An experimental study on transfer function estimation using acoustic modelling and singular value decomposition

Qiaoxi Zhu[1], Xiaojun Qiu[1], Philip Coleman[2], Ian Burnett[1]

[1] *Centre for Audio, Acoustics and Vibration, Faculty of Engineering and IT, University of Technology Sydney, Sydney, Australia*

[2] *Institute of Sound and Recording, University of Surrey, Guildford, UK*

Transfer functions relating sound source strengths and the sound pressure at field points are important for sound field control. Recently, two modal domain methods for transfer function estimation have been compared using numerical simulations. One is the spatial harmonic decomposition (SHD) method, which models a sound field with a series of cylindrical waves; while the other is the singular value decomposition (SVD) method, which uses prior sound source location information to build an acoustic model and obtain basis functions for sound field modelling. In this paper, the feasibility of the SVD method using limited measurements to estimate transfer functions over densely-spaced field samples within a target region is demonstrated experimentally. Experimental results with various microphone placements and system configurations are reported to demonstrate the geometric flexibility of the SVD method compared to the SHD method. It is shown that the SVD method can estimate broadband transfer functions up to 3099 Hz for a target region with a radius of 0.083 m using three microphones, and allow flexibility in system geometry. Furthermore, an application example of acoustic contrast control is presented, showing that the proposed method is a promising approach to facilitating broadband sound zone control with limited microphones.

I. INTRODUCTION

The transfer functions between loudspeakers and the sound pressure at field points over a region are important for work in sound field control, sound field reproduction (Spors *et al.*, 2013; Betlehem *et al.*, 2015; Choi, 2016), spatial active noise control (Zhang *et al.*, 2018; Qiu, 2019), and sound field equalization in rooms (Radlovic *et al.*, 2000). A simple implementation is to directly measure the transfer functions between the loudspeakers and field points over the control region with the spacing of the field points smaller than half a wavelength (Poletti, 2008). Unfortunately, dense measurements are required for high frequencies, making the approach impractical for general applications.

Various approaches have been proposed for acquiring the transfer functions effectively with a practical number of microphones. According to the Kirchhoff-Helmholtz integral equation, an interior sound field can be expressed by the integral of the sound pressure and its gradient over the boundary of the region (Williams, 1999). Therefore, the transfer functions from a source to the sound field samples inside a source-free region can be estimated from those around the boundary, as long as there are sufficient measurement points on that boundary. The spatial harmonic decomposition (SHD) method expresses the sound field over a spatial region as a set of orthogonal and continuous basis fields (i.e., cylindrical/spherical harmonics) (Williams, 1999; Ward and Abhayapala, 2001; Ahrens and Spors, 2010; Ueno *et al.*, 2019). This method requires the number of boundary samples to be proportional to the radius of the region and the frequency. For a small listener-sized zone (i.e., a circular region with a radius of 0.083 m), a minimum of 15 microphones are required at 3000 Hz, and a minimum of 27 microphones are required at 6000 Hz.

Further approaches have been developed in recent years, such as the distributed higher-order microphone approach (Samarasinghe *et al.*, 2014; Fahim *et al.*, 2017; Ueno *et al.*, 2017) and the

compressive sensing approach (Vinceslas *et al.*, 2020). The former gathers microphones in groups, as compact arrays at several distributed locations for convenient use, but does not reduce the required number of microphone measurements. The latter makes use of optimizations with a clear mathematical background, but the physical underneath is unclear when it comes to the acoustic transfer function measurements. The dictionary matrix involves physical models (i.e., point source(s), plane waves or spatial harmonics). However, derived from mathematically fitting the models with the limited measured values, the basis functions of the sensing matrix do not have a clear link with the physical modes of a sound field. Moreover, the performance of compressive sensing depends on the selection of models, spatial sampling and sparsity strategies (Verburg and Fernandez-Grande, 2018; Koyama and Daudet, 2019; Vinceslas *et al.*, 2020), which might vary from case to case.

Apart from the spatial harmonic method, the singular value decomposition (SVD) method also represents an interior sound field with the basis fields through measurement over boundary samples, which has been applied to both loudspeaker weight design (Zhu *et al.*, 2020b) and loudspeaker placement optimization (Zhu *et al.*, 2018) in a sound zone reproduction system. In contrast to the SHD method, the number of basis functions used in the SVD method is no more than the number of sound sources used to generate a sound field. The small set of basis functions effectively represents the spatial radiation pattern of a loudspeaker array with regard to its geometric distribution [Fig. 2 in (Zhu *et al.*, 2018)]. These studies assume that sufficient transfer function measurements are available. However, challenges remain for the high-frequency range when only a limited number of microphone measurements are available.

As a solution, we recently presented initial work on an improved SVD method using acoustics sensing to acquire the transfer functions over a region (from a test source), when only a few

microphones are available on the boundary of the region (Zhu *et al.*, 2019). The approach derives a set of basis functions by performing SVD on the spatial responses of an acoustic model, which is constructed by prior geometric information, i.e., the sound source location. While the previous SVD method (Zhu *et al.*, 2020b) obtained basis functions directly from measured transfer functions, the proposed SVD method obtains basis functions from an acoustic model composed of virtual monopole point sources around the known source location, the responses of which do not need measurement. With the source location information incorporated, the basis functions obtained by the virtual sources approximate those of the test source.

Initial simulations (Zhu *et al.*, 2019) showed that the proposed SVD method allows transfer function estimation with four microphones equivalent to the full measurement for sound field reproduction, by incorporating acoustic modelling with known source direction [Fig. 2(b) in (Zhu *et al.*, 2019)]. The numerical simulations also showed that the SVD method outperforms the SHD method if some prior geometric information of the sound source location is available (Zhu *et al.*, 2020a). However, as the results were based on ideal simulations, it is unclear whether the measurement-less approach would work in a practical scenario with real loudspeakers and measurement noise.

In this paper, the proposed method is introduced fully, and the feasibility of using the SVD method for transfer function estimation over a small target region is investigated experimentally. The theoretical background and the experiment setup is introduced in Section II. The experimental results on the broadband performance are presented in Section III, including the results with various microphone placements and system configurations to demonstrate the geometric flexibility of the SVD method. In Section IV, practical concerns including the robustness of the SVD method against inaccurate prior information, and potential applications of the approach to broadband sound

zone control with limited microphones, are discussed. The performance compared with the previous method (Zhu *et al.*, 2020b) is also discussed. Conclusions are drawn in Section V.

II. THEORETICAL BACKGROUND AND EXPERIMENTAL SETUP

The objective is to estimate the acoustic transfer functions between a sound source and field points over a region from the measurements at a few points on the boundary of the region. This paper considers the interior sound field problem, where the sound source is outside the region. The basic concepts and equations of the SHD and the SVD methods are summarized in this section to enhance understanding of the experimental results and discussions.

A. Spatial harmonic decomposition (SHD) method

In the SHD method, the sound pressure at $\boldsymbol{r} = (r,\theta)$ within a circular region ($r \le R$ in a polar coordinate system) is estimated with the sound pressure measured at M microphones evenly distributed over the boundary of the region, and the microphone locations are at (R, θ_m), where $\theta_m = 2\pi m/M$, $m = 1, 2, \ldots, M$. The discrete spatial Fourier transform coefficients of the sound pressure in the region can be obtained by (Williams, 1999; Betlehem and Abhayapala, 2005)

$$a_n(k, r_s, \theta_s) = \frac{1}{MJ_n(kR)} \sum_{m=1}^{M} p(k, r_s, \theta_s, R, \theta_m) \mathrm{e}^{\mathrm{j}n\theta_m}, \quad (1)$$

where $k = 2\pi f/c$ is the wavenumber, f is the frequency, c is the speed of sound, $p(k, r_s, \theta_s, R, \theta_m)$ is the sound pressure at the mth microphone from a test source with unit source strength at (r_s, θ_s), and $J_n(*)$ is the nth Bessel function.

The transfer function value between the test sound source and a microphone at any location (r, θ) within the circular region is estimated as

$$p(k, r_s, \theta_s, r, \theta) \approx \sum_{n=-N}^{N} a_n(k) J_n(kr) \mathrm{e}^{-\mathrm{j}n\theta}, \quad (2)$$

where the truncation order $N \geq \lceil k\mathrm{Re}/2 \rceil$, e ≈ 2.718 is the Euler's number and $\lceil * \rceil$ represents the ceiling function. The truncation error of a regional field is upper bounded by $\eta \mathrm{e}^{-\Delta}$ for $N = \lceil k\mathrm{Re}/2 \rceil + \Delta$, where $\Delta \geq 0$ and $\eta \approx 0.16127$ (Kennedy *et al.*, 2007). Thus, the truncation error bound exponentially decays with the increase of the truncation order N. To obtain a sufficient number of coefficients a_n, the required number of measured boundary samples is then

$$M \geq 2N + 1 = 2\lceil k\mathrm{Re}/2 \rceil + 1. \quad (3)$$

For a small listener-sized zone (i.e., a circular region with a radius of 0.083 m), a minimum of 15 microphones are required at 3000 Hz, and a minimum of 27 microphones are required at 6000 Hz. The cutoff frequency for an accurate transfer function estimation is

$$f_{cut,SHD} = \frac{c}{\pi \mathrm{Re}} \left\lfloor \frac{M-1}{2} \right\rfloor, \quad (4)$$

where $\lfloor * \rfloor$ represents the floor function.

B. Transfer function estimation using acoustic modelling and SVD

In the SVD method, the basis functions are obtained through SVD on the spatial responses of an acoustic model constructed by the prior geometric information, i.e., the sound source location. In the acoustic model, a set of I virtual monopole point sources with positions $\boldsymbol{r}_\mathrm{A}(i)$, $i = 1, 2, \ldots, I$, are randomly distributed around the location of the test sound source, as shown in Figure 1, with

$$\boldsymbol{r}_\mathrm{A}(i) = [x_\mathrm{A}(i), y_\mathrm{A}(i)] = (x_s, y_s) + (\Delta x_i, \Delta y_i), \quad (5)$$

where (x_s, y_s) is the known location of the test source in the Cartesian coordinate system and $(\Delta x_i, \Delta y_i)$ is the location deviation of the ith virtual source. The subscript "A" refers to acoustic modelling, to indicate that $\boldsymbol{r}_\mathrm{A}$ refers to the location of the virtual source in acoustic modelling. The transfer function between the ith virtual source and any field point $\boldsymbol{r} = (x, y)$ in the free field is

$$h[k, \boldsymbol{r}_{\mathrm{A}}(i), \boldsymbol{r}] = \frac{1}{d[\boldsymbol{r}_{\mathrm{A}}(i), \boldsymbol{r}]} \mathrm{e}^{-\mathrm{j}kd[\boldsymbol{r}_{\mathrm{A}}(i), \boldsymbol{r}]}, \quad (6)$$

where k is the wavenumber and $d[\boldsymbol{r}_{\mathrm{A}}(i), \boldsymbol{r}] = \sqrt{[x_{\mathrm{A}}(i) - x]^2 + [y_{\mathrm{A}}(i) - y]^2}$ is the distance between the virtual source and the field point.

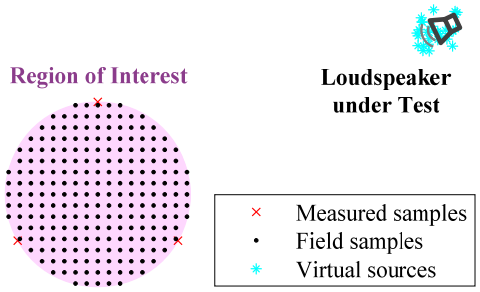


Figure 1 (Color online) Spatial transfer function sensing using a few microphones over the boundary of the region of interest. The transfer functions are measured at boundary samples (×), as the input to estimate the transfer functions received at the field samples (·). In the SVD method, the loudspeaker under test is modelled as a cluster of virtual point sources (*) randomly spaced around the known loudspeaker location.

The location deviations Δx_i and Δy_i are determined based on the uncertainty of the test source location and its acoustics center. It can be difficult in practice to determine the exact acoustic centre of a source. For example, the acoustic centre of a source is not always at its geometric centre and can vary with frequency. Instead of determining the exact acoustic centre, a practical way is to use a cluster of virtual sources with randomized locations, which are averaged at the assumed source centre and distributed within the boundaries of the sound source, as a robust fuzzy representation of the actual acoustic centre. In the experiment in this paper, Δx_i and Δy_i are assumed as independent random variables with a uniform distribution within [−0.02, 0.02] m. It is important

to note that the number of virtual sources I should be sufficiently large to avoid performance perturbation due to the randomness of the virtual source locations; in this paper $I = 100$ is used.

The cluster of the virtual point sources is used to model the test source, because the exact location and spatial radiation features of the test sound source are not always available in practice. On the one hand, randomized location deviations of the virtual sources can model the potential mismatch between the assumed and actual test source location. On the other hand, the virtual sources as an array can reproduce any sound field, within the region of interest, that is generated by any sound source around the known location (Zhu *et al.*, 2020b). Hence, the basis functions obtained from the virtual sources can represent the sound field generated by a test source with an unknown spatial radiation pattern.

Assume $\mathbf{H}_{\mathrm{MA}}$ and $\mathbf{H}_{\mathrm{VA}}$ are the transfer function matrices between the virtual sources and the M measured samples and the V sound field samples, respectively. The subscript "A" refers to acoustic modelling, to indicate that $\mathbf{H}_{\mathrm{MA}}$ and $\mathbf{H}_{\mathrm{VA}}$ are obtained from acoustic modelling rather than transfer function measurement. The sound field samples are densely distributed over the region ($V>>M$). The mth row and the ith column element of $\mathbf{H}_{\mathrm{MA}}$ is the transfer function from the ith virtual source to the mth measured boundary sample, which is $h[k, \boldsymbol{r}_{\mathrm{A}}(i), \boldsymbol{r}_m]$ according to Eq. (6), where $\boldsymbol{r}_m$ is the location of the mth measured sample. The vth row and the ith column element of $\mathbf{H}_{\mathrm{VA}}$ is the transfer function from the ith virtual source to the vth field sample ($v = 1, 2, \ldots, V$), which is $h[k, \boldsymbol{r}_{\mathrm{A}}(i), \boldsymbol{r}_v]$ according to Eq. (6), where $\boldsymbol{r}_v$ is the location of the vth field sample. The basis functions are obtained through SVD, such that

$$\mathbf{H}_{\mathrm{MA}} = \mathbf{U}_{\mathrm{MA}}\boldsymbol{\Sigma}_{\mathrm{MA}}\mathbf{V}_{\mathrm{MA}}^{\mathrm{H}}, \tag{7}$$

$$\mathbf{H}_{\mathrm{VA}} = \mathbf{U}_{\mathrm{VA}}\boldsymbol{\Sigma}_{\mathrm{VA}}\mathbf{V}_{\mathrm{VA}}^{\mathrm{H}}, \tag{8}$$

where $\mathbf{U}_{\mathrm{MA}}$ and $\mathbf{U}_{\mathrm{VA}}$ are unitary matrices whose columns represent the basis functions of the receiver spaces, i.e., the responses at the M measured samples or the V sound field samples; $\mathbf{V}_{\mathrm{MA}}$ and $\mathbf{V}_{\mathrm{VA}}$ are unitary matrices whose columns represent the basis functions of the source space of the virtual sources; and $\mathbf{\Sigma}_{\mathrm{MA}}$ and $\mathbf{\Sigma}_{\mathrm{VA}}$ are diagonal matrices whose elements represent the capability of each source space mode to excite the corresponding receiver space mode (Fazi, 2010; Zhu *et al.*, 2018).

In the SVD method, the transfer function estimation at the field samples is (Zhu *et al.*, 2019)

$$\widetilde{\boldsymbol{p}}_{\mathrm{V}} = \mathbf{H}_{\mathrm{VA}} \cdot \mathbf{H}_{\mathrm{MA}}^{-1} \boldsymbol{p}_{\mathrm{M}}, \tag{9}$$

where the M-element vector $\boldsymbol{p}_{\mathrm{M}}$ represents the measured transfer functions between the test source and the M measured samples, the V-element vector $\widetilde{\boldsymbol{p}}_{\mathrm{V}}$ represents the estimated transfer functions between the test source and the V field samples, and the superscript "–1" denotes the matrix inverse operator. Using the feature of SVD, $\mathbf{H}_{\mathrm{MA}}^{-1} = \mathbf{V}_{\mathrm{MA}} \mathbf{\Sigma}_{\mathrm{MA}}^{-1} \mathbf{U}_{\mathrm{MA}}^{\mathrm{H}}$, where the non-zero elements of $\mathbf{\Sigma}_{\mathrm{MA}}^{-1}$ are the reciprocal of the corresponding elements of $\mathbf{\Sigma}_{\mathrm{MA}}$. Substituting Eqs. (7) and (8) into (9), there is

$$\widetilde{\boldsymbol{p}}_{\mathrm{V}} = \mathbf{U}_{\mathrm{VA}} \mathbf{\Sigma}_{\mathrm{VA}} \mathbf{V}_{\mathrm{VA}}^{\mathrm{H}} \mathbf{V}_{\mathrm{MA}} \mathbf{\Sigma}_{\mathrm{MA}}^{-1} \mathbf{U}_{\mathrm{MA}}^{\mathrm{H}} \boldsymbol{p}_{\mathrm{M}}. \tag{10}$$

Equation 10 shows the process of the SVD based interior sound field sensing. First, $\mathbf{U}_{\mathrm{MA}}^{\mathrm{H}} \boldsymbol{p}_{\mathrm{M}}$ results in the coefficients of the receiver space of the measured samples. Second, the coefficients are translated to those of the source space of the virtual sources by applying the term $\mathbf{\Sigma}_{\mathrm{MA}}^{-1}$. Third, $\mathbf{V}_{\mathrm{VA}}^{\mathrm{H}} \mathbf{V}_{\mathrm{MA}}$ is the projection matrix between the two virtual spaces, allowing the relationship between the measured and estimated microphone positions to be established. When the M measured samples sufficiently sample the sound field at a certain frequency, $\mathbf{V}_{\mathrm{VA}}^{\mathrm{H}} \mathbf{V}_{\mathrm{MA}}$ will lead to a matrix whose first M diagonal elements are ones and the other elements are zeros. See Appendix A for a detailed discussion. Finally, the coefficients of the receiver space of the field samples are obtained

by applying the term $\mathbf{\Sigma}_{\mathbf{VA}}$ and representing the sound field with basis functions as columns of $\mathbf{U}_{\mathrm{VA}}$. Measurement over M samples will result in M coefficients and the M dominant basis functions (first M columns of $\mathbf{U}_{\mathrm{V}}$) are used to represent the transfer functions over V densely spaced field samples. Though the number of basis functions applied is the same as the SHD method, the SVD method uses different basis functions, derived from the test source's geometric features.

C. Experiment Setup

The transfer functions between sound sources to three reproduction zones were measured in a listening room (an acoustically treated room with the reverberation time T_{60} = 0.315 s averaged over octave bands centered at 0.5, 1, and 2 kHz) to evaluate the two approaches for transfer function acquisition (TFA) over the region for a target listener. As illustrated in Fig. 2, the experimental system included 26 sound sources (Genelec 8020B), which were mounted on an arc bar of a sphere frame with a radius of 1.68 m [see Fig. 8 in (Coleman *et al.*, 2014)]. The sound sources were uniformly distributed with an angular spacing of 5.14°. Sound zones with three typical locations were involved in the evaluation. Zone 1 in Fig. 2 was concentric with the arc of the sound source placement. All sources had the same distance to the center of this zone. Zone 2 and Zone 3 were separately 0.5 m leftwards and downwards from Zone 1. The sound zones were circular regions with a radius of 0.083 m.

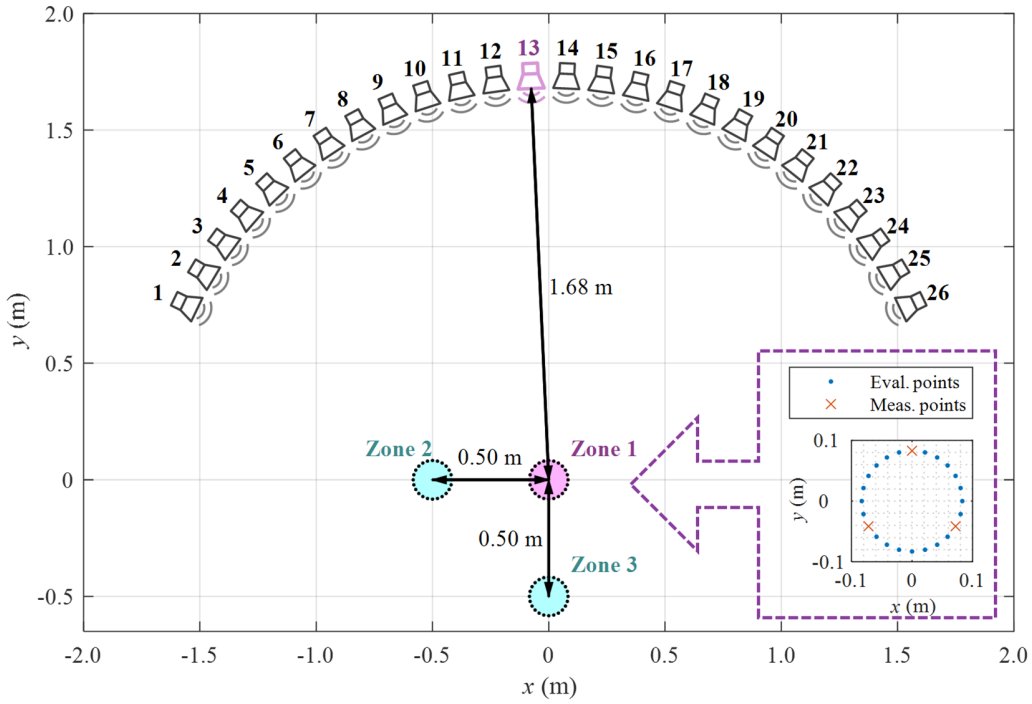


Figure 2 (Color online) The geometry of the experimental system in an acoustically treated room. The 26 sound sources were mounted on an arc bar with a radius of 1.68 m and uniformly distributed with a spacing of 5.14°. Three sound zones with a radius of 0.083 m were separately concentric with the arc of the sound source placement (Zone 1), 0.5 m leftwards (Zone 2) or 0.5 m downwards (Zone 3). The transfer functions over the boundary of each sound zone were measured with a 24-microphone circular array. "×" denotes measurement microphones providing inputs for transfer function estimation and "·" denotes monitor microphones for evaluating transfer function estimation performance.

The transfer functions between each sound source and each sound zone on the horizontal plane were measured using two microphone arrays. First, a uniform circular array composed of 24 omnidirectional microphones (Countryman B3 omni) on the circle with a radius of 0.083 m. This array was used to investigate TFA over a region with different microphone placements over the

boundary, as illustrated in Figs. 2, 8(a) and 9(a), where "×" denotes measurement microphones providing inputs for transfer function estimation and "·" denotes monitor microphones for evaluating transfer function estimation performance. Second, a 10×8 grid array was used to densely measure the transfer function over a region with a spacing of 2.5 cm, fulfilling the Nyquist spatial sampling criterion up to 6.8 kHz, as shown in Fig. 5(a). The 20 omnidirectional microphones (Countryman B3 omni) were assembled to a 5×4 grid array with a spacing of 5 cm on a microphone stand. In order to achieve the required sampling density of microphone locations, four positions of the microphone stand were measured per zone.

The "audioPlayerRecorder" utility in Matlab was used to play and record sound with the sound source and microphone arrays. A multichannel soundcard (MOTU PCIe 424) served as the analog to digital interface, and the microphone inputs were passed through a pre-amplifier (PreSonus Digimax D8). Level differences between microphones were compensated through calibration (114 dB SPL at 1 kHz). The whole system was placed in the listening room. The impulse responses between each microphone position and each sound source were measured using the swept-sine technique (Farina, 2000). Cropping at 4 ms after the impulse onset was applied to the measured room impulse responses (RIRs) to avoid the presence of unnecessary room effects in the transfer functions. The cropping envelope includes a raised cosine ramp at 2 ms duration after the crop time. The frequency domain transfer function was calculated from the cropped impulse responses.

The mean square error is evaluated, which is defined as

$$\mathrm{MSE}(f) = 10\log_{10}\frac{\sum_{v=1}^{V}|\tilde{p}_v(f)-p_v(f)|^2}{\sum_{v=1}^{V}|p_v(f)|^2}, \tag{11}$$

where $\tilde{p}_v(f)$ is the estimated sound pressure and $p_v(f)$ is the measured sound pressure at the vth evaluation point. For clarity and conciseness, the result on TFA between one center sound source (no. 13 in Fig. 2) and the center zone (Zone 1) will be presented in Secs. III.A and B on broadband

TFA performance over the region's boundary or throughout the region, and in Sec. III.C on the effects of microphone placement. The result on TFA between multiple sound sources and multiple zones will be presented in Sec. III.D on the effects of sound source and zone geometry.

III. BROADBAND PERFORMANCE RESULTS

A. Transfer Function Estimation over the Boundary of a Region

Figure 3 presents the measured and estimated transfer functions between the 13th sound source and the 24 samples on the circular boundary of Zone 1 up to 6000 Hz. As illustrated in Fig. 2, three measurement microphones were used to provide input to the transfer function estimation over the other 21 evaluation points on the boundary of the region, using either the SHD method or the SVD method. The magnitude and phase errors were calculated as the absolute value of the differences between the measured and the estimated values, and the results between the 13th sound source and the 24 samples on the circular boundary of Zone 1 up to 6000 Hz is shown in Figure 3, where the transfer function estimation performance degrades with increasing frequency.

The SHD method suffered a magnitude error over 3 dB from 600 Hz, while the 3 dB error occurred from 1746 Hz when using the SVD method. Compared to the magnitude, the phase information was easier to be retrieved by the SVD method. The SVD method maintained phase error within 1.0 Rad over the observed frequency range. However, the SHD method has a phase error over 1.0 Rad from 1184 Hz. Thus, the SVD method achieves better broadband performance in both magnitude and phase using three microphones. Moreover, the SVD method results in high-precision phase estimation over the observed range (up to 6000 Hz), indicating good retrieval of sound orientation and propagation.

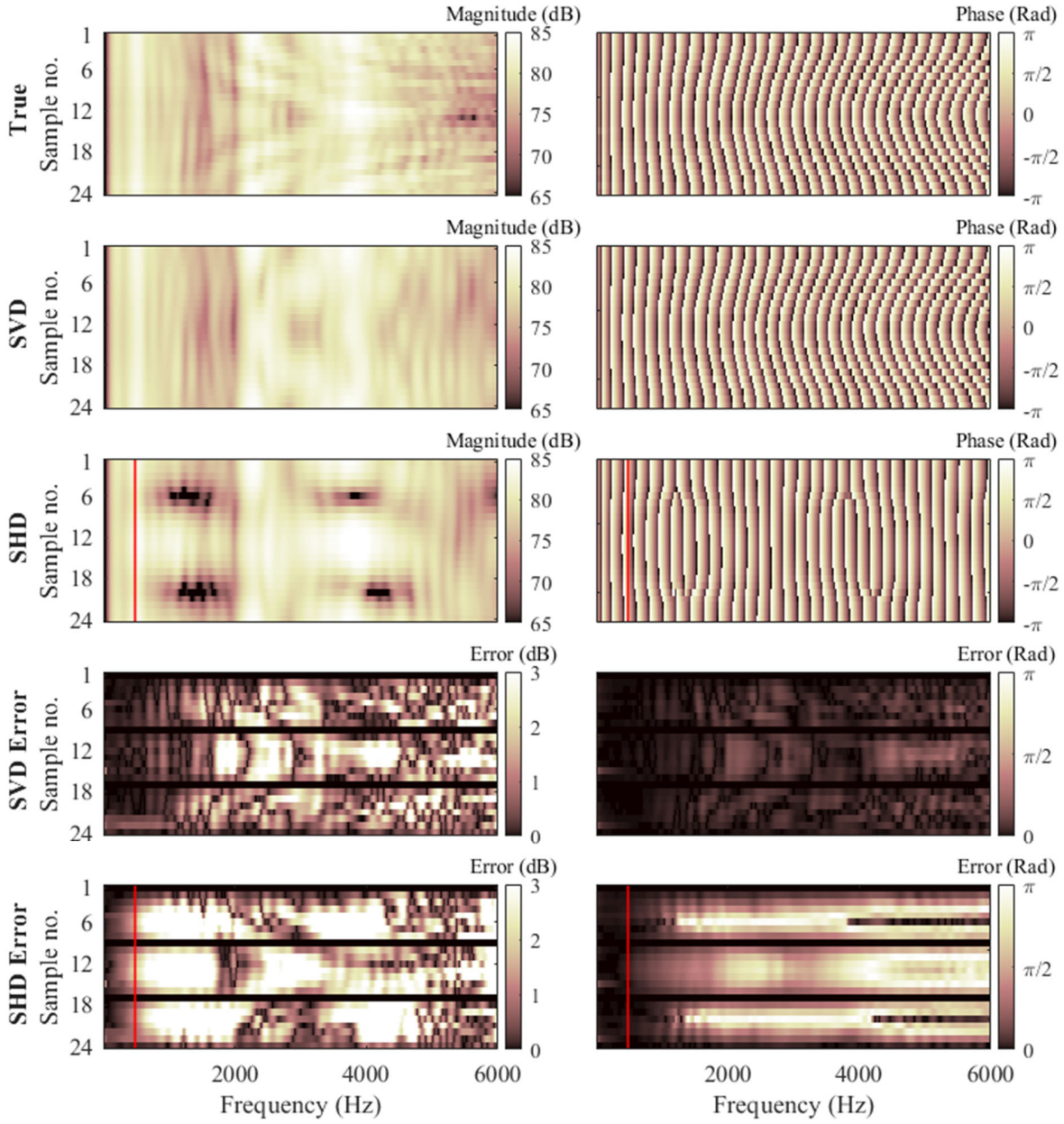


Figure 3 (Color online) The measured and estimated transfer functions from the 13th sound source to the samples over the boundary of Zone 1. Samples are numbered clockwise. The first sample is located at (0, 0.083) m in Fig. 2. Left column presents the magnitude and right column presents phase. The error is also presented. The vertical solid line denotes the cutoff frequency 484 Hz for the SHD method using three microphones.

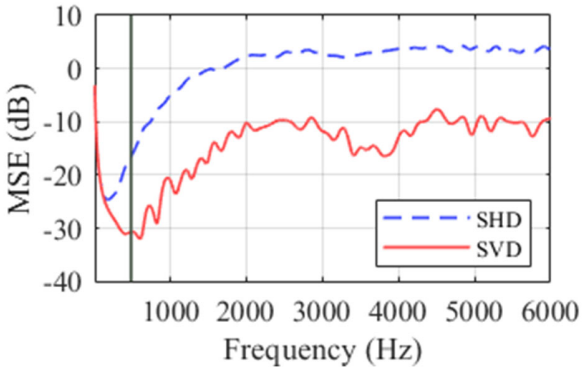


Figure 4 (Color online) The mean square error (MSE) of the sound pressure estimation at the boundary samples with three microphones using the SHD and SVD methods, separately. The vertical solid line denotes the cutoff frequency 484 Hz for the SHD method using three microphones.

Figure 4 presents the mean square error (MSE) of the sound pressure estimation at the boundary samples, separately using the SHD and SVD methods. It is observed that the SVD method has an obvious advantage over the SHD method, with 14.4 dB MSE improvement averaged over the frequency range up to 6000 Hz. Furthermore, the SVD method maintains MSE less than −10 dB for most frequencies over the observed frequency range up to 6000 Hz.

B. Transfer Function Estimation over an Interior Region

In Figure 5 the SVD method and SHD method are compared in terms of local sound field sensing over a square region (using three microphones). The SVD method leads to the same sound field distribution pattern as the measured values, while the SHD method does not. This is because the approaches apply different basis sound fields to represent the local sound field. Figure 6 shows the three basis sound fields used in the SHD and SVD methods, with their coefficients at 1300 Hz presented in Fig. 7.

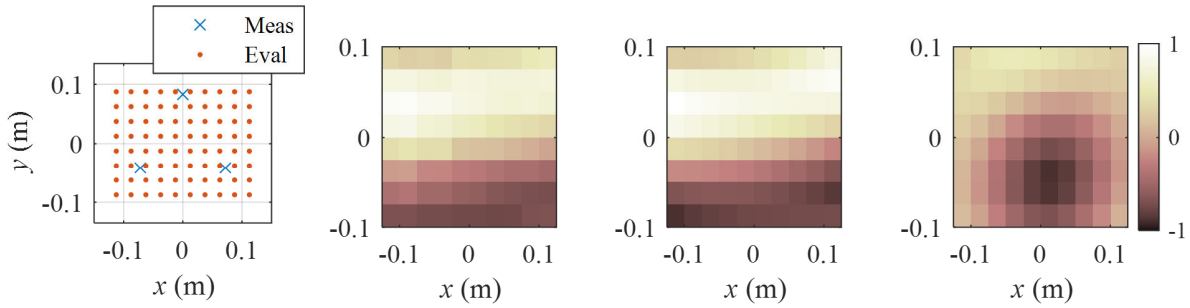


Figure 5 (Color online) The normalized real part of the transfer function distribution at 1300 Hz. (a) The 10×8 evaluation samples are evenly distributed over a grid with a spacing of 0.025 m, and three microphones (denoted as "×") are used for transfer function estimation. (b) The measured values for the evaluation samples. (c) The estimated values for the evaluation samples using SVD with three microphones. (d) The estimated values for the evaluation samples using SHD with three microphones.

Figure 6 presents the spatial patterns of the three modes used in the SHD and SVD methods separately. The SHD modes are the first three cylindrical modes, $J_n(kr)\mathrm{e}^{-\mathrm{j}n\theta}$, $n = -1, 0, 1$, respectively; while the SVD modes are the first three column vectors of $\mathbf{U}_{\mathrm{VA}}$ in Eq. (8). For comparison, the magnitude of both the SHD modes and the SVD modes is normalized over the observed field samples. The wavelength is 0.26 m at the observed frequency of 1300 Hz. The observed area in Figure 6 is 0.25 m long and 0.20 m wide. It is observed that the spatial resolution of the SHD modes is half the wavelength, while that of the SVD modes is finer. The SHD modes are sensitive to, and change dramatically with the distance to the center of the region. In contrast, the SVD modes change relatively gently with space without losing the orthogonality. The distribution of a homogeneous source-free sound field should not be affected by a 'virtual' region

center and, thus, SVD modes can better match the actual sound field with a finer and center-free sound field decomposition.

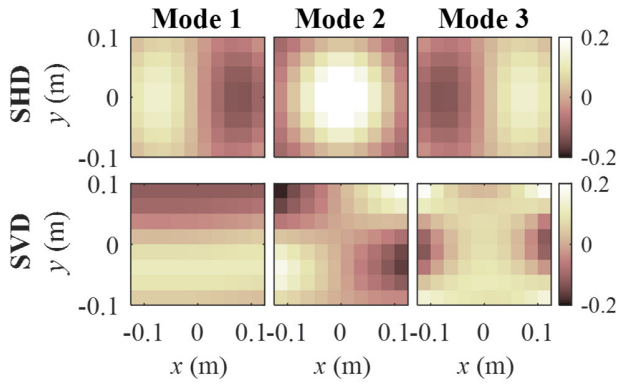


Figure 6 (Color online) The real part of the sound field distribution of each normalized mode at 1300 Hz, when using the SHD and SVD methods, separately, with three microphones.

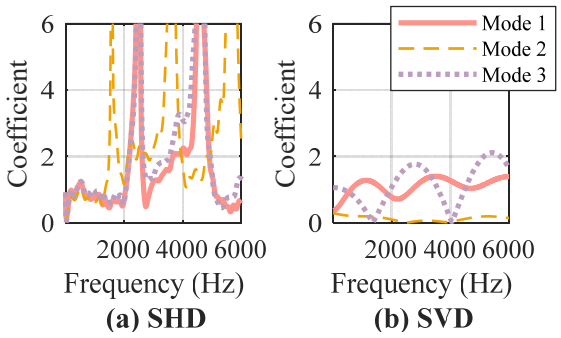


Figure 7 (Color online) The magnitude of the coefficients of the three normalized modes when using the SHD and SVD methods, separately, with three microphones.

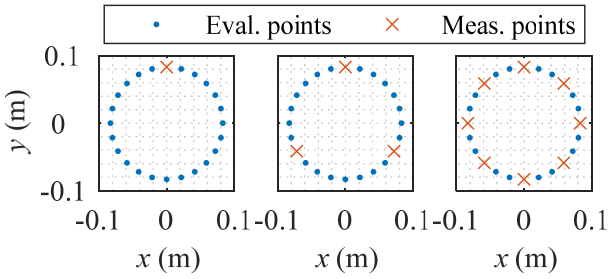
Figure 7 presents the coefficients of the three normalized modes in the SHD and SVD methods using Eqs. (1) and (8), respectively. Only the magnitude is presented to indicate the contribution of each mode. The figure shows that the SHD modes work equally, except for frequencies when

the microphone configuration meets the null(s) of an SHD mode. In contrast, the SVD modes take the major share by turns in a periodical way, which is in accordance with the modal behavior of the regional sound field. That is, the intensity of each mode changes with frequency and the peaks are achieved when the frequency meets an eigen-frequency of the corresponding mode.

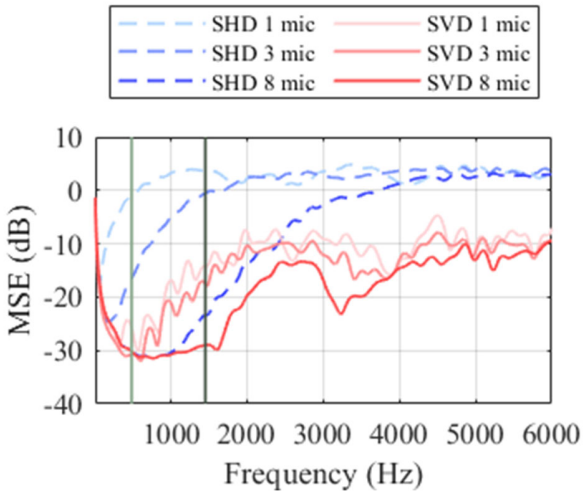
C. Microphone placement

The experimental results show that the SVD method can facilitate TFA over a region for a broader frequency band than the SHD method, when only three microphones are available, because the three modes applied in the SVD method match the actual sound field better than those in the SHD method. Figure 8 presents the performance comparison with different numbers of microphones. Three microphone configurations, namely one, three and eight microphone(s), are investigated for TFA at the boundary samples. As expected, the mean square error decreases with the increasing number of microphones. However, compared with the SHD method, the performance of the SVD method is less affected by the number of microphones. Furthermore, even with eight microphones, the MSE performance of the SVD method is at least 6 dB better than the SHD method from 2 kHz.

Figure 9 presents the performance comparison with different placements of three microphones. Even spacing is kept for the implementation of the SHD method. The SVD method is seen to be affected by the microphone placement, due to the changing geometric relationship with the test sound source. However, –10 dB performance is almost maintained for the three placements over the frequency range up to 3099 Hz when using the SVD method. This critical frequency is where the diameter of the sound zone is equal to half the wavelength times the number of microphones.

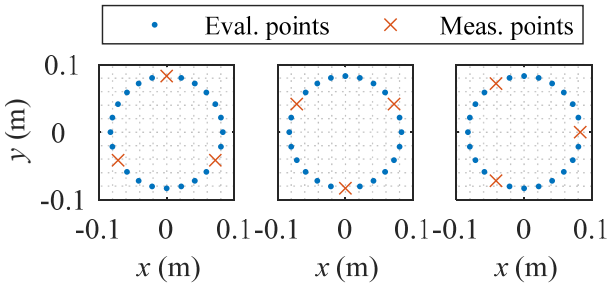


(a)

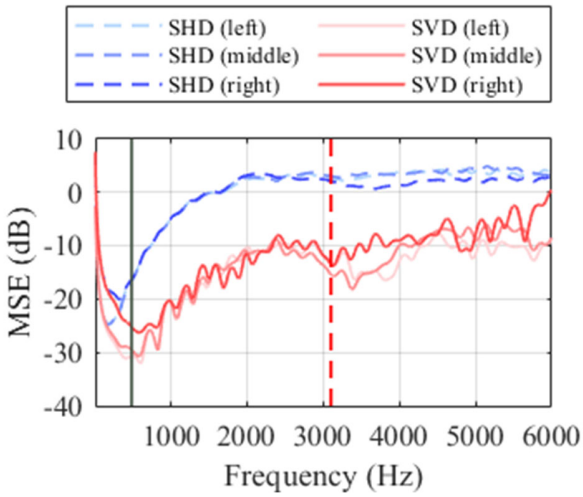


(b)

Figure 8 (Color online) Performance with different number of microphones. (a) One, three or eight microphone(s) with regular spacing. (b) The mean square error (MSE) of the sound pressure estimation at the boundary samples with one, three or eight microphone(s) using the SHD and SVD methods, separately. The vertical solid lines denote the cutoff frequency 484 Hz and 1452 Hz for the SHD method using three or eight microphones, respectively.



(a)



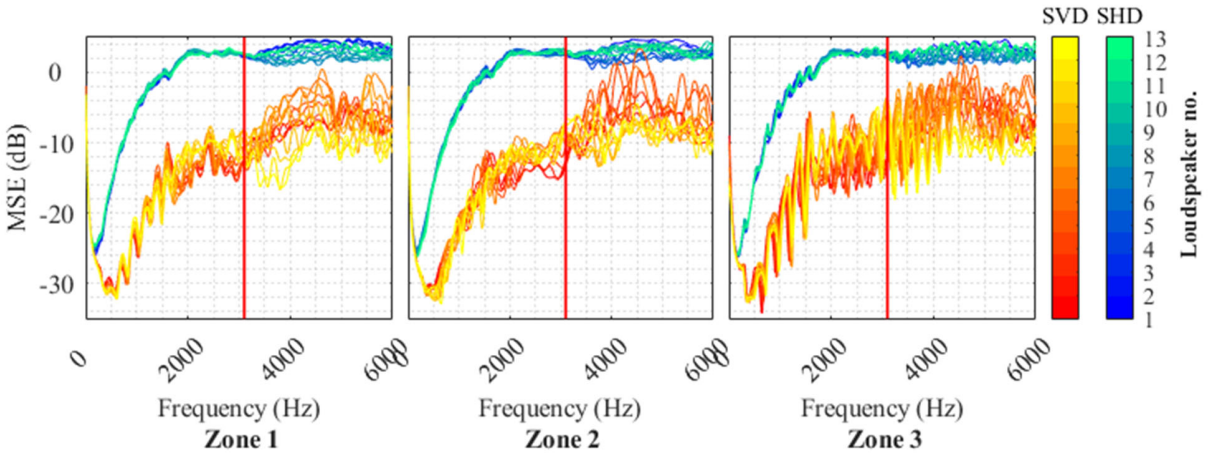
(b)

Figure 9 (Color online) Performance with different placement of three microphones. (a) Upward-pointing triangle placement (left), downward-pointing triangle placement (middle) and right-pointing triangle placement. (b) The mean square error (MSE) of the sound pressure estimation at the evaluated boundary samples with three microphones of different placements using the SHD and SVD methods, separately. The vertical solid line denotes the cutoff frequency 484 Hz for the SHD method using three microphones. The vertical dashed line denotes the critical frequency of 3099 Hz.

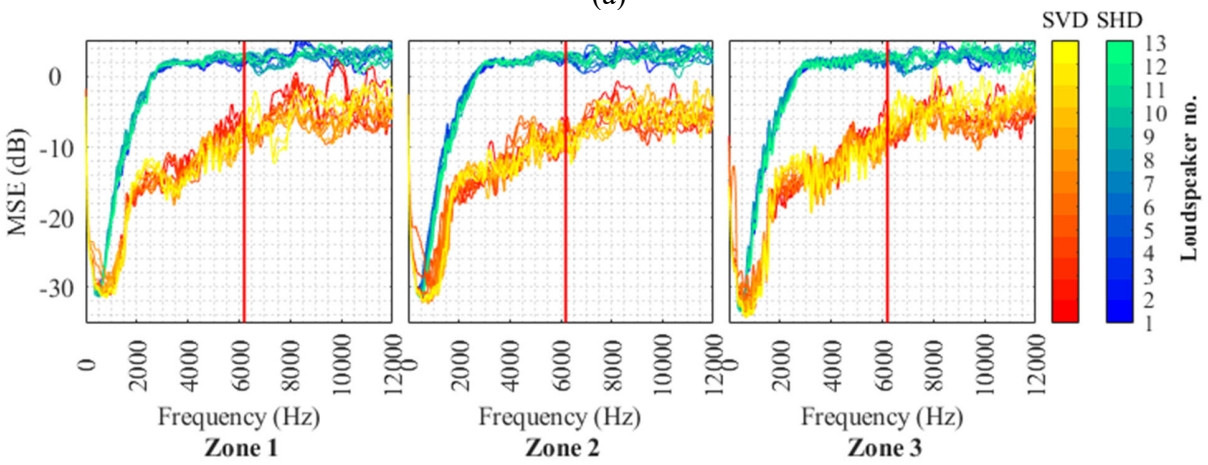
D. Effect from sound source and sound zone configuration

Figure 10(a) presents the MSE performance per sound source, per zone, using three microphones with the geometry illustrated in Figure 2. Compared to Zone 1, the distance between sound sources to Zone 2 or Zone 3 has an increased variation. Thus, the curves of the MSE performance fluctuate even more dramatically. However, the SVD method maintains an MSE less than −10 dB for most cases below 3099 Hz. With the doubled number of microphones evenly distributed over the circular boundary, the increased critical frequency of 6199 Hz can be observed in Fig. 10(b). Besides, the narrowed microphone spacing also results in reduced MSE performance fluctuation. For clarity and conciseness, only the result of the first 13 sound sources (no. 1~13 in Fig. 2) are presented. It was observed that the above findings held for the other 13 sound sources (no. 14~26).

The performance fluctuations resulting from the system geometry change mainly occurs above a critical frequency, which is determined by the size of the region and the number of the microphones. The geometry of the microphone placement or system configuration does not significantly affect the TFA below the critical frequency. The increased number of microphones not only increases the critical frequency but also reduces the performance fluctuations.



(a)



(b)

Figure 10 (Color online) Performance over 26 sound source locations and three region locations, respectively. (a) using three microphones and observed up to 6 kHz and (b) using six microphones and observed up to 12 kHz. The vertical solid lines denote the critical frequencies 3099 Hz in (a) and 6199 Hz in (b).

IV. DISCUSSIONS

A. Effect due to the accuracy of the prior information

Due to the incorporation of prior geometric information, the SVD method reduces the required number of microphones for TFA over a region for a target listener. But the performance is affected when the sound source location information is not accurate. Figure 11 shows the performance when incorrect location information is applied to estimate the transfer function of the 13th sound source. The acoustic model of the 13th sound source is centered around the location of the 13th, 10th, 7th, 4th and 1st sound source, respectively, denoted as AM13, AM10, AM7, AM4 and AM1 in Figures 11. The location mismatch is separately 0°, 15.42°, 30.84°, 61.68° and 123.36°. The corresponding deviation range is 0° (no deviation, for ±0°), 30.84° (for ±15.42°), 61.68° (for ±30.84°), 123.36° (for ±61.68°) and 246.72° (for ±123.36°), where zero deviation is the center of each deviation range. The performance of the SVD method degrades dramatically with increasing deviation between the assumed and actual sound source location.

When the deviation range is larger than the microphone spacing (120° for the 3-microphone array and 60° for the 6-microphone array), the SVD performance deteriorates, being the same or worse than SHD. Thus, the SVD method outperforms the SHD method when

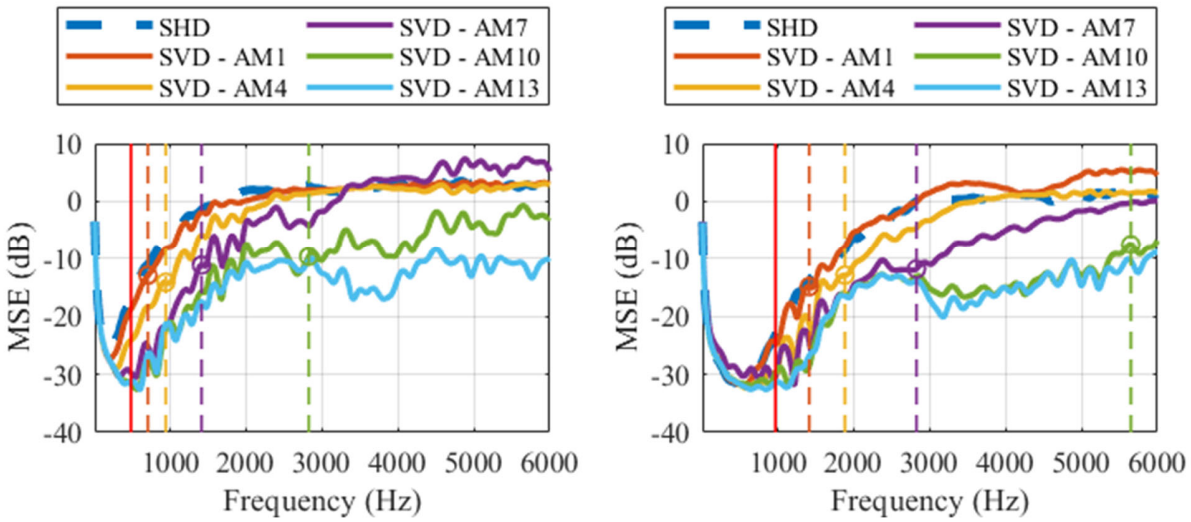
$$\Delta \leq \frac{2\pi}{M}, \tag{12}$$

where Δ (Rad) is the deviation range of the assumed sound source location away from the actual location ($0 \leq \Delta \leq \pi$) and M is the number of the microphones. Under this condition, the cutoff frequency of the SVD method is estimated as

$$f_{\text{cut,SVD}} = \frac{\pi}{\Delta} f_{\text{cut,SHD}} \geq \frac{M}{2} f_{\text{cut,SHD}} = \frac{M(M-1)c}{4\pi R\text{e}}. \tag{13}$$

The cutoff frequency is inversely proportional to the deviation range of the location mismatch, and is larger than that of the SHD method. As shown in Fig. 11, the MSE performance of the SVD

method maintains approximately or below −10 dB below the cutoff frequency. Thus, apart from increasing the number of microphones, improving the accuracy of geometric information also helps maintain acceptable broadband performance.



(a) (b)

Figure 11 (Color online) Performance of the 13th sound source's transfer function estimation using (a) three microphones or (b) six microphones. "AMx" denotes the location of the xth sound source being assumed as the location of the 13th sound source, when using the SVD method. The vertical solid lines denote 484 Hz and 968 Hz, which are the cutoff frequencies for the SHD method using three and six microphones, respectively. The vertical dashed lines denotes the cutoff frequencies resulting from different sound source location mismatches [Eq. (13)] and the markers "○" denote the SVD performance at these cutoff frequencies.

B. Potential application in broadband sound zone control

Figure 12 shows the performance when the 26 loudspeakers in Fig. 2 are used to deliver sound to a target listener at Zone 2 (as the listening zone) while minimizing sound propagation to the other user at Zone 3 (as the quiet zone). The loudspeaker weights are calculated using the acoustic contrast control optimization [Eq. 11 in (Zhu *et al.*, 2017)]. The inputs for the optimization are the transfer functions between the 26 loudspeakers to the listening zone and the quiet zone, respectively. "Direct" in Figure 12 denotes the transfer functions measured from three microphones per zone (as in Fig. 2); "SHD" and "SVD" denote the estimated transfer functions of 32 field points per zone, evenly distributed with a spacing of 0.025 m, from three microphone measurement using either the SHD or SVD method; and "Full Meas." denotes transfer functions measured using 32 field points per zone, as the reference case usually utilized in the literature.

The reproduced sound zones applying the four different transfer functions are evaluated by the acoustic contrast metric and the array effort metric [Eqs. (16) and (17) in (Zhu *et al.*, 2017)] in Figure 12. A high acoustic contrast value refers to a high sound energy difference between the listening and quiet zone, and a low array effort value refers to reproduction with high energy efficiency. The averaged performance over the observed frequency range up to 6000 Hz is, respectively, 18.3 dB, 17. 3 dB, 28.1 dB and 36.8 dB in acoustic contrast and −8.2 dB, −7.6 dB, −9.3 dB and −9.8 dB in array effort for the "Direct", SHD, SVD and full measurement approaches. Compared to the direct method, the SVD method has improved acoustic contrast and array effort performance at high frequencies above 1900 Hz, while the SHD method has no benefit when only a few microphones are available.

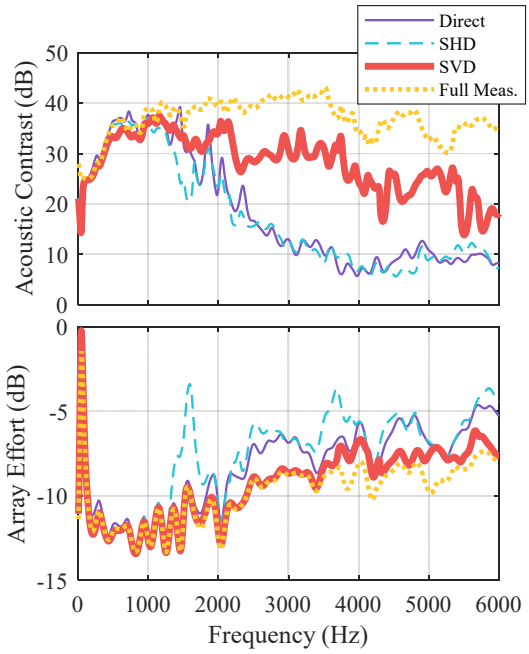


Figure 12 (Color online) Performance of two zone reproduction with transfer functions measured from three microphones per zone ("Direct"), estimated from three microphone measurement using the SHD or SVD method, and measured from 32 field points per zone evenly distributed with a spacing of 0.025 m ("Full Meas").

The "Direct" case applies the previous SVD method (Zhu *et al.*, 2020b), which does not apply transfer function sensing, and only uses the information from the three microphone measurement to implement the sound zone control. In addition to the information from the three microphone measurement, the SVD method in this paper also incorporates acoustic modelling with the source location information. So the SVD method can model and represent the acoustic responses with actual loudspeakers and microphones and thus achieve sound zone control over a broader frequency range.

Figure 13 shows the reproduced sound zones at 3000 Hz, where it is clear that when the microphones are insufficient, the SHD method degrades to the direct method, while the SVD

method has better control over both the listening and quiet zones. The SVD method obtains better performance because the interior sound field is better estimated through geometric information based acoustic modelling.

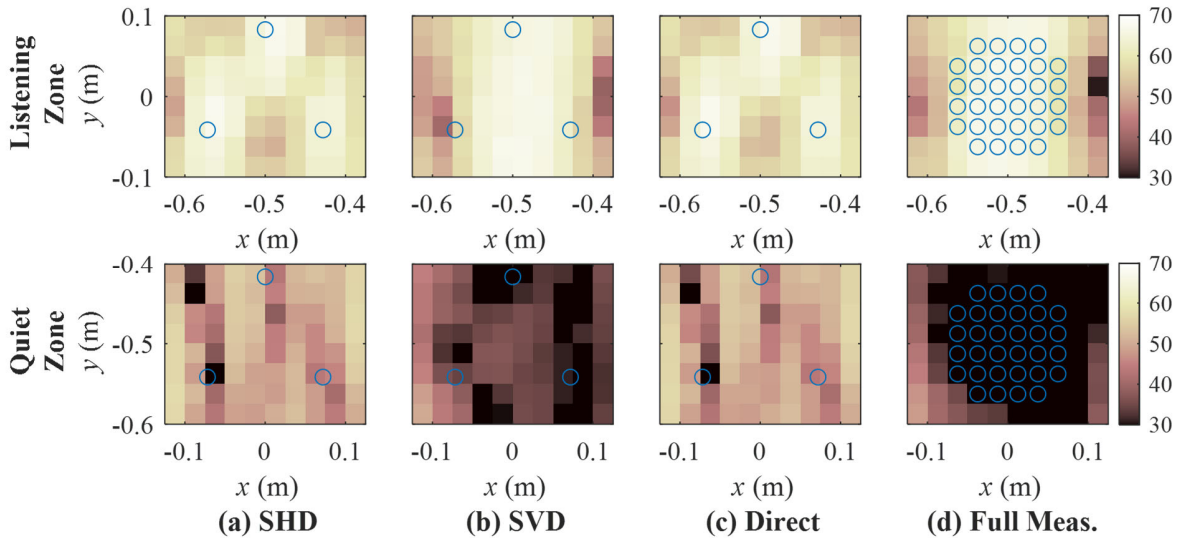


Figure 13 (Color online) The reproduced sound pressure levels at the listening zone and the quiet zone at 3000 Hz with transfer functions measured from three microphones per zone ("Direct"), estimated from three microphone measurement using the SHD or SVD method, and measured from 32 field points per zone evenly distributed with a spacing of 0.025 m ("Full Meas").

It should be noted that, as mentioned in Sec. II. C, the RIRs are truncated in this experimental evaluation, indicating the performance under a heavily acoustically treated environment dominated by direct sound. Nevertheless, this situation facilitates a clear demonstration of the utility of the SVD method for transfer function estimation over a region.

V. CONCLUSIONS

This work presented an experimental validation of the singular value decomposition method for transfer function acquisition over a region. This method shows an accurate transfer function acquisition (with less than −10 dB error) up to 3099 Hz over the region with a radius of 0.83 m, using only three microphones. The performance comparison demonstrates that the singular value decomposition method facilitates accurate transfer function acquisition over a much broader frequency band than the spatial harmonic decomposition method. The performance has been evaluated with various microphone placements and system configurations to demonstrate the geometric flexibility of the singular value decomposition method. Furthermore, the effect of incorporating inaccurate sound source location information has been analyzed, and the potential application to broadband sound zone control with a limited number of microphones has been considered. Future work includes incorporating early reflection detection and modelling to apply the approach in more general acoustic environments.

APPENDIX A

Both the M measured samples and the V sound field samples are sampling the same regional sound field generated by the same I virtual sound sources. The columns of $\mathbf{U}_{MA}$ and $\mathbf{U}_{VA}$ are the basis fields obtained from the two samplings. Suppose the M measured samples can sufficiently sample the sound field at a certain frequency. In that case, the columns of $\mathbf{U}_{MA}$ and the first M columns of $\mathbf{U}_{VA}$ will both represent the M most efficient modes generated by the given array (I virtual sound sources) to the given receiving region at that frequency. Besides, the order of the column vectors in $\mathbf{U}_{MA}$ will be the same as the first M column vectors in $\mathbf{U}_{VA}$. This is because the order of the column vectors in $\mathbf{U}_{MA}$ and $\mathbf{U}_{VA}$ and the order of the diagonal elements in $\mathbf{\Sigma}_{MA}$ and

$\mathbf{\Sigma}_{VA}$, are determined by the radiation efficiency of the source array composed of the same I virtual sound sources due to the feature of SVD.

Accordingly, the columns of $\mathbf{V}_{MA}$ will represent the same modes as the first M columns of $\mathbf{V}_{VA}$. Since the columns of $\mathbf{V}_{MA}$ and $\mathbf{V}_{VA}$ represent the modes of the same source spaces composed of the I virtual sources, the columns of $\mathbf{V}_{MA}$ will be equal to the first M columns of $\mathbf{V}_{VA}$. Since $\mathbf{V}_{MA}$ and $\mathbf{V}_{VA}$ are unitary matrices, $\mathbf{V}_{VA}^{H}\mathbf{V}_{MA}$ will lead to a matrix whose first M diagonal elements are ones and other elements are zeros.

ACKNOWLEDGEMENTS

This research is supported under the Australian Research Council's Linkage Project funding scheme (LP160100616), the FEIT Tech Lab Blue Sky Research Grant from the University of Technology Sydney, and the FASS pump-priming Travel Fund 2019-20 from the University of Surrey.

REFERENCES

Ahrens, J., and Spors, S. (**2010**). "Sound field reproduction using planar and linear arrays of loudspeakers," IEEE transactions on audio, speech, and language processing **18**, 2038-2050.

Betlehem, T., and Abhayapala, T. D. (**2005**). "Theory and design of sound field reproduction in reverberant rooms," The Journal of the Acoustical Society of America **117**, 2100-2111.

Betlehem, T., Zhang, W., Poletti, M. A., and Abhayapala, T. D. (**2015**). "Personal sound zones: Delivering interface-free audio to multiple listeners," IEEE Signal Processing Magazine **32**, 81-91.

Choi, J.-W. (**2016**). "Recent advances in sound field control systems: From personal audio to smart sound systems," in *2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)* (IEEE), pp. 1-2.

Coleman, P., Jackson, P. J., Olik, M., and Abildgaard Pedersen, J. (**2014**). "Personal audio with a planar bright zone," The Journal of the Acoustical Society of America **136**, 1725-1735.

Fahim, A., Samarasinghe, P. N., and Abhayapala, T. D. (**2017**). "Sound field separation in a mixed acoustic environment using a sparse array of higher order spherical microphones," in *2017 Hands-free Speech Communications and Microphone Arrays (HSCMA)* (IEEE), pp. 151-155.

Farina, A. (**2000**). "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Audio Engineering Society Convention 108* (Audio Engineering Society).

Fazi, F. M. (**2010**). "Sound field reproduction," (University of Southampton).

Kennedy, R. A., Sadeghi, P., Abhayapala, T. D., and Jones, H. M. (**2007**). "Intrinsic limits of dimensionality and richness in random multipath fields," IEEE Transactions on Signal processing **55**, 2542-2556.

Koyama, S., and Daudet, L. (**2019**). "Sparse representation of a spatial sound field in a reverberant environment," IEEE Journal of Selected Topics in Signal Processing **13**, 172-184.

Poletti, M. (**2008**). "An investigation of 2-d multizone surround sound systems," in *Audio Engineering Society Convention 125* (Audio Engineering Society).

Qiu, X. (**2019**). *An Introduction to Virtual Sound Barriers* (CRC Press).

Radlovic, B. D., Williamson, R. C., and Kennedy, R. A. (**2000**). "Equalization in an acoustic reverberant environment: Robustness results," IEEE Transactions on Speech and Audio Processing **8**, 311-319.

Samarasinghe, P., Abhayapala, T., and Poletti, M. (**2014**). "Wavefield analysis over large areas using distributed higher order microphones," IEEE/ACM Transactions on Audio, Speech, and Language Processing **22**, 647-658.

Spors, S., Wierstorf, H., Raake, A., Melchior, F., Frank, M., and Zotter, F. (**2013**). "Spatial sound with loudspeakers and its perception: A review of the current state," Proceedings of the IEEE **101**, 1920-1938.

Ueno, N., Koyama, S., and Saruwatari, H. (**2017**). "Sound field recording using distributed microphones based on harmonic analysis of infinite order," IEEE Signal Processing Letters **25**, 135-139.

Ueno, N., Koyama, S., and Saruwatari, H. (**2019**). "Three-Dimensional Sound Field Reproduction Based on Weighted Mode-Matching Method," IEEE/ACM Transactions on Audio, Speech, and Language Processing **27**, 1852-1867.

Verburg, S. A., and Fernandez-Grande, E. (**2018**). "Reconstruction of the sound field in a room using compressive sensing," The Journal of the Acoustical Society of America **143**, 3770-3779.

Vinceslas, L., Lim, H., and Kondoz, A. (**2020**). "Evaluation of Sparse Sound Field Models for Compressed Sensing in Multiple Sound Zones," in *Audio Engineering Society Convention 148* (Audio Engineering Society).

Ward, D. B., and Abhayapala, T. D. (**2001**). "Reproduction of a plane-wave sound field using an array of loudspeakers," IEEE Transactions on speech and audio processing **9**, 697-707.

Williams, E. G. (**1999**). *Fourier acoustics: sound radiation and nearfield acoustical holography* (Elsevier).

Zhang, J., Abhayapala, T. D., Zhang, W., Samarasinghe, P. N., and Jiang, S. (**2018**). "Active noise control over space: A wave domain approach," IEEE/ACM Transactions on audio, speech, and language processing **26**, 774-786.

Zhu, Q., Coleman, P., Qiu, X., Wu, M., Yang, J., and Burnett, I. (**2018**). "Robust personal audio geometry optimization in the SVD-based modal domain," IEEE/ACM Transactions on Audio, Speech, and Language Processing **27**, 610-620.

Zhu, Q., Coleman, P., Wu, M., and Yang, J. (**2017**). "Robust acoustic contrast control with reduced in-situ measurement by acoustic modelling," Journal of the Audio Engineering Society **65**, 460-473.

Zhu, Q., Qiu, X., and Burnett, I. (**2020a**). "An acoustic modelling based remote error sensing approach for quiet zone generation in a noisy environment," in *Proceedings.(ICASSP'20). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2020.* (IEEE).

Zhu, Q., Qiu, X., Coleman, P., and Burnett, I. (**2019**). "Reducing number of transfer function measurement in local sound field reproduction using acoustic modeling," in *the 23rd International Congress on Acoustics (ICA 2019)* (Aachen, Germany).

Zhu, Q., Qiu, X., Coleman, P., and Burnett, I. (**2020b**). "A comparison between two modal domain methods for personal audio reproduction," The Journal of the Acoustical Society of America **147**, 161-173.