1ST-ORDER MICROPHONE ARRAY SYSTEM FOR LARGE AREA SOUND FIELD RECORDING AND RECONSTRUCTION: DISCUSSION AND PRELIMINARY RESULTS

Federico Borra [2*], *Steven Krenn* [1], *Israel Dejene Gebru* [1], *Dejan Marković* [1]

[1] Facebook Reality Labs, 4420 Bayard Street, Suite 100, Pittsburgh, PA 15213
[2] Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy

ABSTRACT

The process of capturing, analyzing and predicting sound fields is finding novel areas of applications in AR/VR. One of the key processes in such applications is to estimate the sound field at locations that differ from the actual measurement points—i.e., the sound field reconstruction. However, it's a difficult spatial audio processing problem. Though theoretical solutions exist to reconstruct sound fields, they are practically infeasible due to hardware and computational requirements. This paper discusses the implementation of a system for large area sound field recording and reconstruction and proposes an improved sound field reconstruction algorithm. The proposed algorithm introduces a practical improvement in order to overcome implementation issues. In addition, we present a preliminary real-world results on an innovative but highly challenging application.

Index Terms— Sound Field Recording, Microphone Arrays,Virtual Microphone, Ambisonics, HRTF.

1. INTRODUCTION

In this paper we consider an audio capture system in which an array of microphones surrounds the acoustic scene of interest, with the goal of estimating the sound field at arbitrary points in space. Referred here as sound field reconstruction, it is a challenging problem that has recently gained increased attention [1–7]. Possible applications include: synthesis of binaural signals for navigation of captured scenes in VR; study of spatial properties of sound sources; generation of datasets for machine learning algorithms; etc. However, setting up such a capture system poses a number of research and engineering challenges and trade-offs. Particularly important is the spatial sampling, especially for 3D reconstruction methods that do not suppose any models regarding sound sources composing the acoustic scene [1, 7, 8]. Larger capture areas can easily require thousands of microphones, bringing challenges such has synchronization and accurate positioning. Furthermore, if the capture environment is not anechoic, the reverberation should be taken into account, especially considering that the hardware infrastructure is likely to add (early) reflections.

The purpose of this paper is to: offer an insight into the challenges that arise when building a spatial sound field recording system and propose practical solutions; introduce an improvement to a recently developed sound field reconstruction algorithm [7]; and, to the best of the authors knowledge, present one of the first real-world experimental evaluations of a sound field reconstruction system of this kind and scale, including a highly challenging application that uses the system for HRTF measurement. In particular, our system encompasses a relatively large capture area of radius $2.74\,\mathrm{m}$ and, in order to reduce the number of spatial samples, we employ higher-order microphones. Our implementation uses the 1st-order (ambisonic) microphones, while, for completeness, the theoretical derivation considers arbitrary microphone orders.

The paper is organized as follows. Section 2 presents a quick theoretical overview of sound field reconstruction and its formulation, and discusses the real-world implementation issues. Section 3 describes the practical improvement we propose. Our capture system is described in Section 4. And finally, Section 5 discusses the performance analysis of the proposed method and presents preliminary results on simulated and real-world data.

2. PROBLEM FORMULATION AND RELATED WORK

The problem of estimating a sound field at arbitrary points in space is formulated in the Spherical Harmonics (SH) domain [9]. In general, we can estimate the sound field without ambiguities in regions that do not contain any sources. For convenience we breakdown a particular setup into two spherical regions: a smaller region with radius R_s containing sources inside and a larger region with radius R_o having sources outside. Though not necessary concentric, the smaller sphere should be contained within the bigger one so that the region in between is source-free, with microphones placed inside this area. The sound field component generated inside R_s and outside R_o are called, respectively, exterior and interior sound fields, and are indicated here with p_E and p_I. Let $k = 2\pi f/c$ be the wave number, f the frequency, and c the speed of sound. At at any point $\boldsymbol{x} = (r, \theta, \phi)$, the two components can be represented as [9]:

$$\begin{aligned} p_E(\boldsymbol{x}, k) &= \sum_{n=0}^{N_E=\lceil keR_s/2\rceil} \sum_{m=-n}^{n} \beta_{nm}(k) h_n(kr) Y_{nm}(\theta, \phi)\,, \\ p_I(\boldsymbol{x}, k) &= \sum_{n=0}^{N_I=\lceil keR_0/2\rceil} \sum_{m=-n}^{n} \alpha_{nm}(k) j_n(kr) Y_{nm}(\theta, \phi)\,, \end{aligned} \tag{1}$$

where $Y_{nm}(\theta, \phi)$ represents the complex spherical harmonic of order n and degree m, $h_n(\cdot)$ and $j_n(\cdot)$ being, respectively, nth-order spherical Hankel and Bessel functions, $\beta_{nm}(k)$ and $\alpha_{nm}(k)$ denoting, respectively, the exterior and interior sound field coefficients, $\lceil\cdot\rceil$ is the ceiling operator and e is the Euler's number. The global sound field is given by the superposition of p_E and p_I. Therefore, if the global coefficients $\beta_{nm}(k)$ and $\alpha_{nm}(k)$ are known (up to the order N_E and N_I), the sound field can be easily reconstructed at arbitrary points in the region of interest using (1).

Given Q number of Vth-order microphones, each microphone captures local sound field coefficients $\alpha^{(q)}_{\nu\mu}(k)$, $q = 1, \ldots, Q$, up to the order V, i.e $\nu = 0, \ldots, V$. The SH addition theorem [10] links the local sound field coefficients with the global

*Work done during an internship at Facebook Reality Labs, Pittsburgh.

978-1-7281-1123-0/19/$31.00 ©2019 IEEE

coefficients. Let $\boldsymbol{d}(k) = [\alpha_{00}^{(1)}(k), \ldots, \alpha_{VV}^{(Q)}(k)]^T$ denote the vector containing $Q(V+1)^2$ observed local sound field coefficients, and $\boldsymbol{\beta}(k) = [\beta_{00}(k), \ldots, \beta_{N_E N_E}(k)]^T$ and $\boldsymbol{\alpha}(k) = [\alpha_{00}(k), \ldots, \alpha_{N_I N_I}(k)]^T$ the vectors containing, respectively, $(N_E+1)^2$ external and $(N_I+1)^2$ internal global sound field coefficients. The local and global coefficients are related by:

$$\boldsymbol{d}(k) = [\boldsymbol{T}_E(k), \boldsymbol{T}_I(k)] \begin{bmatrix} \boldsymbol{\beta}(k) \\ \boldsymbol{\alpha}(k) \end{bmatrix} \tag{2}$$

where $\boldsymbol{T}_E(k)$ and $\boldsymbol{T}_I(k)$ are, respectively, external and internal global to local translation matrices, defined as in [1]. The method proposed in [1] solves the inverse problem of (2) and obtains the estimate of the global coefficients $\hat{\boldsymbol{\beta}}(k)$ and $\hat{\boldsymbol{\alpha}}(k)$.

In practice, it can become quite challenging to solve (2) for large areas, from both computational and hardware perspective. In fact, the radius R of the given sphere determines the relationship between the number of microphones and the aliasing frequency for the corresponding sound field component [8], i.e. $f_{alias} = \frac{c[\sqrt{Q}(V+1)-1]}{\pi e R}$. In the considered setup, R_s is the sphere that encompass the active acoustic scene and it can be reduced by tracking objects/sources and adjusting the origin of the reference system [11]. On the other hand, R_0 is delimited by the microphone array and corresponds roughly to the size of the capture area. Given the region with $R_0 = 2.74\,\mathrm{m}$, theoretically we would need 13×10^3 1st-order microphones and 54×10^3 coefficients to estimate the interior field at $3.4\,\mathrm{kHz}$. On the other hand, if generated within a volume with $R_s = 0.5\,\mathrm{m}$ the exterior field requires around 450 1st-order microphones for the same frequency.

Since the scene of interest is confined within R_s and the interior field consists mostly of reverberation, a naive solution is to ignore the interior field and estimate only the exterior coefficients using:

$$\hat{\boldsymbol{\beta}}(k) = \boldsymbol{T}_E(k)^\dagger \boldsymbol{d}(k)\,, \tag{3}$$

with $(\cdot)^\dagger$ denoting the pseudo-inverse. This approach can give reasonable results in practice, but the estimate is inevitably affected by the interior sound field components present in the data $\boldsymbol{d}(k)$.

Recently in [7], we proposed a method that tackles the problem of reverberation leaking inside $\hat{\boldsymbol{\beta}}(k)$. For a given environment, the reverberant field can be considered as a function of the exterior field. We encode this function by measuring a set of room impulse responses (RIRs), effectively linking the exterior field coefficients $\boldsymbol{\beta}(k)$ to local coefficients $\boldsymbol{d}(k)$ while taking reverberation into account. However, a large number of RIR measurements is required for this method. While it is not necessary to repeat measurements as long as the environment does not change, it is necessary to densely sample the capture area in order to avoid aliasing problems. This can be time consuming and impractical. To adress this problem, here we propose to augment the available RIR measurements with points modeled as free-field impulse responses. By adding "dummy" measurement points we avoid the aliasing problblems in [7]. It will be shown in simulations that when no RIR measurements are available the proposed method performs comparably to the naive approach in (3). Intuitively, the results gradually improve as real RIR measurements are added. The following section presents the details of the proposed approach.

3. PROPOSED RECONSTRUCTION METHOD

We assume that RIR measurements are available; either measured or modeled as free-field impulse responses. Let $L = L_m + L_f$ be the total number of points, with L_m points from actual measurement and L_f points from free-field modeling. The RIR measurements can be considered as a virtual array of speakers whose environment response is known for each microphone. In the forward problem we assume that the exterior sound field coefficients $\boldsymbol{\beta}(k)$ are known. Hence, we can use the virtual loudspeakers to reproduce this sound field. Supposing omni-directional loudspeakers, the unknowns are the virtual loudspeaker modal weights: $\mathbf{w}(k) = [w_{00}^{(1)}(k), \ldots, w_{00}^{(L)}(k)]^T$ [12]. The relation between $\mathbf{w}(k)$ and $\boldsymbol{\beta}(k)$ can be written in a matrix form as:

$$\mathbf{J}(k)\mathbf{w}(k) = \boldsymbol{\beta}(k) \;\Rightarrow\; \mathbf{w}(k) = \mathbf{J}^\dagger(k)\boldsymbol{\beta}(k)\,, \tag{4}$$

where $\mathbf{J}(k)$ is the translation matrix [7, 12]. Because the environment response is given for all loudspeakers, we can relate the weights of the loudspeaker to the local sound field coefficient acquired by the qth microphone at order ν and mode μ as [13–15]

$$\tilde{\alpha}_{\nu\mu}^{(q)}(k) = \sum_{l'=1}^{L_m} w_{00}^{(l')}(k) \frac{\gamma_{\nu\mu}^{(l',q)}(k)}{ik} + \sum_{l''=1}^{L_f} w_{00}^{(l'')}(k) \eta_{\nu\mu}^{(l'',q)}(k)\,, \tag{5}$$

where $\gamma_{\nu\mu}^{(l,q)}(k)$ is the transfer function obtained from the RIR measurement, $\eta_{\nu\mu}^{(l'',q)}(k)$ is the free-field transfer function [9] and $\tilde{\alpha}_{\nu\mu}^{(q)}(k)$ is the real equivalent of complex local coefficient $\alpha_{\nu\mu}^{(q)}(k)$ [16]. Using (5) we can relate the signals acquired by the Vth-order microphones $\tilde{\boldsymbol{d}}(k) = [\tilde{\alpha}_{00}^{(1)}(k), \ldots, \tilde{\alpha}_{VV}^{(Q)}(k)]^T$ to the exterior sound field coefficients $\boldsymbol{\beta}(k)$ as:

$$\tilde{\boldsymbol{d}}(k) = \boldsymbol{\Gamma}(k)\mathbf{w}(k) = \boldsymbol{\Gamma}(k)\mathbf{J}^\dagger(k)\boldsymbol{\beta}(k), \tag{6}$$

where $\boldsymbol{\Gamma}(k) = [\boldsymbol{\Gamma}_m(k), \boldsymbol{\Gamma}_f(k)]$, with $\boldsymbol{\Gamma}_m(k) \in \mathbb{C}^{Q(V+1)^2 \times L_m}$ and $\boldsymbol{\Gamma}_f(k) \in \mathbb{C}^{Q(V+1)^2 \times L_f}$.

In the inverse problem, the exterior sound field coefficients are unknown and are thus estimated using:

$$\hat{\boldsymbol{\beta}}(k) = \left[\boldsymbol{\Gamma}(k)\mathbf{J}^\dagger(k)\right]^\dagger \tilde{\boldsymbol{d}}(k)\,. \tag{7}$$

If $\boldsymbol{\Gamma}(k)$ is full column rank and $\boldsymbol{J}^\dagger(k)$ is full row rank, the expression in (7) can be simplified as $\hat{\boldsymbol{\beta}}(k) = \mathbf{J}(k)\boldsymbol{\Gamma}^\dagger(k)\tilde{\boldsymbol{d}}(k)$. The matrix multiplication $\hat{\mathbf{w}}(k) = \boldsymbol{\Gamma}^\dagger(k)\tilde{\boldsymbol{d}}(k)$ can be interpreted as the widely studied multi-channel deconvolution problem [17–22]. Here we adopt the solution proposed in [17] where the deconvolution is accomplished in the frequency domain and the duration of the inverse filter is controlled by a regularization term.

4. AUDIO CAPTURE SYSTEM

Our capture system consists of 420 1st-order microphones spatially distributed on the surface of a sphere-shaped dome of radius $2.74\,\mathrm{m}$. The microphones are grouped into 20 distinct distributed recording nodes. Each node consists of 21 1st-order microphones connected to 11 daisy-chained RME MADI A-to-D converters. The A-to-D converters are in phase-lock with each other syncing to a master word-clock generator. The system samples up to $96\,\mathrm{kHz}$ and generates a 24-bit serial data for each channel. A master IRIG signal is distributed to each node as a control signal and used to time-align the audio recordings. A distributed computer system connected over a local area network is used to handle large amount of audio data; one computer per node. To control and manage audio recordings over the network, a custom recording software was developed.

978-1-7281-1123-0/19/$31.00 ©2019 IEEE

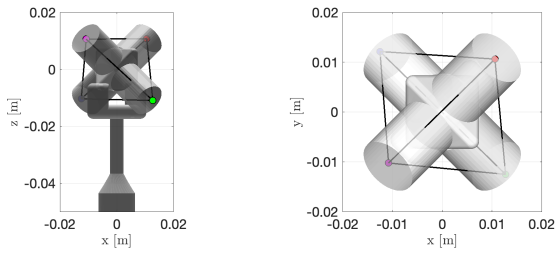

Figure 1: Localized capsule positions (colored dots) superimposed on the model of Core Sound TetraMic; during calibration capsules are considered independent with no geometry constraints.

Many multi-microphone audio processing algorithms require a precise knowledge of the microphones' locations and orientations. Our approach for acoustic geometry calibration is based on times of arrival (ToA) [23]. The calibration is performed offline with dedicated recordings. We gathered audio recordings by moving a B&K OmniSource loudspeaker mounted on a robot arm inside of the dome. Sensors monitor the temperature inside the capture area during the calibration procedure. The microphones are calibrated with a precision of $\sim 3\,\mathrm{mm}$. An example is shown in Figure 1.

5. EXPERIMENTAL VALIDATION

For validation we perform experiments on both simulated and real data. The purpose of using simulated data is to demonstrate the validity of the proposed method and to analyze the effect of the number and distribution of RIR measurements. Furthermore, to show the effectiveness of the proposed approach in separating reverberation from the exterior field and give an insight into the accuracy of the reconstruction, we performed experiments on real data.

5.1. Simulations

We simulated a shoe-box room of dimensions 6m×6m×6m with 150 1-st order microphones placed on a sphere of 1 m radius centered in the room. The microphone positions are generated through the Fibonacci sphere algorithm, and RIR generator from [24] is used to get the impulse responses for a source placed in the center. The reflection order and coefficient are set to 2 and 1, respectively.

We compared the proposed approach against the naive solution stated in (3). The performance is evaluated by comparing the estimated exterior sound field pressure with the ground truth using the Normalized-Mean-Square Error NMSE, defined as follows:

$$\mathrm{NMSE}(k) = 10\log_{10}\frac{\sum_{p=1}^{P}\left(p_E(\boldsymbol{x}_p,k)-\hat{p}_E(\boldsymbol{x}_p,k)\right)^2}{\sum_{p=1}^{P}\left(p_E(\boldsymbol{x}_p,k)\right)^2}, \quad (8)$$

where $\boldsymbol{x}_p$ denotes the position of the pth point, and $p_E(\boldsymbol{x}_p,k)$ is the the ground truth exterior sound field, $\hat{p}_E(\boldsymbol{x}_p,k)$ is the estimated exterior sound field. The experiments are performed to analyze the effect of the number and distribution of RIR measurements. We simulated a total of $L = L_m + L_f = 50$ virtual speakers for which RIRs are either known or assumed as free-field. The positions of the virtual speakers are sampled from a multivariate Gaussian with mean located at the room center, and covariance matrix of $\mathrm{diag}(\sigma^2)$.

To validate the effect of RIR availability on the proposed method, we set $\sigma^2 = 10^{-3}$ and varied the ratio L_m/L. The results are reported in Figure 2. As expected, the best performance is achieved when RIRs are known for all points. The error increases with frequency. If only half of the available points have known RIRs the error remains relatively low. When no RIRs are available, the performance of the proposed method is similar to that of (3).

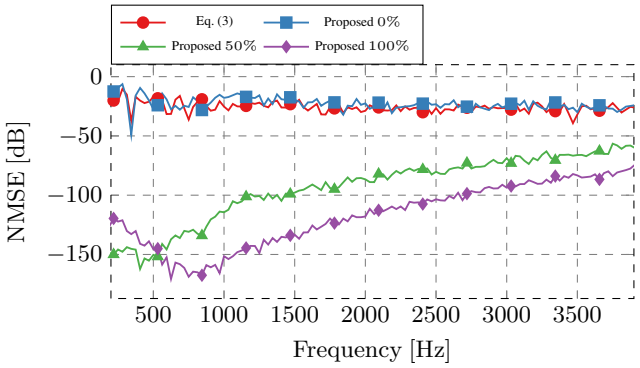

Figure 2: NMSE for the solution in (3) and the proposed method for different percentages of available RIRs (0%, 50%, 100%).

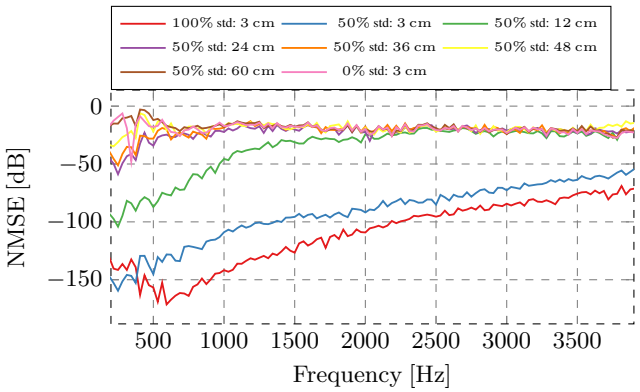

Figure 3: NMSE for the proposed method for different values of the standard deviation σ.

To validate the effect of RIR distribution on the proposed method, we set $L_m = L_f = 25$ and varied the standard deviation σ of points with known RIRs (the distribution of points assumed free-field is unchanged). The performance of the proposed method for different values of the standard deviation is shown in Figure 3. For comparison we also show the two extreme cases from the previous simulation, i.e. all and no RIRs available. As the spread of RIR points around the source region increases, the performance decreases. Nonetheless, the availability of RIRs always remains advantageous for lower frequencies.

5.2. Real-world experiments

We present two preliminary real-world experiments as a benchmark of our capture system described in Section 4 and to show the feasibility of our proposed reconstructing method. For both experiments, we measured RIRs using an omni-directional speaker. We collected $L_m = 37$ RIR measurements around the center. During the measurement session, omni-directional reference microphones are placed at 21 randomly chosen positions within the capture area.

978-1-7281-1123-0/19/$31.00 ©2019 IEEE

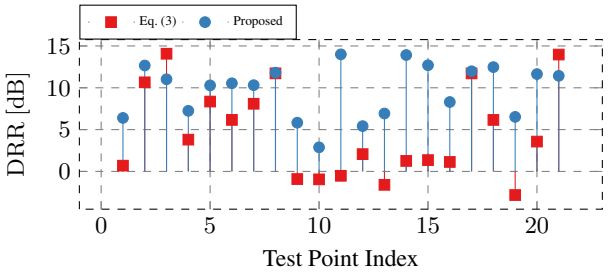

Figure 4: DRR at each test point.

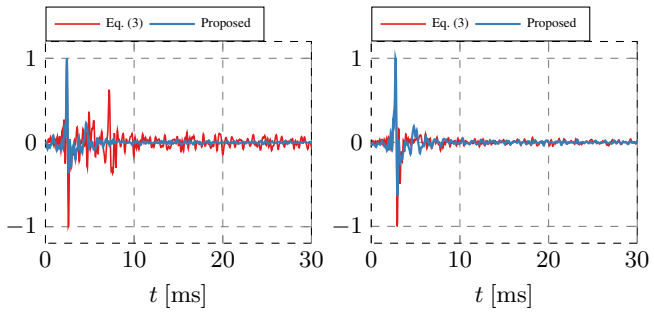

Figure 5: Example of two impulse responses $v_p(t)$ computed at the test point using the approach in (3) and the proposed method.

Sound field separation In this experiment a speaker is placed roughly in the center of the capture area. We used $Q = 67$ microphones that are uniformly spaced around the capture area with heights ranging between $0.4\,\mathrm{m}$ and $2.4\,\mathrm{m}$. In order to evaluate the amount of reverberation leaking into the exterior field estimates, we used the speaker signal to extract the impulse responses and then compute the Direct-to-Reverberant Ratio (DRR)

$$\mathrm{DRR} = 10\log_{10}\frac{\int_{T_0-C}^{T_0+C} v_p(\tau)^2 d\tau}{\int_{T_0+C}^{+\infty} v_p(\tau)^2 d\tau}, \tag{9}$$

where $v_p(t)$ is the RIR between the speaker and the sound field estimate at the reference position $\boldsymbol{x}_p$, t is time, T_0 the time of the direct impulse, and $C = 2.5\,\mathrm{ms}$ [25]. The results obtained at each test point are reported in Figure 4. To illustrate the differences between estimates obtained using (3) and the proposed method, the impulse responses for the test points with highest and lowest difference in DRR between the two approaches are shown in Figure 5. In the case of highest DRR difference we can observe a strong early reflection (probably due to the structure supporting the capture hardware) that gets removed using the proposed approach. Overall, the preliminary results show the effectiveness of the proposed approach in reducing the effects of reverberation.

Reciprocal HRTF measurement A possible application of the reconstruction system is the study of spatial properties of sound sources. By placing the sound source inside the ear, the Head Related Transfer Function (HRTF) can be measured based on reciprocity principle [26]. By reconstructing the exterior sound field, the system essentially estimates the SH representation of the HRTF [27–29]. We use $Q = 132$ microphones more densely concentrated around the azimuth plane, with heights between $1.6\,\mathrm{m}$ and $1.86\,\mathrm{m}$.

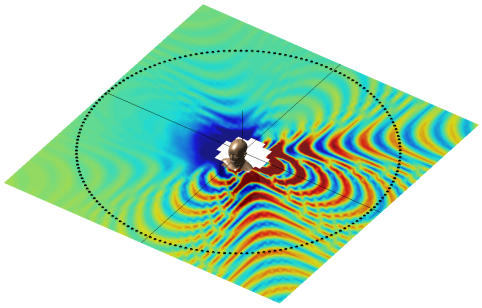

Figure 6: Reconstruction snapshot showing the sound field generated by the in-ear speaker placed in the left ear of the KEMAR manikin and producing a sine sweep. Black dots indicate the virtual microphones used for HRTF estimation.

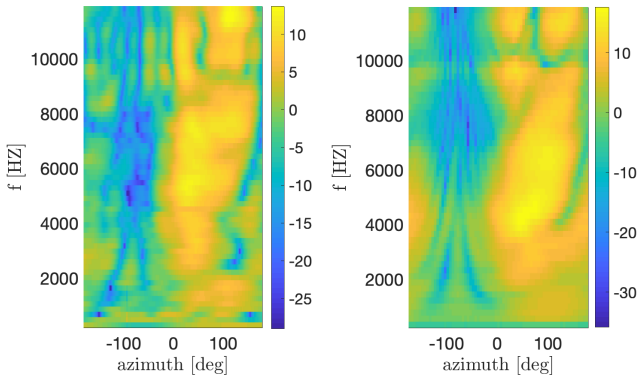

Figure 7: HRTF of the KEMAR manikin obtained: (a) using our experimental measurement procedure; (b) from the CIPIC dataset.

A KEMAR manikin with an in-ear speaker is placed in the center of the capture area. Figure 6 shows a snapshot of the reconstructed sound field in the azimuth plane and depicts the positions of the virtual microphones used for the HRTF estimation. Finally, Figure 7a shows the preliminary HRTF measurement result for the azimuth plane. For visual comparison, Figure 7b shows the KEMAR HRTF for the same frequency range obtained from the CIPIC dataset [30].

6. CONCLUSIONS

In this work we discussed the sound field reconstruction problem and presented a method that uses measured and free-field impulse responses to avoid frequency-aliasing problem. Moreover, we presented a first order microphone array system and show some preliminary results. We showed that the audio capture system together with the proposed reconstruction method can potentially be used to estimate HRTFs. In the near future, we plan to perform experiments to measure the HRTFs over the whole sphere and evaluate performance and perceptual fidelity of the measurements.

ACKNOWLEDGMENT

The authors would like to thank Nicholas Pourazima for designing and prototyping the in-ear speaker used for HRTF measurements.

978-1-7281-1123-0/19/$31.00 ©2019 IEEE

7. REFERENCES

[1] A. Fahim, P. N. Samarasinghe, and T. D. Abhayapala, "Sound field separation in a mixed acoustic environment using a sparse array of higher order spherical microphones," in *IEEE Hands-free Speech Communications and Microphone Arrays (HSCMA)*, 2017.

[2] H. Zuo, P. N. Samarasinghe, and T. D. Abhayapala, "Exterior - interior 3D sound field separation using a planar array of differential microphones," in *Proc. IEEE Int. Workshop on Acoustic Signal Enhancement (IWAENC)*, 2018.

[3] M. Pezzoli, F. Borra, F. Antonacci, A. Sarti, and S. Tubaro, "Estimation of the sound field at arbitrary positions in distributed microphone networks based on distributed ray space transform," in *IEEE Int. Conf. on Acoustics, Speech and Signal Process. (ICASSP)*, 2018.

[4] M. Pezzoli, F. Borra, F. Antonacci, A. Sarti, and S. Tubaro, "Reconstruction of the virtual microphone signal based on the distributed ray space transform," in *Proc. European Signal Process. Conf. (EUSIPCO)*, 2018.

[5] Y. Takida, S. Koyama, and H. Saruwataril, "Exterior and interior sound field separation using convex optimization: Comparison of signal models," in *European Signal Process. Conf. (EUSIPCO)*, 2018.

[6] Y. Takida, S. Koyama, N. Ueno, and H. Saruwatari, "Robust gridless sound field decomposition based on structured reciprocity gap functional in spherical harmonic domain," in *IEEE Int. Conf. on Acoustics, Speech and Signal Process. (ICASSP)*, 2019.

[7] F. Borra, I. D. Gebru, and D. Markovic, "Soundfield reconstruction in reverberant environments using higher-order microphones and impulse response measurements," in *IEEE Int. Conf. on Acoustics, Speech and Signal Process. (ICASSP)*, 2019.

[8] P. N. Samarasinghe and T. D. Abhayapala, "3D spatial soundfield recording over large regions," in *Proc. Int. Workshop on Acoustic Signal Enhancement (IWAENC)*, 2012.

[9] E. G. Williams, *Fourier Acoustics*, Academic Press, 1999.

[10] P. A. Martin, *Multiple Scattering: Interaction of Time-Harmonic Waves with N Obstacles*, Cambridge University Press, 2006.

[11] B. Rafaely, "Spatial alignment of acoustic sources based on spherical harmonics radiation analysis," in *IEEE Int. Symposium on Communications, Control and Signal Process. (ISCCSP)*, 2010.

[12] P. N. Samarasinghe, M. A. Poletti, S. M. A. Salehin, T. D. Abhayapala, and F. M. Fazi, "3D soundfield reproduction using higher order loudspeakers," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Process. (ICASSP)*, 2013.

[13] P. N. Samarasinghe, T. D. Abhayapala, and M. A. Poletti, "Synthesis of room transfer function over a region of space by multiple measurements using a higher-order directional microphone," in *IEEE China Summit Int. Conf. on Signal and Information Process. (ChinaSIP)*, 2014.

[14] M. Poletti P. N. Samarasinghe, T. Abhayapala and T. Betlehem, "An efficient parameterization of the room transfer function," *IEEE/ACM Trans. Audio, Speech, Language Process.*, 2015.

[15] P. N. Samarasinghe and T. D. Abhayapala, "Room transfer function measurement from a directional loudspeaker," in *IEEE Int. Workshop on Acoustic Signal Enhancement (IWAENC)*, 2016.

[16] M. Poletti, "Unified description of ambisonics using real and complex spherical harmonics," in *Proc. Ambisonics Symposium*, 2009.

[17] Ole Kirkeby, Philip A Nelson, Hareo Hamada, and Felipe Orduna-Bustamante, "Fast deconvolution of multichannel systems using regularization," *IEEE Trans. on Speech and Audio Process.*, 1998.

[18] O. Kirkeby, P. Rubak, and A. Farina, "Analysis of ill-conditioning of multi-channel deconvolution problems," in *Proc. IEEE Workshop on Applications of Signal Process. to Audio and Acoustics. (WASPAA)*, 1999.

[19] A. Gonzalez and J. J. Lopez, "Fast transversal filters for deconvolution in multichannel sound reproduction," *IEEE Trans. on Speech and Audio Process.*, 2001.

[20] J. J. Lopez, A. Gonzalez, and L. Fuster, "Room compensation in wave field synthesis by means of multichannel inversion," in *IEEE Workshop on Applications of Signal Process. to Audio and Acoustics (WASPAA)*, 2005.

[21] Etienne Corteel, "Equalization in an extended area using multichannel inversion and wave field synthesis," *J. Audio Eng. Soc*, 2006.

[22] M. Kolundzija, C. Faller, and M. Vetterli, "Reproducing sound fields using mimo acoustic channel inversion," *J. Audio Eng. Soc*, 2011.

[23] Axel Plinge, Florian Jacob, Reinhold Haeb-Umbach, and Gernot A Fink, "Acoustic microphone geometry calibration: An overview and experimental evaluation of state-of-the-art algorithms," *IEEE Signal Processing Magazine*, vol. 33, no. 4, pp. 14–29, 2016.

[24] E. A. P. Habets, "RIR generator, version 2.1.20141124," 2014.

[25] Pavel Zahorik, "Direct-to-reverberant energy ratio sensitivity," *The J. Acoust. Soc. Am.*, 2002.

[26] D. N. Zotkin, R. Duraiswami, E. Grassi, and N. A. Gumerov, "Fast head-related transfer function measurement via reciprocity," *The Journal of the Acoustical Society of America*, 2006.

[27] W. Zhang, T. D. Abhayapala, R. A. Kennedy, and R. Duraiswami, "Modal expansion of HRTFs: Continuous representation in frequency-range-angle," in *IEEE Int. Conf. on Acoustics, Speech and Signal Process (ICASSP)*, 2009.

[28] M. Pollow, KV Nguyen, O. Warusfel, T. Carpentier, M. Müller-Trapet, M. Vorländer, and M. Noisternig, "Calculation of head-related transfer functions for arbitrary field points using spherical harmonics decomposition," *Acta acustica united with Acustica*, 2012.

[29] R. A. Kennedy, W. Zhang, and T. D. Abhayapala, "Comparison of spherical harmonics based 3D-HRTF functional models," in *Int. Conf. on Signal Process. and Communication Systems (ICSPCS)*, 2013.

[30] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF database," in *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No.01TH8575)*, Oct 2001, pp. 99–102.

978-1-7281-1123-0/19/$31.00 ©2019 IEEE