computer Engineering
Ben-Gurion University of the Negev, Israel.

ABSTRACT

The technique of rendering binaural room impulse responses from spatial data captured by spherical microphone arrays has been recently proposed and investigated perceptually. The finite spatial resolution enforced by the microphone configuration restricts the available frequency bandwidth and, accordingly, modifies the perceived timbre of the played-back material. This paper presents a feasibility study investigating the use of filters to correct such spectral artifacts. Listening tests are employed to gain a better understanding of how equalization affects externalization, source focus and timbre. Preliminary results suggest that timbre correction filters improve both timbral and spatial perception.

1. INTRODUCTION

Binaural technology [1] provides a means to render headphone-presented stimuli that mimic sounds as if they were heard in a regular, headphone-free listening situation. It has applications in psychoacoustics research [2], auditory neuroscience [3], architectural acoustics [4] and audio technology [5]. In its simplest form, binaural technology utilizes free field head-related transfer-functions (HRTFs) of a mannikin or an individual subject to spatially filter an audio signal. While such processing provides a basic means to simulate sound localization, the absense of reverberation and time-varying auditory cues have a negative effect with regard to achieving an accurate sense of sound externalization [6, 7].

To obtain a more complete set of auditory cues, one may opt to directly measure the transfer function of a room using a manikin, which results in a Binaural Room Impulse Response (BRIR), e.g. see [8]. This, however, results in a single transfer function combining the effects of the room itself as well as the head, ears and torso, and thus represents the anthropometric features of a specific listener. Additionally, if one wishes to reproduce the effects of head movements, the BRIR needs to be measured in a range of head orientations hence making the measurement procedure inefficient for many practical applications.

Rafaely and Avni [9] suggested a method to render BRIRs in the spherical harmonics (SH) domain, by making use of pre-measured HRTFs and a Spatial Room Impulse Response (SRIR) which can be obtained either by means of numerical simulation or by direct room measurement using a spherical microphone array. More recently, Avni et al. [10] studied the perceptual effects of recording and reproducing sound fields at different spatial resolutions. Among their findings, they discovered that limiting the SH order of the recorded sound field has a prominent effect on the frequency bandwidth of the resulting BRIR, and hence on the timbre of the played-back material. In other words, a BRIR of low spatial resolution also has a limited frequency bandwidth, which indicates that the spatial and spectral design parameters of BRIRs should not be seen in isolation. To address this, Villeval [11] suggested a timbre equalization filter, compensating for the average change in frequency response between BRIRs constructed at two different spatial resolutions.

The results presented in [10] showed that there is an inherent trade-off between the spatial resolution of the sound field recorded with a spherical array and the spectral representation of the resulting BRIR. As a first step towards addressing this trade-off, this paper presents a feasibility study on the effects of correcting low-order BRIRs with a timbre equalization filter. The paper is structured as follows: Sec. 2 and 3 briefly outline the procedure for computing a BRIR in the SH domain, and for equalizing it to a desired SH order. This is followed by experimental results from a preliminary listening test in Sec. 4, which are further discussed in Sec. 5.

2. RENDERING BINAURAL ROOM IMPULSE RESPONSES

To render a BRIR from sound pressure measured by a spherical microphone array, the method suggested in [9] is followed in this paper. Let $H^l(k,\Omega)$ and $H^r(k,\Omega)$ denote a set of pre-measured HRTFs for the left and the right ear, respectively, where $k = 2\pi f/c$ is the acoustic wavenumber, f is the frequency and c is the speed of sound in air. Here, $\Omega \equiv (\theta,\phi) \in S^2$ denotes the angle in a standard spherical coordinate system [12] in which (r,θ,ϕ) denote radial distance, elevation and azimuth, respectively. By applying the spherical Fourier transform [13] to $H^l(k,\Omega)$ and $H^r(k,\Omega)$, one obtains their respective representations in the SH domain, $H^l_{nm}(k)$ and $H^r_{nm}(k)$.

Similarly, let $p(k,r,\Omega)$ denote some pressure function on a sphere that is square integrable over Ω and whose spherical Fourier transform yields the function $p_{nm}(k,r)$. In a room, this function represents spatial information on a continuum of plane waves arriving at the receiving position from the sound source and the different reflecting surfaces. The complex amplitudes of the spherical harmonic components of these plane waves, $a_{nm}(k)$, can be obtained by performing a plane-wave decomposition of the sound field as follows [14]:

$$a_{nm}(k) = \frac{p_{nm}(k,r)}{b_n(kr)}. \tag{1}$$

In this paper all spherical array measurements are performed directly over the surface of a rigid sphere and, as such, $b_n(kr)$ is given by [15]

$$b_n(kr) = 4\pi i^n \left[j_n(kr) - \frac{j_n'(kr)}{h_n'(kr)} h_n(kr) \right], \tag{2}$$

where $j_n(\cdot)$ is the spherical Bessel function, $h_n(\cdot)$ is the spherical Hankel function and $j_n'(\cdot)$ and $h_n'(\cdot)$ represent their first derivatives with respect to the argument. For high values of n, the result of $b_n(kr)$ approaches zero for low values of kr, which requires a large calculation dynamic range. To overcome this numerical limitation, in this paper $b_n(kr)$ is soft-limited according to the procedure suggested in [16].

Once a plane wave decomposition is performed, a BRIR can be calculated as follows [9]:

$$p^l(k) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \tilde{a}_{nm}^*(k) H_{nm}^l(k), \tag{3}$$

where $\tilde{a}_{nm}(k) = (-1)^m a_n^{-m}(k)$ is the representation of $a^*(k,\Omega)$ in the SH domain, and $p^l(k)$ denotes the resulting pressure at the left ear. For the right ear, Eq. (3) is computed with the corresponding right ear HRTF in a similar fashion. In the limiting case, evaluation of the sum in Eq. (3) results in a plane wave representation of the BRIR. In practice, however, the functions $p(k,r,\Omega)$, $H^l(k,\Omega)$ and $H^r(k,\Omega)$ are sampled in space with finite resolution, which implies that the infinite series in (3) must be truncated at some order N to avoid introducing any detrimental effects of spatial aliasing.

3. TIMBRE EQUALIZATION

The practical constraint regarding this series truncation motivates a perceptual comparison of BRIRs generated with different truncation orders. Avni et al. [10] showed that truncating (3) not only restricts the spatial resolution of $p^l(k)$, but also affects its frequency content due to the explicit dependency of b_n on kr and the increased truncation error for $kr > N$ [17], with N being the order limit. This direct impact on the resulting timbre may affect perception and obscure psychoacoustic investigations. To compensate for this effect, Villeval [11] suggested an equalization method, which shall be briefly described in this section.

As a first approximation, the transfer function of the human head can be seen as that of a rigid sphere, which provides an analytic means to quantify the frequency related effects of constructing an incident wave with a finite series of spherical harmonics. In a reverberant setting, the sound field is comprised of a large number of plane waves of random incidence directions. Assuming that a receiver, representing the sound pressure at the ear, is placed at some point Ω_0 on the surface of a rigid sphere, then the average magnitude response over all incident waves is given by [11]

$$\overline{p(kr,\Omega_0)} = \sqrt{\frac{1}{4\pi} \sum_{n=0}^{N} \sum_{m=-n}^{n} |b_n(kr)|^2 \, |Y_n^m(\Omega_0)|^2}, \tag{4}$$

which reduces to

$$\overline{p(kr)} = \frac{1}{4\pi} \sqrt{\sum_{n=0}^{N} |b_n(kr)|^2 \, (2n+1)}, \tag{5}$$

where $b_n(kr)$ is as defined before, and $Y_n^m(\cdot)$ are the spherical harmonics [15]. Accordingly, one can describe the transfer function of a *timbre correction filter*, that equalizes the frequency response of some finite series of order N to that of an order N_h, as follows:

$$H(k)\big|_{N \to N_h} = \frac{\overline{p(kr)}_{N_h}}{\overline{p(kr)}_N}. \tag{6}$$

For example, the magnitude response of two timbre correction filters, equalizing orders $N = 3 \to 19$ and $N = 2 \to 10$, are shown in Figure 1.

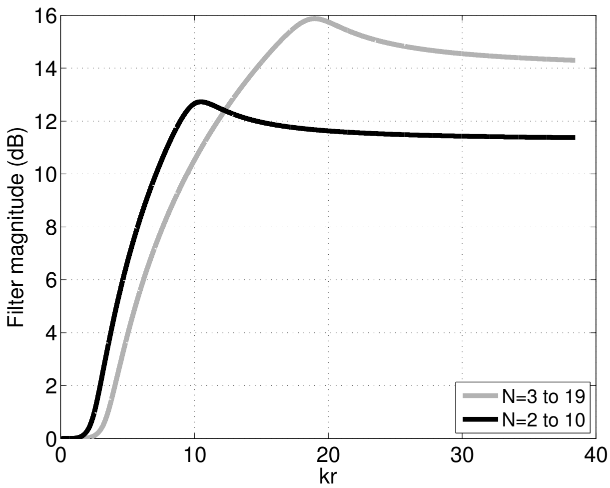

Figure 1: *Magnitude response of two timbre correction filters, equalizing orders* 3 *to* 19 *and* 2 *to* 10.

4. LISTENING TEST

Following the proposed timbre correction filters, the objective of the listening test is to investigate the filters' effects on three perceptual attributes, namely:

1. *Externalization.* When binaurally reproducing a sound field over headphones, the listener may or may not experience a sense of externalization and, accordingly, judges whether the sound is arriving from within the headphones or from a more distant location.
2. *Localization blur / focus.* Whether the sound is externalized or not, it is spatially localized with some error. Thus, a source perceived as having a well defined position in space is here referred to as having a low localization blur or, a high localization focus.

3. *Timbre*. As discussed in Sec. 3, limiting the order of Equation (2) has an effect on the frequency bandwidth. Accordingly, the perceived timbre of the recorded material is modified.

4.1. Methodology

A binaural representation of a sound field was rendered using the method described in Sec. 2, by making use of a pre-measured HRTF set and a SRIR measured using a spherical microphone array. The HRTF set consists of measurements of a Neumann KU-100 manikin, based on a Gaussian sampling scheme with a total resolution of 16020 measurement points distributed around the manikin with no spatial gaps. All technical details regarding the used HRIRs, including the measurement procedure and post-processing can be found in [18]. The chosen SRIR was of the WDR small broadcast studio, having a floor area of 201m^2 and a total volume of 1246m^3 [19], and was sampled using a 1202 points nearly-uniform scheme. Both the HRTFs and the SRIR can be found on-line as part of the the WDR impulse response compilation [20].

To account for head movements (which are required for achieving an effective sense of externalization), BRIR sets were computed for a range of head rotation angles (360° in a 1° resolution), by multiplying $H^l_{nm}(k)$ and $H^r_{nm}(k)$ by respective Wigner-D functions [21]. In the sound reproduction stage, a pair of AKG K702 headphones were fitted with an *Attitude and Heading Reference System* (Razor IMU) which was used to obtain real-time data on the subject's head orientation. All stimuli were processed with a matching headphone compensation filter, were generated pre-test and were played-back using the *SoundScape Renderer* auralization engine [22]. The total latency of the playback system was 5.3ms.

Eleven subjects (all male, ages 24-37) participated in a multiple stimuli, hidden-reference listening test. The labeled reference was based on a BRIR constructed using the method described in Sec. 2, with the SH series truncated at $N = 19$. Similarly, the remaining test samples were based on a BRIR constructed at $N = 3$ (also serving as a low-anchor) and the same BRIR equalized to $N = 19$ using the method described in Sec. 3. To obtain the final test samples, these BRIRs were convolved with anechoic recordings of a classical guitar and of speech. In each screen listeners were asked to rank, on a scale of 1 to 5, how similar each test sample is to the reference in terms of externalization, focus and tone (timbre).

Figure 2 shows the sixth-octave smoothed spectrum of the anechoic speech recording used in the listening experiment. To demonstrate the effects of timbre equalization, the anechoic signal was convolved with three left-ear BRIRs based on $N = 19$ (reference curve), $N = 3$ and $N = 3/E19$ (corresponding to order 3 equalized to order 19). Observe that up to $kr \approx 3$ (equivalent to $f = 1.9$kHz for a sphere of $r = 85$mm), which represents the frequency range of the filter's stop-band, the three curves are nearly identical. Above $kr = 3$, which represents the filter's pass-band, the $N = 3/E19$ curve is amplified compared to the $N = 3$ curve. Because the filter is designed based on the pressure magnitude averaged over all incident directions, then for any room other than a perfectly diffuse field, the accuracy of compensation will always be dependent on the specific directional characteristics of the SRIR.

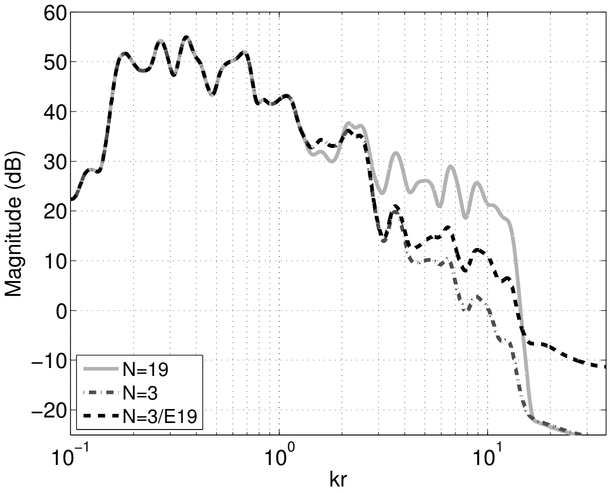

Figure 2: *Sixth-octave smoothed spectra for the anechoic speech recording, after convolving it with left-ear BRIR based on $N = 19$, $N = 3$ and $N = 3/E19$.*

4.2. Results

Figure 3 shows the mean scores ($\bar{x}$) and confidence intervals ($t_{.95,11} = 2.26$) for the different test samples used in the experiment. All eleven subjects gave the hidden reference a score of "5", and as such, this listening condition is excluded from the results.

For the case of timbre, there is a significant difference between equalized and unequalized test samples, for both speech ($\bar{x} = 2.27, \sigma = 0.786$ compared to $\bar{x} = 1.09, \sigma = 0.301$) and classical guitar ($\bar{x} = 2.45, \sigma = 0.82$ compared to $\bar{x} = 1.36, \sigma = 0.67$).

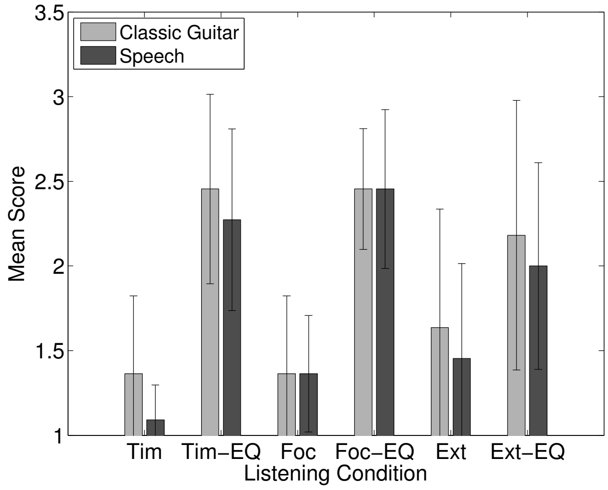

Figure 3: *Mean score and 95% confidence intervals for subjective evaluation of timbre (Tim), focus (Foc) and externalization (Ext), testing for speech (dark bars) and classical guitar (light bars). Equalized BRIRs are marked with "EQ".*

A similar trend is evident in results of the focus test ($\bar{x} = 2.45, \sigma = 0.68$ compared to $\bar{x} = 1.36, \sigma = 0.50$ for speech, and $\bar{x} = 2.45, \sigma = 0.52$ compared to $\bar{x} = 1.36, \sigma = 0.67$ for classical guitar), suggesting a significant spatial improvement.

In the externalization test, however, the improvement in the mean score is smaller (a difference of only $\Delta\bar{x} = 0.54$ points for both classical guitar and speech) and the confidence intervals overlap. While this does not necessarily indicate a lack of statistical significance (the data is dependent), a more rigorous testing is required to investigate this effect.

5. CONCLUDING REMARKS

A preliminary study on the effects of timbre correction filters on low-resolution BRIR rendering was presented in this paper. A perceptual comparison of equalized vs. unequalized samples revealed a significant improvement in the perceived timbre and a reduction of localization blur. Such localization errors are known to decrease as the order of the microphone array is increased [23]. This immediately suggests that spectral equalization of low order array recordings could be beneficial not only for timbre restoration, but also for improving the perceived spatial resolution, most noticeably in terms of sound localization. One possible reason for this result may be related to the recovery of high-frequency auditory cues, which contribute to sound localization [24] as well as the spatial perception of an enclosed space [25].

Future tests will involve a more systematic comparison of listening conditions, taking into account a wider range of truncation orders, room characteristics, source positions and program-materials.

6. ACKNOWLEDGMENTS

The authors would like to thank Barak Ben-Dayan for his technical assistance in implementing the head-tracking system.

7. REFERENCES

[1] Henrik Møller, "Fundamentals of binaural technology," *Applied acoustics*, vol. 36, no. 3, pp. 171–218, 1992.

[2] Frederic L Wightman and Doris J Kistler, "Headphone simulation of free-field listening. i: Stimulus synthesis," *The Journal of the Acoustical Society of America*, vol. 85, pp. 858, 1989.

[3] Simon Carlile, *Virtual auditory space: Generation and applications*, RG Landes New York, 1996.

[4] Hilmar Lehnert and Jens Blauert, "Principles of binaural room simulation," *Applied Acoustics*, vol. 36, no. 3, pp. 259–291, 1992.

[5] Mendel Kleiner, Bengt-Inge Dalenbäck, and Peter Svensson, "Auralization-an overview," *Journal of the Audio Engineering Society*, vol. 41, no. 11, pp. 861–875, 1993.

[6] Nathaniel I Durlach, A Rigopulos, XD Pang, WS Woods, A Kulkarni, HS Colburn, and EM Wenzel, "On the externalization of auditory images," *Presence: Teleoperators and Virtual Environments*, vol. 1, no. 2, pp. 251–257, 1992.

[7] Durand R Begault, Elizabeth M Wenzel, and Mark R Anderson, "Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source," *Journal of the Audio Engineering Society*, vol. 49, no. 10, pp. 904–916, 2001.

[8] Barbara G Shinn-Cunningham, Norbert Kopco, and Tara J Martin, "Localizing nearby sound sources in a classroom: Binaural room impulse responses," *The Journal of the Acoustical Society of America*, vol. 117, pp. 3100, 2005.

[9] Boaz Rafaely and Amir Avni, "Interaural cross correlation in a sound field represented by spherical harmonics," *The Journal of the Acoustical Society of America*, vol. 127, pp. 823, 2010.

[10] Amir Avni, Jens Ahrens, Matthias Geier, Sascha Spors, Hagen Wierstorf, and Boaz Rafaely, "Spatial perception of sound fields recorded by spherical microphone arrays with varying spatial resolution," *The Journal of the Acoustical Society of America*, vol. 133, pp. 2711, 2013.

[11] Shahar Villeval, "Spatial soundfield processing for headphone reproduction," M.S. thesis, Ben-Gurion University of the Negev, To be published in 2014.

[12] George Brown Arfken, Hans-Jurgen Weber, and Lawrence Ruby, *Mathematical methods for physicists*, vol. 6, Academic press New York, 1985.

[13] James R Driscoll and Dennis M Healy, "Computing fourier transforms and convolutions on the 2-sphere," *Advances in applied mathematics*, vol. 15, no. 2, pp. 202–250, 1994.

[14] Boaz Rafaely, "Plane-wave decomposition of the sound field on a sphere by spherical convolution," *The Journal of the Acoustical Society of America*, vol. 116, pp. 2149, 2004.

[15] Earl G Williams, *Fourier acoustics: sound radiation and nearfield acoustical holography*, Access Online via Elsevier, 1999.

[16] Benjamin Bernschütz, Christoph Pörschmann, Sascha Spors, Stefan Weinzierl, and Begrenzung der Verstärkung, "Soft-limiting der modalen amplitudenverstärkung bei sphärischen mikrofonarrays im plane wave decomposition verfahren," *Proceedings of the 37. Deutsche Jahrestagung für Akustik (DAGA 2011)*, pp. 661–662, 2011.

[17] Munhum Park and Boaz Rafaely, "Sound-field analysis by plane-wave decomposition using spherical microphone array," *The Journal of the Acoustical Society of America*, vol. 118, pp. 3094, 2005.

[18] Benjamin Bernschütz, "A spherical far field hrir/hrtf compilation of the neumann ku 100," *Proceedings of the 40th Italian (AIA) Annual Conference on Acoustics and the 39th German Annual Conference on Acoustics (DAGA) Conference on Acoustics*, 2013.

[19] B. Bernschütz P. Stade and M. Ruhl, "Spatial audio impulse response compilation captured at the wdr broadcast studios," *Proc. TONMEISTERTAGUNG - VDT International Convention (TMT) Cologne, Germany*, 2012.

[20] Benjamin Bernschütz, "Spatial audio impulse response compilation captured at the wdr broadcast studios," `http://www.audiogroup.web.fh-koeln.de/wdr_irc.html`, 2013.

[21] Boaz Rafaely and Maor Kleider, "Spherical microphone array beam steering using wigner-d weighting," *Signal Processing Letters, IEEE*, vol. 15, pp. 417–420, 2008.

[22] Jens Ahrens, Matthias Geier, and Sascha Spors, "The soundscape renderer: A unified spatial audio reproduction framework for arbitrary rendering methods," in *Audio Engineering Society Convention 124*. Audio Engineering Society, 2008.

[23] Stéphanie Bertet, Jérôme Daniel, Etienne Parizet, and Olivier Warusfel, "Investigation on localisation accuracy for first and higher order ambisonics reproduced sound sources," *Acta Acustica united with Acustica*, vol. 99, no. 4, pp. 642–657, 2013.

[24] John C Middlebrooks and David M Green, "Sound localization by human listeners," *Annual review of psychology*, vol. 42, no. 1, pp. 135–159, 1991.

[25] Barbara G Shinn-Cunningham and Suraj Ram, "Identifying where you are in a room: Sensitivity to room acoustics," in *Proc. Int. Conf. Auditory Displays*, 2003, pp. 21–24.