

# EXTRACTION OF EXTERIOR FIELD FROM A MIXED SOUND FIELD FOR 2D HEIGHT-INVARIANT SOUND PROPAGATION

Abdullah Fahim, Prasanga Samarasinghe, Thushara D. Abhayapala

Research School of Engineering, CECS, The Australian National University, Canberra, Australia  
Email: {abdullah.fahim, prasanga.samarasinghe, thushara.abhayapala}@anu.edu.au

## ABSTRACT

Any sound field caused by one or more sound sources takes the form of an interior or an exterior sound field based on the source location. The representation of the interior or exterior sound field in terms of harmonics decomposition has been extensively studied in the literature. However, the interior and exterior sound fields often co-exist in nature to form a mixed wave field. The extraction of the exterior sound field from such a mixed field can be very useful in many acoustic processes, which is yet to be explored extensively. In this paper, we discuss a method to extract the exterior field from a mixed sound field for 2D height-invariant sound propagation. Such an extraction method can be employed to record a sound field in a noisy or interfered room or to perform dereverberation in a reverberant room where the sound fields due to the source signal and its reflections superimpose each other. We demonstrate two practical uses of the proposed method in the forms of (i) an exterior sound field recording in a mixed wave field and (ii) speech dereverberation in a simulated reverberant room.

**Index Terms**— Cylindrical harmonics, microphone array, sound field recording, speech dereverberation

## 1. INTRODUCTION

Consider a circular region  $\eta$  where the existing sound field is caused by one or more sources lying outside the circle. Such a sound field is defined as an interior sound field [1]. On the other hand, if all the sources lie within  $\eta$ , the sound field at any point outside  $\eta$  is known as an exterior sound field. The representation of an interior or an exterior sound field has been extensively studied in literature [2, 3, 4, 5]. However, the problem of isolating exterior (or interior) field in a mixed sound field environment and its practical application remain largely unexplored. In this work, we describe a practical method for extracting the exterior sound field from a mixed wave field by means of cylindrical harmonics decomposition for 2D height-invariant sound propagation. We also demonstrate the application of proposed algorithm by means of sound field recording and speech dereverberation examples.

To represent a sound field in modal domain, it is required to obtain all its harmonic components. However, the standard omni-directional microphones can record only zeroth order harmonic. Hence, we need to spatially sample [6] the sound field using a microphone array to record higher order harmonics. Samarasinghe et al. used an array of higher order microphones to record spatial sound field [7]. Chen et al. designed 2D planar microphone arrays to capture 3D spatial sound fields in [8]. Ideally, the order of a sound field extends to infinity, however, the high pass characteristic of Bessel functions allows us to represent a sound field with limited number of harmonics without major deviation [9]. The error analysis for such truncation is presented in [10]. All these methods help us to record an exterior or an interior sound field separately. However, in real applications, the sound field frequently contains both interior and exterior field coefficients, which makes them inseparable under normal circumstances.

In this study, we intend to extract the exterior field from a mixed sound field by means of cylindrical harmonics decomposition. Using two concentric circular microphone arrays, we capture the combined sound field and thereafter, estimate the exterior sound field coefficients using a dual surface approach of scattering holography [1]. We use the sound field coefficients to record the desired exterior sound field. We also demonstrate a simulated method of speech dereverberation to exhibit the practical application of the proposed algorithm.

## 2. PROBLEM FORMULATION

We consider a predefined global origin  $O$  while expressing any position vector or spatial sound field. In cylindrical harmonics decomposition, the sound field at any point  $\mathbf{x} = (r, \phi)$  due to a sound source at  $\mathbf{x}_s = (r_s, \phi_s)$  is expressed by [4]

$$P(\mathbf{x}, k) = \begin{cases} \sum_{n=-\infty}^{\infty} \alpha_n(k) J_n(kr) e^{in\phi}, & r_s > r \quad (1) \\ \sum_{n=-\infty}^{\infty} \beta_n(k) H_n(kr) e^{in\phi}, & r_s < r \quad (2) \end{cases}$$

where  $k = 2\pi f/c$  is the wave number,  $f$  is the frequency,  $c$  is the speed of sound propagation,  $J_n(\cdot)$  and  $H_n(\cdot)$  are the  $n^{\text{th}}$

order Bessel and Hankel functions of first kind, respectively,  $i = \sqrt{-1}$ , and  $\alpha_n(k)$  and  $\beta_n(k)$  are the  $n^{\text{th}}$  order sound field coefficients, respectively. The sound field of (1) is known as an interior sound field whereas the sound field with in (2) is described as an exterior sound field. For a unit amplitude sound source at  $\mathbf{x}_s$ ,  $\alpha_n(k)$  and  $\beta_n(k)$  are defined as [4]

$$\alpha_n(k) = H_n(kr_s)e^{-in\phi_s} \quad (3)$$

$$\beta_n(k) = J_n(kr_s)e^{-in\phi_s}. \quad (4)$$

Let us assume, at point  $\mathbf{x}$ ,  $L$  distinct sound sources create an interior field while  $M$  distinct sound sources result an exterior field. The total sound pressure at  $\mathbf{x}$  is

$$P(\mathbf{x}, k) = \sum_{m=1}^M P_m(\mathbf{x}, k) + \sum_{\ell=1}^L P_\ell(\mathbf{x}, k) \quad (5)$$

where  $P_m(\mathbf{x}, k)$  and  $P_\ell(\mathbf{x}, k)$  are the sound pressures due to  $m^{\text{th}}$  and  $\ell^{\text{th}}$  sources, respectively. We get the modal representation of (5) using (1), (2) and (5) as

$$\begin{aligned} P(\mathbf{x}, k) &= \underbrace{\left( \sum_{m=1}^M \beta_n^m(k) S_m(k) \right)}_{\tilde{\beta}_n(k)} H_n(kr) e^{in\phi} + \\ &\quad \underbrace{\left( \sum_{\ell=1}^L \alpha_n^\ell(k) S_\ell(k) \right)}_{\tilde{\alpha}_n(k)} J_n(kr) e^{in\phi} \quad (6) \\ &= \sum_{n=-\infty}^{\infty} [\tilde{\beta}_n(k) H_n(kr) + \tilde{\alpha}_n(k) J_n(kr)] e^{in\phi} \quad (7) \end{aligned}$$

where  $\alpha_n^\ell(k)$  and  $\beta_n^m(k)$  are the interior field coefficients due to  $\ell^{\text{th}}$  source and exterior field coefficients due to  $m^{\text{th}}$  source, respectively,  $\tilde{\alpha}_n(k)$  and  $\tilde{\beta}_n(k)$  characterize all the interior and exterior fields, respectively, and  $S_\ell(k)$  and  $S_m(k)$  are the  $\ell^{\text{th}}$  and  $m^{\text{th}}$  source signal, respectively. In this work, we intend to estimate  $\tilde{\beta}_n(k)$  from the observed signal  $P(\mathbf{x}, k)$  in a 2D height-invariant cylindrical sound field to extract exterior fields.

### 3. EXTRACTING EXTERIOR FIELD COEFFICIENTS

Let us assume that the freedom of movement of the speaker is a circle with radius  $R$ . We use two circular microphone arrays with radius  $r_1$  and  $r_2$  such that they are concentric with the source circle of radius  $R$  and  $r_2 > r_1 > R$  (Fig. 1). Using (7), we express the sound pressure at any two co-angular points on the microphone arrays by

$$P_1(\phi, k) = \sum_{n=-\infty}^{\infty} (\tilde{\beta}_n(k) H_n(kr_1) + \tilde{\alpha}_n(k) J_n(kr_1)) e^{in\phi} \quad (8)$$

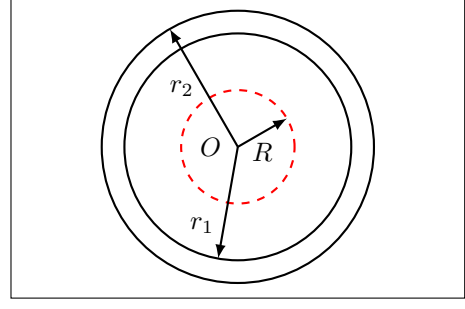


Fig. 1. The room geometry

$$P_2(\phi, k) = \sum_{n=-\infty}^{\infty} (\tilde{\beta}_n(k) H_n(kr_2) + \tilde{\alpha}_n(k) J_n(kr_2)) e^{in\phi} \quad (9)$$

where  $\phi$  is the azimuth angle of the measurement points, and  $P_1(\phi, k)$  and  $P_2(\phi, k)$  are the sound pressures on array 1 and array 2, respectively. Using the orthogonal property of the exponential functions on (8) and (9), we can derive

$$\tilde{\beta}_n(k) H_n(kr_1) + \tilde{\alpha}_n(k) J_n(kr_1) = P_{1n}(k) \quad (10)$$

$$\tilde{\beta}_n(k) H_n(kr_2) + \tilde{\alpha}_n(k) J_n(kr_2) = P_{2n}(k) \quad (11)$$

where

$$P_{1n}(k) = \frac{1}{2\pi} \int P_1(\phi, k) e^{-in\phi} d\phi \quad (12)$$

$$P_{2n}(k) = \frac{1}{2\pi} \int P_2(\phi, k) e^{-in\phi} d\phi. \quad (13)$$

The realization of (12) and (13) requires continuous microphone arrays which is impractical. However, we can estimate  $P_{1n}(k)$  and  $P_{2n}(k)$  with spatial sampling [2] by

$$\hat{P}_{1n}(k) = \frac{1}{2\pi} \sum_{q=1}^{Q_1} A_{1q} P_1(\phi_q, k) e^{-in\phi_q} \quad (14)$$

$$\hat{P}_{2n}(k) = \frac{1}{2\pi} \sum_{q=1}^{Q_2} A_{2q} P_2(\phi_q, k) e^{-in\phi_q} \quad (15)$$

where  $Q_1 = \lceil 2ker_1 + 1 \rceil$  and  $Q_2 = \lceil 2ker_2 + 1 \rceil$  [4] are the minimum number microphones required to avoid spatial aliasing [11] for array 1 and array 2, respectively,  $\lceil \cdot \rceil$  denotes ceiling operation,  $A_{1q}$  and  $A_{2q}$  are corresponding weights to ensure the orthogonal property of exponential functions,  $\phi_q$  denotes the angular location of  $q^{\text{th}}$  microphone. For simplicity, we use a uniform-angle sampling scheme with  $A_{1q} = 2\pi/Q_1$  and  $A_{2q} = 2\pi/Q_2$ . Hence, we solve (10) and (11) for  $\tilde{\beta}_n(k)$  by

$$\hat{\beta}_n(k) = \frac{J_n(kr_1) \hat{P}_{2n}(k) - J_n(kr_2) \hat{P}_{1n}(k)}{J_n(kr_1) H_n(kr_2) - J_n(kr_2) H_n(kr_1)}. \quad (16)$$

With the estimation of  $\hat{\beta}_n(k)$  using (16), it is possible to extract the exterior sound field from a complex wave field, allowing us to use the method in practical applications like spatial sound field recording and speech dereverberation.

## 4. PRACTICAL APPLICATIONS

### 4.1. Sound field recording

Assuming none of the undesired source lies in the close proximity of the intended source region, we model the mixed sound field like (6), where the intended sources contribute to an exterior field. Therefore, we can record the desired sound field using (2) and (16) by

$$\hat{P}_s(\mathbf{x}, k) = \sum_{n=-N_S}^{N_S} \hat{\beta}_n(k) H_n(kr) e^{in\phi} \quad (17)$$

where  $N_S = ker_s$  is the exterior field order [5] and  $r_s$  is assumed to be known (e.g., either  $r_s$  is fixed or estimated through any suitable localization algorithm).

### 4.2. Speech dereverberation

Due to the fact that, all the reflected waves in a reverberant room form an interior field irrespective of the source and microphone positions, we model the room using (6) where  $M = 1$ ,  $S_1(k) = S(k)$  represents the clean speech signal,  $L$  is total number of reflections and  $S_\ell(k) = S(k)$ ,  $\forall \ell$ . Hence,  $\hat{P}_s(\mathbf{x}, k)$  of (17) in a reverberant room represents the spatial filtered version of  $S(k)$ , e.g., the sound field at  $\mathbf{x}$  due to  $S(k)$ . Furthermore, if we localize the source angular position,  $\phi_s$ , we can use (4), (6) and (17) to estimate clean speech  $S(k)$  by

$$\hat{S}(k) = \frac{\hat{P}_s(\mathbf{x}, k)}{\sum_{n=-N_S}^{N_S} \beta_n(k) H_n(kr) e^{in\phi}}. \quad (18)$$

## 5. SIMULATION RESULTS

Unless specified differently, the following parameters are used in the simulations:  $r_1 = 1$  m,  $r_2 = 1.5$  m,  $r_s = 0.5$  m,  $\phi_s = \pi/3$  and  $c = 343$  m/s. Also, the center of the room was taken as the global origin.

### 5.1. Microphone selection

Theoretically, for a perfect reconstruction, we would require  $Q_1$  and  $Q_2$  microphones in array 1 and array 2, respectively. For simplicity, we used a fix number of microphones,  $Q_{\text{opt}}$  for both the arrays. We defined the coefficient estimation error as

$$C_{\text{err}} = \frac{\sum_{\forall n} |\beta_n(k) - \hat{\beta}_n(k)|}{\sum_{\forall n} |\beta_n(k)|} \quad (19)$$

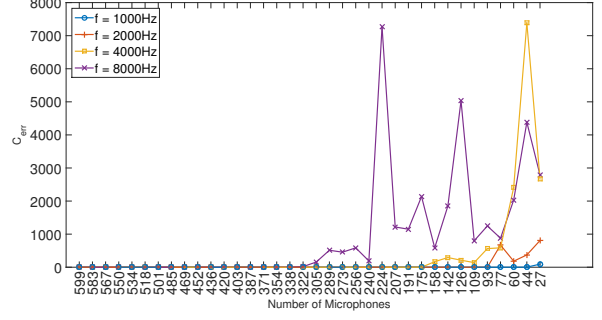


Fig. 2. Coefficient estimation error

where  $|\cdot|$  denotes absolute value. The undesired sound fields were modeled with 20 point sources and plane waves by [12]

$$P(\mathbf{x}, k) = \begin{cases} A_m \frac{i}{4} H_0(k \|\mathbf{x} - \mathbf{x}_s\|) & \text{for point source,} \\ A_m e^{-ik\hat{\mathbf{x}}_i \cdot \mathbf{x}} & \text{for plane wave} \end{cases} \quad (20)$$

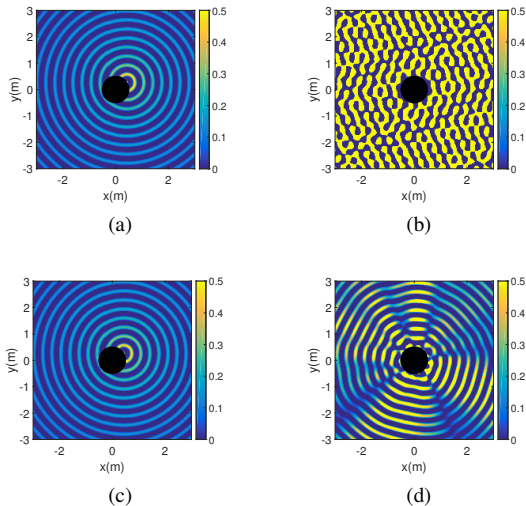
where  $\|\cdot\|$  denotes geometric distance, the signal magnitude  $A_m$ , source location  $\mathbf{x}_s$  outside  $r_2$  and the plane wave incident angle  $\hat{\mathbf{x}}_i$  were chosen randomly.

Fig. 2 plots  $C_{\text{err}}$  against  $Q_{\text{opt}}$  for different frequencies. It is evident from the figure that the  $C_{\text{err}}$  is zero when  $Q_{\text{opt}} \geq Q_2$ , as expected. It also exhibits an interesting trend that, at a fixed  $C_{\text{err}}^{\text{max}}$  (maximum allowable error), the difference between  $Q_2$  and  $Q_{\text{opt}}$  increases with frequency. Hence, to record a sound field at higher frequency, choosing a value of  $Q_{\text{opt}}$  significantly lower than  $Q_2$  should suffice. In this work, the selection of  $Q_{\text{opt}}$  was done in an empirical manner based on the highest frequency component of the desired signal.

### 5.2. Exterior field recording

In this simulation, we extracted the desired sound field from the complex wave field of previous section for  $f = 1$  kHz and  $Q_{\text{opt}} = 70$ , whereas  $Q_2$  was 77. Fig. (3a) and (3b) show the desired and the superimposed sound field, respectively. We ran simulations under two different conditions. In the first case, we assumed no thermal noise at microphone positions and achieved a perfect extraction of the desired sound field (Fig. 3c). However, in the real world applications, thermal noise exists at the pressure sensors of the microphones due to the thermal vibrations in the conducting material. Hence, we added a thermal noise at each microphone signal while maintaining a combined signal to noise ratio (SNR) of 30 dB at each microphone array. The SNR at each microphone array was calculated by

$$\text{SNR}_i = \frac{\sum_{q=1}^{Q_i} |P_i(\phi_q, k)|^2}{\sigma^2 Q_i}, \text{ for } i = 1, 2 \quad (21)$$



**Fig. 3.** Sound field recording with  $f = 1$  kHz,  $Q_{\text{opt}} = 70$ ,  $r_1 = 1$  m,  $r_2 = 1.5$  m,  $r_s = 0.5$  m and  $\phi_s = \pi/3$ . The black circle denotes the source area. (a) Original sound field, (b) combined sound field with 20 undesired sources outside  $r_2$ , (c) reconstructed sound field without thermal noise and (d) reconstructed sound field with 30 dB thermal noise.

where  $\sigma^2$  is the noise power. The estimated sound field in the presence of the thermal noise is shown in Fig. (3d), which also exhibits a good extraction of the desired wave field. Please note that, the proposed method inherently cancels out any noise signal originating outside  $r_2$ .

### 5.3. Application in speech dereverberation

We measured dereverberation in two rooms with (8 m  $\times$  8 m) and (5 m  $\times$  5 m) dimensions. We used image source method [13] to simulate reverberant conditions in those rooms with  $(r_1, r_2, r_s)$  being (1 m, 1.5 m, 0.5 m) and (2.5 m, 3 m, 1 m), respectively. Each image source was modeled as a 2D point source in space. Clean speech data was taken from WSJ-CAM0 corpus [14] and re-sampled at 8 kHz to reduce computational cost. A 256-point discrete Fourier transform was used with a 20 ms window and 50% overlap. We evaluated the performance of dereverberation in terms of frequency-weighted segmental SNR (FWSegSNR) and cepstral distance (CD) [15]. For reference, a higher value of FWSegSNR and a lower value of CD indicate less reverberated signal.

The measured value of  $Q_2$  was 599 and 301 in room 1 and room 2, respectively, for a Nyquist frequency of 4 kHz. Based on the empirical data, we used  $Q_{\text{opt}} = 325$  for room 1 and  $Q_{\text{opt}} = 175$  for room 2. The simulation results are listed in Table 1 for different reflection coefficients. We measured the performance for the recorded sound field  $\hat{P}$  at  $(r_1, \pi/3)$  as well as for the estimated clean speech  $\hat{S}$  in the presence of thermal noise. We recorded up to 1 dB and 1.5 dB im-

**Table 1.** Dereverberation performance.  $\zeta$  denotes the reflection coefficients of the walls. Rev. stands for the reverberated speech,  $\hat{P}$  is the recorded sound field at  $(r_1, \pi/3)$  and  $\hat{S}$  is the estimated clean speech in the presence of thermal noise.

	$\zeta$	FWSegSNR (dB)			CD (dB)		
		Rev.	$\hat{P}$	$\hat{S}$	Rev.	$\hat{P}$	$\hat{S}$
Room 1	0.3	5.1	6.4	13.3	2.8	1.3	1.5
Room 1	0.5	5.1	6.5	12.2	2.7	1.4	1.5
Room 1	0.8	5.9	6.6	9.2	2.3	1.6	1.7
Room 2	0.3	6.4	6.6	12.7	1.7	1.3	1.3
Room 2	0.5	6.8	7.2	9.5	1.7	1.3	1.4
Room 2	0.8	7.6	8.0	6.5	1.8	1.6	1.6

provements in terms of FWSegSNR and CD for  $\hat{P}$ . However, a significant improvement of FWSegSNR (up to 8 dB) was observed for  $\hat{S}$ , except for the smaller room 2 at high reverberant condition. The better result for  $\hat{S}$  came at the cost of an extra requirement of known source angular position whereas  $\hat{P}$  could be estimated only with the knowledge of source radius. The latter case is particularly useful when the speakers are located at fixed radial positions (e.g., in a meeting room).

It is important to note that, the results in Table 1 was based on the simulations that used image source method where each reflected wave was modeled as a point source in a perfectly shaped rectangular room. In practical implementation, the proposed method is unlikely to achieve the same level of improvement due to the deviations from the assumptions made in image source method while simulating reverberation. The impact of the algorithm on the experimental data is pending for investigation. Also, in the case of a 3D sound field, the proposed method requires minimum  $\lceil (ker_2 + 1)^2 / (A + 1)^2 \rceil$  of  $A^{\text{th}}$  order microphones to sample the sound field [5]. The number can be reduced if we consider the reflections from XY plane only (e.g., no reflection from ceiling or floor) [16].

## 6. CONCLUSION

In this work, we demonstrated a practical method of extracting exterior sound field in a reverberant room for the case of height-invariant cylindrical sound field. We successfully recorded the desired exterior sound field using optimum number of microphones. We used the method in a simulated reverberant room to perform speech dereverberation and obtained significant improvement in terms of FWSegSNR and CD. The scope for future work includes the evaluation of the algorithm in practical environment where certain assumptions of the image source method used in the simulation may not hold true.

## 7. REFERENCES

- [1] E. G. Williams, *Fourier acoustics: sound radiation and nearfield acoustical holography*, Academic press, 1999.
- [2] T. D. Abhayapala and D. B. Ward, "Theory and design of high order sound field microphones using spherical microphone array," in *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*. IEEE, 2002, vol. 2, pp. II-1949.
- [3] M. A. Poletti, "Three-dimensional surround sound systems based on spherical harmonics," *Journal of the Audio Engineering Society*, vol. 53, no. 11, pp. 1004-1025, 2005.
- [4] P. Samarasinghe, T. D. Abhayapala, and M. A. Poletti, "Wavefield analysis over large areas using distributed higher order microphones," *Audio, Speech, and Language Processing, IEEE/ACM Transactions on*, vol. 22, no. 3, pp. 647-658, 2014.
- [5] P. Samarasinghe, T. D. Abhayapala, M. A. Poletti, and T. Betlehem, "An efficient parameterization of the room transfer function," *Audio, Speech, and Language Processing, IEEE/ACM Transactions on*, vol. 23, no. 12, pp. 2217-2227, 2015.
- [6] B. Rafaely, *Fundamentals of Spherical Array Processing*, vol. 8, Springer, 2015.
- [7] P. N. Samarasinghe, T. D. Abhayapala, and M. A. Poletti, "Spatial soundfield recording over a large area using distributed higher order microphones," in *Applications of Signal Processing to Audio and Acoustics (WASPAA), 2011 IEEE Workshop on*. IEEE, 2011, pp. 221-224.
- [8] H. Chen, T. D. Abhayapala, and W. Zhang, "Theory and design of compact hybrid microphone arrays on two-dimensional planes for three-dimensional soundfield analysis," *The Journal of the Acoustical Society of America*, vol. 138, no. 5, pp. 3081-3092, 2015.
- [9] P. N. Samarasinghe, T. D. Abhayapala, and M. A. Poletti, "3d spatial soundfield recording over large regions," in *Acoustic Signal Enhancement; Proceedings of IWAENC 2012; International Workshop on*. VDE, 2012, pp. 1-4.
- [10] S. Brown and D. Sen, "Error analysis of spherical harmonic soundfield representations in terms of truncation and aliasing errors," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. IEEE, 2013, pp. 360-364.
- [11] H. M. Jones, R. A. Kennedy, and T. D. Abhayapala, "On dimensionality of multipath fields: Spatial extent and richness," in *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*. IEEE, 2002, vol. 3, pp. III-2837.
- [12] H. Teutsch, *Modal array signal processing: principles and applications of acoustic wavefield decomposition*, vol. 348, Springer, 2007.
- [13] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *The Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943-950, 1979.
- [14] T. Robinson, J. Fransen, D. Pye, J. Foote, and S. Renals, "Wsjcamo: a british english speech corpus for large vocabulary continuous speech recognition," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 1995, vol. 1, pp. 81-84.
- [15] Y. Hu and P. C. Loizou, "Evaluation of objective quality measures for speech enhancement," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 1, pp. 229-238, 2008.
- [16] B. Bu, T. D. Abhayapala, C. Bao, and W. Zhang, "Parameterization of the three-dimensional room transfer function in horizontal plane," *The Journal of the Acoustical Society of America*, vol. 138, no. 3, pp. EL280-EL286, 2015.