Robust Acoustic Contrast Control with Reduced In-situ Measurement by Acoustic Modeling*

QIAOXI ZHU[1,2], PHILIP COLEMAN[3], MING WU[1,2], AND JUN YANG,[1,2] *AES Member*

(qiaoxi.zhu@gmail.com)   (p.d.coleman@surrey.ac.uk)   (mingwu@mail.ioa.ac.cn)   (jyang@mail.ioa.ac.cn)

[1]*Key Laboratory of Noise and Vibration Research, Institute of Acoustics, Chinese Academy of Sciences, Beijing, 100190, China*
[2]*University of Chinese Academy of Sciences, Beijing, 100049, China*
[3]*Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford, Surrey, GU2 7XH, UK*

Personal audio systems generate a local sound field for a listener while attenuating the sound energy at pre-defined quiet zones. In practice, system performance is sensitive to errors in the acoustic transfer functions between the sources and the zones. Regularization is commonly used to improve robustness, however, selecting a regularization parameter is not always straightforward. In this paper a design framework for robust reproduction is proposed, combining transfer function and error modeling. The framework allows a physical perspective on the regularization required for a system, based on the bound of assumed additive or multiplicative errors, which is obtained by acoustic modeling. Acoustic contrast control is separately combined with worst-case and probability-model optimization, exploiting limited knowledge of the potential error distribution. Monte-Carlo simulations show that these approaches give increased system robustness compared to the state of the art approaches for regularization parameter estimation, and experimental results verify that robust sound zone control is achieved in the presence of loudspeaker gain errors. Furthermore, by applying the proposed framework, in-situ transfer function measurements were reduced to a single measurement per loudspeaker, per zone, with limited acoustic contrast degradation of less than 2 dB over 100–3000 Hz compared to the fully measured regularized case.

0 INTRODUCTION

Personal audio reproduction aims to use a loudspeaker array to render desired sound to a listening position with minimal interference to listeners in other regions. Various methods for sound zone reproduction are discussed and compared in [1] and [2]. Several of these algorithms are based on input transfer functions (TFs) between the loudspeaker array and the control microphones, for example: acoustic contrast control (ACC) [3], pressure matching [4, 5], planarity control [6], and sound field reproduction ACC [7]. These systems are vulnerable to perturbations in the assumed TFs, which might result from inconsistencies in the individual sources' sensitivities, complexities in the spatial responses, source position mismatches, and so on [8]. In addition, sound speed changes caused by variations in, e.g., temperature, humidity and airflow, lead to changes in the

TFs and eventually lower the system performance [9]. Besides these time-variant perturbations within a reproduction system, there are also perturbations from system to system, for example due to variations among loudspeakers of the same model. Furthermore, it might not always be possible to conduct in-situ measurements in the zones to calibrate or adjust installed systems.

In order to reduce performance degradation in TF-based methods under such conditions, robustness should be carefully considered in the algorithm design. Kim et al. [10] and Elliott et al. [8] applied constraints on array audible gain or array effort (AE) to increase system robustness. Coleman [9] studied the effect and selection of a regularization parameter to improve overall performance considering acoustic contrast (AC), planarity, and AE. These methods only utilize the characteristics of the array and the application geometry, without considering potential errors in the system. The AE constraints were also employed by Cheer et al. for practical systems in a car cabin [11] and a mobile device [12].

Park et al. [13] discussed the relationship between TF errors and the resulting AC but did not propose a robust

---

*Portions of this paper were presented in Q. Zhu et al., "Robust Personal Audio Reproduction Based on Acoustic Transfer Function Modeling," at the 2016 *AES International Conference on Sound Field Control*, Guildford, UK, July 2016.

control solution. Bai et al. [14] incorporated information about the errors in the system by using Monte-Carlo simulations to find the best regularization parameter for a microphone array. This approach has high computational complexity and is constricted by the pre-given solution form, i.e., diagonal loading.

Error-based robust solutions, utilizing both system geometry and error distribution information, can be derived by employing different optimization strategies, e.g., probability-model optimization (PMO) and worst-case optimization (WCO). The concept of utilizing error information by PMO and WCO to increase system robustness was originally investigated for robust beamforming, e.g., [15] and [16]. For personal audio, Zhang et al. [17] utilized PMO to achieve robust ACC, assuming multiplicative noise in the TFs. Cai et al. [18] also studied robust ACC, considering additive gain and phase variations, using PMO. A special case of PMO [8] treated the reverberant field above the Schroeder frequency as diffuse, and increased robustness by introducing a spatially-averaged mean square uncertainty into diagonal loading regularization. This approach was experimentally validated in [19]. Cai [20] also applied WCO to ACC and discussed its relationship to the PMO approaches. However, the relationship between error-based methods, such as WCO and PMO, and other regularization methods has not been systematically elaborated in the literature.

In addition to errors in the system, the robust control performance depends on both the determined part and the error part of the TF model. The determined part and the error part should both be modeled accurately for the most robust performance. This can be realized by incorporating acoustic modeling. As an example of determined part adjustment, Chang et al. [21] and Tu et al. [22] investigated the incorporation of a rigid sphere model into TFs to help increase robustness against head scattering effects. In our previous work [23] we proposed a general framework for robust sound reproduction making use of acoustic modeling to obtain input for error-based robust optimization (i.e., PMO, WCO). The acoustic modeling described the ideal radiation pattern of the loudspeakers to the control points and additionally represented gain and phase variations in the TFs with an estimated error model.

In this article we extend our work in [23] with three new contributions. First, we present new analysis of the problem of selecting a diagonal loading parameter for robust ACC, including a method based on the maximum singular value of the TF matrix. Second, we include experimental validation results, which evaluate the performance of robust control techniques in the presence of perturbations introduced to the loudspeaker gains. Third, we investigate a practical application of error modeling to reduce the need for in-situ TF measurement. Specifically, for each loudspeaker, we make a single TF measurement to each zone and use a point-source interpolation and uncertainty model to populate the remaining rows of the TF matrices.

In Sec. 1 we outline the acoustic-modeling based robust reproduction framework and provide the robust ACC solutions derived by two optimization strategies (PMO and WCO). In Sec. 2 the robust performance pattern of a typical ACC system is analyzed by Monte-Carlo simulations, which shows the advantage of acoustic-modeling based methods. Experimental results are presented in Sec. 3, comparing robust control strategies when there are errors in the system due to loudspeaker gain errors and missing information in the TFs. We discuss the results in Sec. 4 and summarize in Sec. 5.

1 METHODS

In this section we first introduce the framework for robust reproduction based on acoustic modeling. Then, two different robust control strategies (WCO and PMO) are applied to ACC as a typical application of this framework. Finally, a practical "quick" approach is proposed. The approach uses the estimated bounds of the errors to determine the control parameters for robust ACC.

1.1 Framework for Robust Sound Reproduction Based on Acoustic Modeling

The proposed framework consists of two distinct stages: *acoustic modeling* to obtain an acoustic TF model for a real system with uncertainty and *robust design* to derive source weights for robust control. Acoustic TFs from the sources to the control zones are modeled as source radiation paths with errors. As shown in Fig. 1, after setting the *application scenario* (including the system geometry and the acoustic environment), a set of spatial impulse responses is acquired by *transfer function measurement*. Then in the *sound propagation analysis* module, the sound field radiated by each source is modeled as the superposition of a determined part $\{\mathbf{G}\}$ and a set of potential error $\{\Delta\mathbf{G}\}$ that describes variations in the TFs for certain reproduction scenarios:

$$\tilde{\mathbf{G}} = \mathbf{G} + \Delta\mathbf{G}, \tag{1}$$

where $\tilde{\mathbf{G}}$, $\mathbf{G}$ and $\Delta\mathbf{G}$ are $M \times L$ TF matrices, with M control points in the target zone(s) and L loudspeakers.

To model the determined part, recent technologies for source radiation measurements and modeling can be considered, such as holographic nearfield measurements [24] or sound field separation [25]. In this paper we focus on the direct radiation paths and consider the room effects to be errors, although these could additionally be modeled.

The error term obtained in the *error estimation* module is a certain probability distribution function $p(*)$ or bound $b(*)$ describing the variances in TF amplitude and phase due to potential errors. The error term represents any source directivity components that cannot be fully represented with limited resolution measurements and also the spatial response variances resulting from non-ideal conditions, e.g., the room effect.

With the determined TF matrix $\mathbf{G}$ and the error descriptions $p(*)$ or $b(*)$, several error-model-based robust control strategies can be applied. In the following sub-sections, the solutions of robust ACC are derived separately for WCO and PMO.
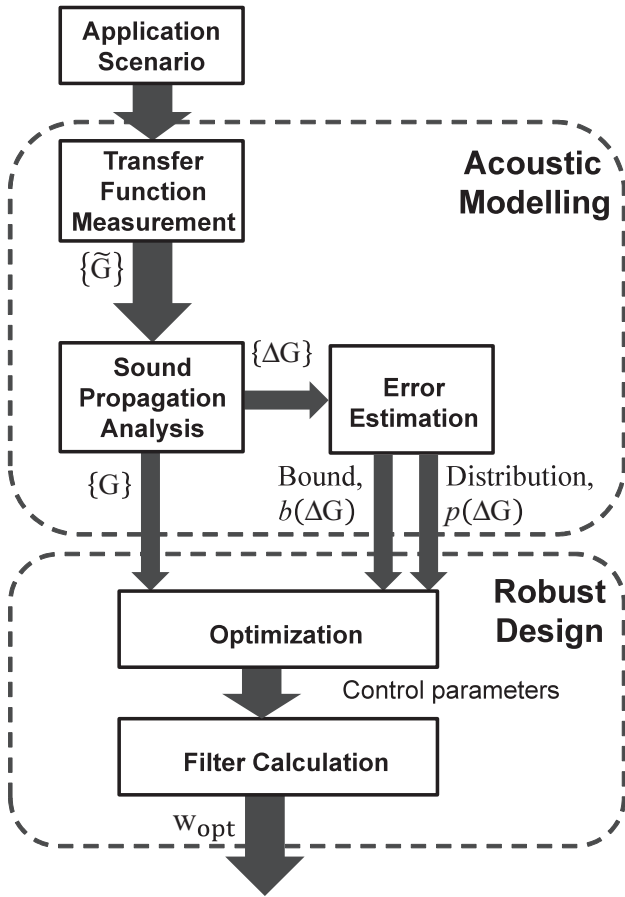
Fig. 1. Framework of acoustic-modeling based robust sound reproduction.

1.2 Acoustic Contrast Control

ACC maximizes the ratio of the spatially averaged sound energy between the listening zone and the quiet zone [3]:

$$\max \frac{\mathbf{w}^H \mathbf{R}_L \mathbf{w}}{\mathbf{w}^H \mathbf{R}_Q \mathbf{w}}, \tag{2}$$

where $\mathbf{w}$ is the $L \times 1$ loudspeaker weight vector to be obtained, and $\mathbf{R}_L = \mathbf{G}_L^H \mathbf{G}_L$ and $\mathbf{R}_Q = \mathbf{G}_Q^H \mathbf{G}_Q$ are the spatial correlation matrices defined by TF matrices $\mathbf{G}_L$ and $\mathbf{G}_Q$ for the listening zone and the quiet zone, respectively.

1.3 Worst-Case Optimization

In order to improve the overall performance of ACC applied to all the predictable situations, WCO [20, 23] aims to find and optimize the situation giving the worst performance among all the possible variations. Considering uncertainties in the TFs, the quiet zone spatial correlation matrix is written as $\tilde{\mathbf{R}}_Q = \mathbf{R}_Q + \Delta \mathbf{R}_Q$, where $\Delta \mathbf{R}_Q$ represents the perturbations in $\tilde{\mathbf{R}}_Q$. Assuming $\Delta \mathbf{R}_Q$ has a norm bound δ_Q, the optimization problem is to maximize the worst AC under all potential $\Delta \mathbf{R}_Q$,

$$\max_{\mathbf{w}} \min_{\Delta \mathbf{R}_Q} \frac{\mathbf{w}^H \mathbf{R}_L \mathbf{w}}{\mathbf{w}^H \tilde{\mathbf{R}}_Q \mathbf{w}} \quad s.t. \left\| \Delta \mathbf{R}_Q \right\|_F \le \delta_Q, \tag{3}$$

where $\|*\|_F$ denotes Frobenius norm. The solution is [15]:

$$\mathbf{w}_{WC} = \Phi_1 \left\{ (\mathbf{R}_Q + \delta_Q \mathbf{I})^{-1} \mathbf{R}_L \right\}, \tag{4}$$

where $\Phi_1(*)$ denotes the principal eigenvector, corresponding to the maximum eigenvalue and $\mathbf{I}$ is a $L \times L$ identity matrix. Eq. (4) has a standard regularization form that may be found by applying an energy constraint [8, 26]. However, the diagonal loading parameter δ_Q is derived directly from the error bound $\left\| \Delta \mathbf{R}_Q \right\|_F \le \delta_Q$, rather than from white noise gain or AE constraints. As mentioned in [23], it might also be useful to apply a weighting to the estimated error bound of $\left\| \Delta \mathbf{R}_Q \right\|_F$ to avoid "over-robust" solutions that maintain robustness but lead to poor AC performance. Errors in the listening zone TFs could be incorporated into the WCO, replacing $\mathbf{R}_L$ in Eq. (4) with $(\mathbf{R}_L - \delta_L \mathbf{I})$ [15]. However, our previous work [23, Fig. 4] showed that applying a diagonal loading parameter δ_L to $\mathbf{R}_L$ did not obviously increase robust performance. In general, if the ratio δ_L / δ_Q is much smaller than the potential AC, the effect of diagonal loading on $\mathbf{R}_L$ can be ignored.

1.4 Probability-Model Optimization

Assuming the errors in the acoustic TFs have a predictable distribution, PMO aims to improve the average performance according to that distribution [16]. Here we consider two kinds of error: multiplicative error and additive error.

1.4.1 Multiplicative Error PMO

Following [17], multiplicative errors can be incorporated into PMO for ACC. The form of the TF with multiplicative error is $\tilde{\mathbf{G}}_{(m,l)} = \mathbf{G}_{(m,l)} a e^{j\phi}$, where $j = \sqrt{-1}$, $\mathbf{G}_{(m,l)}$ denotes the TF between the lth loudspeaker and mth control point in a zone, and a and ϕ are the amplitude and phase of multiplicative TF errors. The statistical features of the TF errors are

$$\begin{aligned} \sigma_a &= \int_a a^2 \, p(a) \, da, \\ \mu_a &= \int_a a \, p(a) \, da, \\ \sigma_\phi &= (\int_\phi \cos\phi \, p(\phi) \, d\phi)^2 + (\int_\phi \sin\phi \, p(\phi) \, d\phi)^2, \end{aligned} \tag{5}$$

and $p(a)$ and $p(\phi)$ are the probability density functions describing the amplitude and phase errors, respectively. $p(a)$ and $p(\phi)$ can be obtained by acoustic modeling. Assuming errors in each TF are independent and identically distributed, the average spatial correlation matrix for the quiet zone is

$$\mathbf{R}_{Q,PMO} = \mathbf{R}_Q \otimes \mathbf{E}_Q, \tag{6}$$

where $\otimes$ is pointwise multiplication, and $\mathbf{E}_Q$ is a $L \times L$ matrix with diagonal elements equal to σ_a and non-diagonal elements equal to $\mu_a^2 \sigma_\phi$. The average spatial correlation matrix $\mathbf{R}_{L,\,PMO}$ for the listening zone is obtained in the same way. ACC with PMO for multiplicative error is

$$\max_{\mathbf{w}} \frac{\mathbf{w}^H (\mathbf{R}_L \otimes \mathbf{E}_L) \mathbf{w}}{\mathbf{w}^H (\mathbf{R}_Q \otimes \mathbf{E}_Q) \mathbf{w}}, \tag{7}$$

that has the solution

$$\mathbf{w}_{ME} = \Phi_1 \{ (\mathbf{R}_Q \otimes \mathbf{E}_Q)^{-1} (\mathbf{R}_L \otimes \mathbf{E}_L) \}. \tag{8}$$

A special case assumes that the amplitudes of the error a are uniformly distributed over the range $[a_{min}, a_{max}]$ and the phase error is uniformly distributed over the range $[\phi_{min}, \phi_{max}]$. In this case, the entries of $\mathbf{E}_Q$ and $\mathbf{E}_L$ are calculated as

$$\begin{aligned}
\sigma_a &= (a_{min}^2 + a_{min}a_{max} + a_{max}^2)/3, \\
\mu_a &= (a_{min} + a_{max})/2, \\
\sigma_\phi &= [2 - 2\cos(\phi_{max} - \phi_{min})]/(\phi_{max} - \phi_{min})^2.
\end{aligned} \quad (9)$$

Therefore, PMO introduces regularization, which does not directly have diagonal loading form, into the spatial correlation matrices for both zones.

1.4.2 Additive Error PMO

Following [18], additive errors can also be incorporated into PMO for ACC. The form of the TF with additive error is $\tilde{\mathbf{G}}_{(m,l)} = \mathbf{G}_{(m,l)} + ae^{j\phi}$, where a and ϕ are now the amplitude and phase of the additive TF errors. The statistical features of the additive TF errors are σ_a, μ_a, σ_ϕ and $\mu_\phi = \int_\phi e^{-j\phi} p(\phi)\,d\phi$. We define $\mathbf{E}_{1,Q}$ as a $L \times L$ matrix with elements in the lth column equal to $\sum_{i=1}^M \mathbf{G}_Q^{i,l}$, i.e., the sum of TFs between the lth loudspeaker and the M control points in the quiet zone. We further define $\mathbf{E}_{2,Q}$ as a $L \times L$ matrix with diagonal elements equal to σ_a and non-diagonal elements equal to $\mu_a^2 \sigma_\phi$. The average spatial correlation matrix of the quiet zone is

$$\mathbf{R}_{Q,PMO} = \mathbf{R}_Q + \mu_a\mu_\phi\mathbf{E}_{1,Q} + \mu_a\mu_\phi^*\mathbf{E}_{1,Q}^H + M\mathbf{E}_{2,Q}, \quad (10)$$

where μ_ϕ^* is the complex conjugate of μ_ϕ. The solution is

$$\mathbf{w}_{AE} = \Phi_1\{(\mathbf{R}_Q + \mathbf{E}_{3,Q})^{-1}(\mathbf{R}_L + \mathbf{E}_{3,L})\}, \quad (11)$$

where $\mathbf{E}_{3,Q} = \mu_a\mu_\phi\mathbf{E}_{1,Q} + \mu_a\mu_\phi^*\mathbf{E}_{1,Q}^H + M\mathbf{E}_{2,Q}$ and $\mathbf{E}_{3,L} = \mu_a\mu_\phi\mathbf{E}_{1,L} + \mu_a\mu_\phi^*\mathbf{E}_{1,L}^H + M\mathbf{E}_{2,L}$.

As before, a special case assumes uniform error distribution with amplitude a uniformly distributed in the range $[0, a_{max}]$ and phase uniformly distributed in the range $[0, 2\pi]$. In this case,

$$\mathbf{w}_{AE} = \Phi_1\left\{\left(\mathbf{R}_Q + \frac{Ma_{max}^2}{3}\mathbf{I}\right)^{-1}\left(\mathbf{R}_L + \frac{Ma_{max}^2}{3}\mathbf{I}\right)\right\}. \quad (12)$$

In this special case, PMO has a diagonal loading form acting on the spatial correlation matrices for both zones.

1.5 Quick Parameter Estimation for Robust ACC

In the following, we refer to errors occurring in the system as the *actual errors*, while the *assumed errors* are the estimated errors used for filter calculations.

In the WCO approach, the robust parameter δ_Q (Eq. (4)) used to calculate $\mathbf{w}_{WC}$ is frequency-varying and determined by the bound of errors in the spatial correlation matrix,

$$\delta_Q = b(\|\Delta\mathbf{R}_Q\|_F)/\varepsilon, \quad (13)$$

where ε is a scale parameter to prevent "over-robust" performance in WCO. If the assumed error has the form of multiplicative error, then

$$\begin{aligned}
b(\|\Delta\mathbf{R}_Q\|_F) &\approx [\|(a_{max,ME}\mathbf{G}_Q)^H(a_{max,ME}\mathbf{G}_Q) - \mathbf{R}_Q\|_F/2 \\
&+ \|(a_{min,ME}\mathbf{G}_Q)^H(a_{min,ME}\mathbf{G}_Q) - \mathbf{R}_Q\|_F/2] \\
&= (|a_{max,ME}^2 - 1| + |a_{min,ME}^2 - 1|)\|\mathbf{R}_Q\|_F/2, \quad (14)
\end{aligned}$$

where $a_{max,ME}$ and $a_{min,ME}$ are the maximum and minimum error gains, respectively. If the assumed error has the form of additive error, then

$$\begin{aligned}
b(\|\Delta\mathbf{R}_Q\|_F) &\approx \|(\mathbf{G}_Q + a_{max,AE})^H(\mathbf{G}_Q + a_{max,AE}) \\
&- \mathbf{R}_Q\|_F, \quad (15)
\end{aligned}$$

where $a_{max,AE}$ is the maximum error amplitude. The parameters $a_{max,ME}$, $a_{min,ME}$, $a_{max,AE}$ are determined by $b(\Delta\mathbf{G})$, the bound of errors in TFs. Eqs. (14) and (15) lead to a quick parameter estimation in WCO incorporating prior knowledge, filling in the gap between the outcomes of acoustic modeling ($\mathbf{G}_Q$ and $b(\Delta\mathbf{G})$), and directly lead to δ_Q (Eq. (13)), which in turn populates the WCO solution (Eq. (4)).

Furthermore, though PMO is an error distribution based solution, it can also be formulated based on TF error bounds by selectively sacrificing accuracy in the error distribution estimation. Since it is difficult to accurately estimate the actual error, a practical "quick" approach is to use a simple error distribution (e.g., Eqs. (9) and (12) assuming uniform distribution) to approximate the actual error. This only requires the TF error bound parameters (a_{max}, a_{min}, ϕ_{max}, ϕ_{min}) to calculate loudspeaker weights. These TF error bound parameters have a clear physical meaning and a relationship to the actual error.

These "quick" approaches employing coarse error estimation (that does not fully describe the actual errors) are included in the following simulation and experimental results.

2 SIMULATIONS

In this section we present simulation results comparing the performance of robust control approaches in the presence of TF errors.

2.1 Simulation Setup and Baseline Geometry

As shown in Fig. 2, we use an arc-shaped array with 11 loudspeakers. The loudspeakers are uniformly arranged with a distribution angle of 6° around a radius of 1.68 m. The control points for the listening and quiet zones are defined on dual circles with 24 microphones in each ring, with radii of 0.083 m and 0.104 m. In the simulations, we suppose each loudspeaker acts as a point source. That is, the spatial response is defined by e^{jkd}/kd, where k is the wave number and d is the distance between a loudspeaker and a receiver. Errors in the system are simulated by perturbing the spatial responses, assuming that the error has multiplicative form. In the Monte-Carlo trials, the amplitude errors are drawn from a Gaussian distribution with
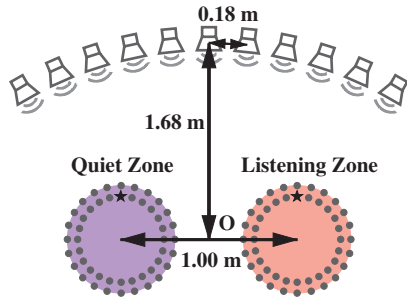
Fig. 2. Simulation geometry with 11 loudspeakers oriented towards the origin in an arc array and 48 monitor microphones (reference microphone denoted by star, other microphones denoted by dot) in each zone. The listening zone and the quiet zone are symmetric with respect to the loudspeaker array.

standard deviation of 3 dB, and the phase errors are drawn from a uniform distribution between $-10°$ and $+10°$.

2.2 Diagonal Loading

The solutions WCO (Eq. (4)) and PMO for assumed special additive errors (Eq. (12)), derived above have a similar form to common regularization approaches, which employ diagonal loading and add a small parameter to the diagonal of $\mathbf{R}_Q$. Since the diagonal loading form is common among many robust methods, we first directly study the relationship between the value of the diagonal loading parameter applied and the average AC performance, using Monte-Carlo simulations [14] with 1000 trials for each frequency sample. The AC evaluation is defined for a single frequency as

$$\text{AC} = 10 \log_{10} \frac{\mathbf{p}_L^H \mathbf{p}_L}{\mathbf{p}_Q^H \mathbf{p}_Q}, \quad (16)$$

which is the ratio of the sound energy in the listening zone and the quiet zone, with $\mathbf{p}_L = \mathbf{G}_L \mathbf{w}$ and $\mathbf{p}_Q = \mathbf{G}_Q \mathbf{w}$. The diagonal loading parameter δ_Q was varied from 10^{-15} to 10^5 at 1000 logarithmically spaced values, and the resulting AC, averaged over all trials for the error conditions described above, is shown in Fig. 3. Diagonal loading regularization on $\mathbf{R}_Q$ is seen to be effective in improving robustness with a suitable selection of the parameter. A certain pattern of average AC performance under varying δ_Q is observed, which is composed of three stages. If the regularization parameter is too small, the AC performance is reduced ("under-robust"); with the regularization parameter in the proper range, the AC performance stabilizes ("robust") and there is a peak that corresponds to the best regularization parameter (denoted by the solid line in Fig. 3); after that there is a sharp drop in AC and the trend eventually becomes stable (with minimum AE), where the performance is very robust but the AC achieved is relatively low ("over-robust").

The AC performance within 1 dB of the peak is also illustrated in Fig. 3 by the contoured dashed lines. The system has different sensitivity to δ_Q over different frequency ranges. The "middle" frequencies, approx. 740–2600 Hz and 4500–5200 Hz have relatively broad acceptable δ_Q range, while at low frequencies under 430 Hz and spe-
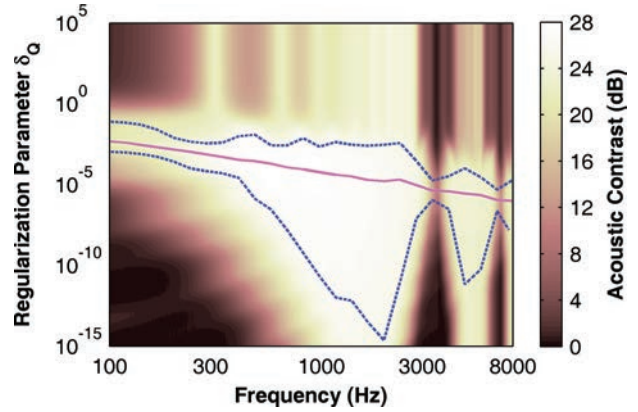

Fig. 3. Simulated AC performance with varying regularization parameter δ_Q on $\mathbf{R}_Q$, by Monte-Carlo simulation over 1000 trials for each frequency sample between 100–8000 Hz. Solid line denotes the peak value of the mean AC performance and dashed contours show peak AC –1 dB.

cial frequencies such as 2900–4100 Hz and 6300–7700 Hz, where grating lobes exist, the suitable δ_Q range is quite narrow, leading to sensitive performance.

It can be observed that the optimal δ_Q is proportional to $1/k^2$ for most frequencies, since the optimal regularization value is related to the interaction between the system geometry and the errors. As equal amplitude multiplicative errors are assumed in the whole frequency range for the simulated system geometry, δ_Q largely follows the square of the adopted source radiation gain $1/kd$. However, selecting δ_Q is more complex in practice, where the source radiation pattern may not be as simple as a point source, and error will also vary with both frequency and space (e.g., due to room reflections). Incorporating error information into control could therefore give a δ_Q close to the AC peak. We explore this in the following section through comparison with alternative methods to calculate a diagonal loading parameter.

2.3 Estimation of Regularization Parameters

In this section regularization parameter calculation is described for various robust control strategies, including acoustic-modeling based methods and frequency-dependent diagonal loading methods; and the resultant performance is compared. The peak AC at each frequency (i.e., the value corresponding to the solid line in Fig. 2) achieved by Monte-Carlo simulation with full knowledge of error is used as a reference denoted as **MCS**.

2.3.1 Parameter Calculation

The methods under test, and their corresponding parameters, are defined in this section. Three acoustic-modeling based methods were tested:

WCO represents WCO substituting the assumed error bounds ($a_{\text{max,ME}} = \sqrt{2}$ and $a_{\text{min,ME}} = \sqrt{2}/2$) into Eqs. (13) and (14), $\delta_Q = 0.75 \|\mathbf{R}_Q\| /\varepsilon$, with the scale parameter $\varepsilon = 10^2$.

PMO$_M$ represents PMO for assumed multiplicative errors, assuming the special case of uniform error distribution (Eqs. (8) and (9)). Though the error distribution features are different between the actual error and the assumed error, we directly apply the minimum and maximum values of the error gain and phase as a_{min}, a_{max}, ϕ_{min}, and ϕ_{max} in Eq. (9). So, $a_{max} = \sqrt{2}$, $a_{min} = \sqrt{2}/2$, and the phase parameters take the bound values $\phi_{max} = 10°$, $\phi_{min} = -10°$.

PMO$_A$ represents PMO for assumed additive errors assuming the special case (Eq. (12)). The error distribution features are also different between the actual error and the assumed error. As the actual error in the simulation is multiplicative, PMO$_A$ here also represents a situation where the error type is wrongly estimated. The maximum value of additive error gain in Eq. (12) among all the potential TF variances is $a_{max, AE}$. To make the estimated additive error set contain the real multiplicative error set, $a_{max,AE} = \max\{\mathbf{G}\} \times a_{max,ME}$, where $\max\{\mathbf{G}\}$ is the maximum among all the TFs in $\mathbf{G}$, and $a_{max,ME} = \sqrt{\mu^2 - 2\cos(\phi_{max})\mu + 1}$ with $\mu = 10^{\frac{3}{20}}$.

For the three acoustic-modeling based methods above, the assumed error models are not exactly the same as the actual error model in simulation. Nevertheless, this is indicative of the practical situation when applying error-based regularization.

In addition to these methods, three frequency-dependent regularization methods were tested:

SV denotes regularization based on σ_{max}, the maximal singular value of $\mathbf{R}_Q$. It was noted in the discussion of Fig. 3 that over-robust performance occurs after δ_Q exceeded σ_{max}. Therefore, we set $\delta_Q = \sigma_{max}/10$. This method is also adopted in, e.g., [27].

EL0 and **ELM** represent the approaches based on AE control. The AE [8] is defined as:

$$AE = 10 \log_{10}\left[\frac{(\mathbf{w}^H\mathbf{w})(\mathbf{p}_{L,0}^H\mathbf{p}_{L,0})}{\mathbf{p}_L^H\mathbf{p}_L}\right], \tag{17}$$

where $\mathbf{p}_{L,0} = \mathbf{G}_L\mathbf{w}_0$, in which $\mathbf{w}_0 = [0\,0\,0\,0\,0\,1\,0\,0\,0\,0]$ is the weighting vector for a reference source at the center of the array. EL0 and ELM limit AE towards 0 dB and the minimal value, respectively. EL0 was chosen to match previous studies (e.g., [19]). SV, EL0, and ELM represent the situations where the geometry is known but the error features are not.

2.3.2 Performance Comparison

In Fig. 4 we plot the mean AC and AE performance over frequency (based on Monte-Carlo simulations, with the same error conditions as above), for no regularization (NR), alongside SV, EL0, ELM, and PMO$_A$, with MCS presented as an upper reference. The parameters estimated for WCO and PMO$_M$ gave very similar results to PMO$_A$ (except that PMO$_M$ was slightly better at low frequencies) and are therefore omitted from the figure for clarity.

In Fig. 4 all robust approaches improve performance over NR. Among them, PMO$_A$ has the closest performance to the optimal MCS performance, both in terms of AC and AE, over a broad frequency range 100–8000 Hz, with lim-
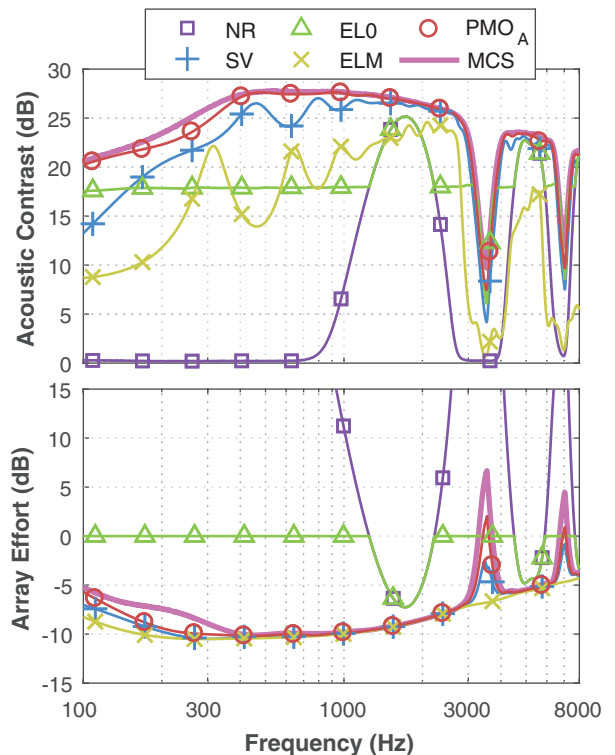

Fig. 4. Simulated performance of NR (no regularization), SV (maximum singular-value-based regularization), EL0 (AE limited to 0 dB), ELM (AE minimized), PMO$_A$ (PMO for additive error with quick parameter estimation), and MCS (optimal regularization from Fig. 3), using Monte-Carlo Simulation over 10000 trials.

ited AC degradation, mainly at low frequencies. According to the sensitivity concerns discussed in Sec. 2.2, we chose three typical frequencies, namely a low frequency (200 Hz), a middle frequency (1000 Hz), and a grating lobe frequency (3538 Hz), to make a detailed comparison between the different regularization methods. The performance of each approach at these frequencies is listed in Table 1. To aid the interpretation of Table 1, the methods are ranked from "under-robust" (left) to "over-robust" (right). Metrics "mean AC" and "min AC" (denoted by $\overline{AC}$ and $\wedge$ AC, respectively) are the average and minimum AC performance over 10000 Monte-Carlo trials. Metric "Δ AC" is the difference between "mean AC" and "min AC," i.e., the difference between average performance and worst performance over the 10000 trials. The regularization parameter "δ_Q" adopted in each method, which is roughly inversely proportional to "AE," is also shown. It can be noted that, among all the methods, PMO$_A$ achieved the best performance for both $\overline{AC}$ and $\wedge$ AC and has the median "Δ AC" value among all the methods. Comparing δ_Q values with reference to Fig. 3, PMO$_A$ nearly reaches the optimal regularization parameter, while EL0 is a little "under-robust" and SV is a little "over-robust." Since "over-robust," SV and ELM have smaller "Δ AC" than PMO$_A$, with correspondingly weaker $\overline{AC}$ and $\wedge$ AC performance. Table 1 and Fig. 4 also demonstrate that very close AE values can give very different AC performance, for example comparing SV and PMO$_A$ at 1000 Hz.

Table 1. Simulated performance of various robust strategies at 200, 1000, and 3538 Hz. The mean/minimal AC ($\overline{AC}$/$\wedge$ AC) and their difference (Δ AC) over 10000 Monte-Carlo trials, array effort (AE), and regularization parameter (δ_Q) are shown. The methods are ordered in terms of their performance, from "under-robust" (left) to "over-robust" (right).

Freq. (Hz)	Eval.	NR	EL0	PMO_A	SV	ELM
200	$\overline{AC}$	0.2	17.8	**22.3**	20.4	11.8
	$\wedge$ AC	−3.5	15.3	**20.5**	19.2	11.3
	Δ AC	3.7	2.5	1.8	1.2	0.5
	AE	99.4	0.0	−9.4	−9.7	−10.3
	δ_Q	0.0	1.5e-5	1.0e-2	1.0e-1	1.0*
1000	$\overline{AC}$	7.7	17.9	**27.5**	26.1	22.0
	$\wedge$ AC	4.4	15.1	**25.1**	23.9	20.3
	Δ AC	3.3	2.8	2.4	2.2	1.7
	AE	11.0	0.0	−9.9	−9.9	−10.0
	δ_Q	0.0	5.5e-17	4.1e-4	3.6e-3	1.0*
3538	$\overline{AC}$	0.3	6.0	**7.6**	4.3	0.9
	$\wedge$ AC	−3.8	5.1	**6.3**	3.7	0.6
	Δ AC	4.1	0.9	1.3	0.6	0.3
	AE	56.3	0.0	2.0	−3.3	−6.8
	δ_Q	0.0	4.8e-5	3.3e-5	1.4e-4	1.0*

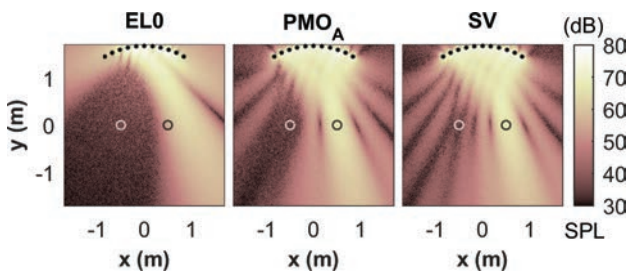*: Array effort method picks δ between 10^{-20} and 10^0, so ELM picks 1.0.

Fig. 5. Simulated sound pressure level of ACC sound field at 1 kHz with EL0 (AE limited to 0 dB), PMO_A (PMO for additive error with quick parameter estimation), and SV (maximum singular-value-based regularization). The illustrated sound field corresponds to a single Monte-Carlo simulation trial.

This emphasizes the difficulty in achieving optimal robust control directly from a certain AE limit.

The reproduced sound fields using EL0, PMO_A, and SV with normalized loudspeaker weights at 1000 Hz are illustrated in Fig. 5. These represent a visualization of a random trial from the 10000 Monte-Carlo simulations. EL0 (representing "under-robust") maintains a deep quiet zone yet does not ensure that the main lobe is transmitted directly to the required direction. On the other hand, SV and PMO_A create shallower quiet zones, yet they manipulate the array to emit a relatively sharp main lobe towards the listening zone. While SV (representing "over-robust") is affected by energy lobes across the quiet zone, PMO_A still creates sufficient cancellation and is a good compromise between quiet zone cancellation and efficient listening zone focusing under the error conditions.

2.4 Summary

Acoustic-modeling based solutions WCO, PMO_A, and PMO_M (representing WCO and PMO with assumed special error models) are compared with AE control and singular-value-based regularization by simulations. Since PMO_A and WCO share a diagonal loading form with regularization methods, we first observed the robust performance by varying the regularization parameter under a certain error through Monte-Carlo simulations. The pattern shows the system's sensitivity to the regularization parameter and that good parameter estimation leads to robust AC and AE performance. By incorporating coarse error information in addition to geometry information, PMO and WCO derived diagonal loading parameters with clear physical meaning and reduced computational complexity compared to the AE-based method.

3 EXPERIMENTAL VALIDATION

Experimental results are presented in this section to validate the simulation results in a real-world reproduction system and investigate the feasibility of using acoustic-modeling based robust control to reduce the need for in-situ measurements.

The experimental reproduction and performance measurement system is introduced in Sec. 3.1. Three TF models and error estimation parameters for populating various filters are elaborated in Sec. 3.2, and the error conditions are described. Finally, three main observations are made from the experimental performance measurements and are presented in Secs. 3.3–3.5. Observation 1 concerns the comparison of different robust strategies against a certain error; Observation 2 explores the potential of using acoustic-modeling based robust control to compensate TF mismatches arising from reduced in-situ measurements; Observation 3 further compares the solution with reduced in-situ measurements and the solution with full in-situ measurements with different levels of error.

3.1 Experimental Setup

For comparison with the simulation study, the same loudspeaker array and zone geometry was adopted in the experiments (as shown in Fig. 2). The 11 loudspeakers (Genelec 8020b) were mounted on an arc truncated from a 60-element circular array (see [6]). The 48 monitor microphones (Countryman B3 omni) were assembled in a bi-circular array and used to measure each zone at the listening plane. The "playrec" utility in Matlab was used to play and record sound with the loudspeaker 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. The whole system was placed in an acoustically treated recording studio [6] with size 6.55 ×
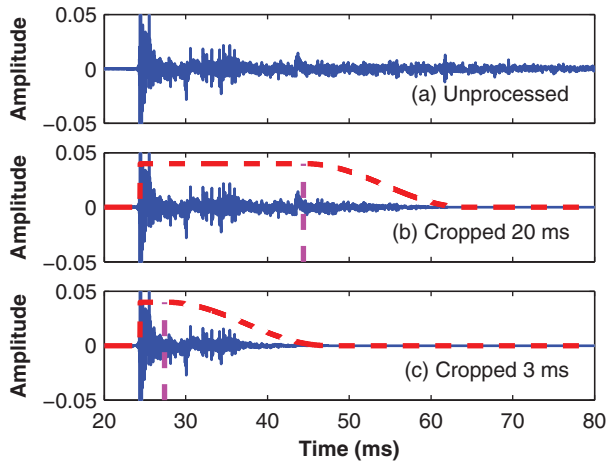
Fig. 6. Measured RIR: (a) unprocessed; (b) cropped 20 ms; (c) cropped 3 ms. The length shape of the cropping windows, including the crop limit and a 15 ms raised-cosine taper, are illustrated by the dashed lines. The maximal amplitude value is 0.30.

8.78×4.02 m (RT60 255 ms averaged over 0.5, 1, and 2 kHz octave bands[1]).

To achieve the goal of rendering different audio content to two listeners, one in each sound zone, two target situations were measured: **Target A** – the left zone as the quiet zone and the right zone as the listening zone; **Target B** – the left zone as the listening zone and the right zone as the quiet zone.

Though a symmetric geometry setup was employed, the performance was not expected to be identically symmetric, being influenced by the asymmetric room effect, source position mismatches, source sensitivity inconsistencies, and so on.

3.2 Transfer Function Modeling and Filter Calculation with Error Estimation

The measurement and pre-processing of TFs between loudspeakers and zones are introduced in this section. The errors manually introduced in realization and the filter calculation are also described.

3.2.1 Transfer Function Measurements

Room impulse responses (RIRs) between each microphone position and each loudspeaker were measured using the maximum length sequence approach (15th order at 48 kHz). Cropping was applied to the measured RIRs to avoid the presence of unnecessary room effects in the TFs. The RIR from the center loudspeaker to the reference microphone (⋆, Fig. 2) in the listening zone is shown in Fig. 6, (a) without cropping, (b) cropped at 20 ms after the impulse onset, and (c) cropped at 3 ms after the impulse onset. The cropping envelopes, which include a raised cosine ramp at 15 ms duration after the crop time, are also shown. The

unprocessed RIR (Fig. 6(a)) contains the direct sound, reflections, and reverberant tail, illustrating the complexity of the TFs in this system realization, whereas the 20 ms cropped RIR (Fig. 6(b)) retains the direct sound and early reflections, and the 3 ms cropped RIR (Fig. 6(c)) retains the direct sound and the first early reflections.

Three TF models, implying different levels of in-situ measurement, are adopted in the loudspeaker weight calculation and are compared in Sec. 3.4 and Sec. 3.5.

RIR: All the TFs (between each loudspeaker and 48 control points in each zone) used for the filter design are measured in-situ and cropped at 20 ms (i.e., as in Fig. 6(b)).

RIR-PS: Only TFs between each loudspeaker and the two reference microphones are measured and cropped at 3 ms (i.e., as in Fig. 6(c)). The TFs to the other microphone positions are obtained by interpolation through the point source model, $H_{m,l} = H_{\text{ref},l} \times e^{-jk(d_{m,l}-d_{\text{ref},l})} \times d_{\text{ref},l}/d_{m,l}$, where $d_{\text{ref},l}$ and $d_{m,l}$ are the distances from the lth loudspeaker to the reference and mth monitor microphones, respectively. $d_{\text{ref},l}/d_{m,l}$ and $(d_{m,l} - d_{\text{ref},l})/c$ are assumed to be the gain and delay differences between the TF from the lth loudspeaker to the mth control point ($H_{m,l}$) and that to the reference control point ($H_{\text{ref},l}$) in a zone, and c is the speed of sound.

PS: All the TFs are calculated by the point source model $e^{-jkd_{m,l}}/d_{m,l}$, without any measurement.

The distances ($d_{\text{ref},l}$ and $d_{m,l}$) used in RIR-PS and PS are those according to the pre-defined geometry (in Sec. 3.1) rather than experimentally measured distances.

Compared with PS, RIR-PS contains the basic source equalization and partial loudspeaker directivity information. Compared with RIR, RIR-PS lacks detailed source directivity, source inconsistency, and room information but requires less measurements to be made. RIR-PS therefore represents the case where limited measurements are included in the source model. This approach can be compared to previous work that has used either full TF measurement (e.g., [6]) or purely a point-source model (e.g., [28]).

In practice, pseudo-anechoic responses from loudspeakers towards the reference microphone, similar to RIR-PS could potentially be obtained from the manufacturer. Furthermore, instead of the simple point source model, a source directivity model (e.g., [24]) could be measured and applied into the interpolation for better source modeling.

3.2.2 Errors in Realization

Two errors are considered in the experiments:

1) **Inconsistency error over the loudspeaker array**
   We investigate perturbation errors by artificially introducing source inconsistencies in our experiments. Three sets of multiplicative gains for the loudspeaker responses were applied to explore the performance under this systematic error. The gains for each loudspeaker channel were randomly generated under error bounds ±0.1, ±0.5, and ±1.0 and are shown in Fig. 7. Note that the gain sets are not symmetric, contributing to performance differences between the Target A and Target B cases.
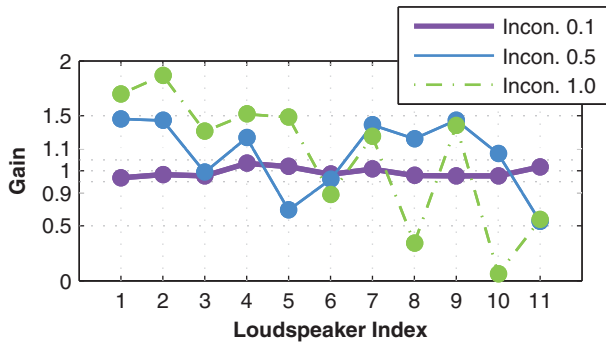
Fig. 7. Artificial inconsistency gain applied to each loudspeaker randomly generated by three levels of multiplicative error bounds (0.1, 0.5, 1.0).

2) Mismatch error in TF modeling

Reducing in-situ TF measurements can potentially reduce the amount of time taken to realize a sound zone system. However, mismatches might arise between the modeled TFs for filter design and the real TFs in reproduction.

The artificially applied inconsistency error is adopted as the main error in Observation 1 (Sec. 3.3). The mismatch error is considered as the main error in Observation 2 (Sec. 3.4). Both errors are considered together in Observation 3 (Sec. 3.5).

3.2.3 Error Estimation Parameters for Filter Calculation

We use [–3 dB, 3 dB] uniform multiplicative error (identical to the simulation study), and also a tightened [–1 dB, 1 dB] uniform multiplicative error, with [–10°, 10°] uniform phase error, as rough estimations of system errors (introduced in Sec. 3.2.2).

The quick estimation (e.g., Sec. 2.3) was again applied to obtain the control parameters for WCO, PMO_A, and PMO_M in experiments. With the three types of determined model (RIR, RIR-PS or PS) and the two levels of error estimation (evenly-distributed multiplicative error within ± 3 dB or ± 1 dB), six sets of WCO/PMO_A/PMO_M filters were prepared. Due to the similarity in measured AC performance between WCO, PMO_A, and PMO_M, which was also observed in simulations, WCO is omitted from the figures in the following observations for clarity.

3.3 Observation 1: Robust Strategies under Loudspeaker Inconsistency Error

This experimental observation corresponds to the simulation study in Sec. 2.3.2. TFs RIR were used for filter calculation. Errors were assumed to be uniform multiplicative errors with ± 1 dB error bounds and the set of loudspeaker gain inconsistencies with error bound 0.5 was applied. The methods under test were SV, EL0, PMO_A, PMO_M, and WCO. In the simulations in Sec. 2.3.2, the average/minimum AC performance was evaluated by Monte-Carlo simulation while the experimental observations are analogous to a single case study.
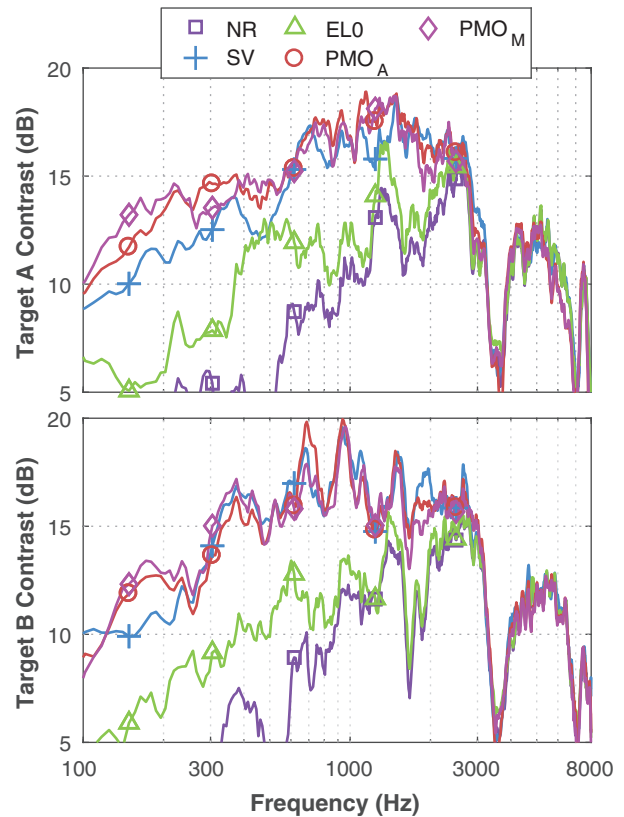

Fig. 8. Measured AC performance for Target A and Target B with NR (no regularization), SV (maximum singular-value-based regularization), EL0 (AE limited to 0 dB), PMO_A (PMO for additive error with quick parameter estimation), and PMO_M (PMO for multiplicative error with quick parameter estimation), with loudspeaker gain inconsistency error bound of 0.5 and RIR TFs.

Following Fig. 4, Fig. 8 shows the comparison of the five robust strategies. Among all the strategies, NR and EL0 were not effective at most frequencies; SV acted close to PMO_A; PMO_A and PMO_M performed well at most frequencies and had an obvious advantage at low frequencies for both Target A and Target B. PMO_M and PMO_A gave AC at or above 15 dB over 300–2800 Hz, and above 10 dB over 125–3000 Hz (for both targets), maintaining acceptable zone separation [29] in these ranges. It can be observed that even though a symmetric geometry was defined, the asymmetric errors led to differences in the performance for Target A and Target B.

Following Table 1, Table 2 shows $\overline{AC}$ and $\overline{AE}$ for Target A and Target B at three typical frequencies according to the robust performance pattern (cf., Fig. 3). At 200 Hz and 1000 Hz, PMO_M achieved the best $\overline{AC}$ and also low $\overline{AE}$; PMO_A and SV both performed much better than EL0 and NR. However, at the grating lobe frequency 3538 Hz, EL0 performs slightly better than the others. However, no method is effective for delivering sound zones at this frequency.

From the case study conducted with the measured RIRs against loudspeaker inconsistency errors generated from a partially known error set, PMO_A and PMO_M (representing acoustic-modeling based robust ACC) gave competitive performance.

Table 2. Measured AC and AE of various robust strategies, averaged over Target A and Target B, with loudspeaker inconsistency gain with multiplicative error bound of 0.5 using RIR TFs.

Freq. (Hz)	Eval.	NR	EL0	PMO_M	PMO_A	SV
200	$\overline{AC}$	0.8	6.2	**13.4**	12.9	10.8
	$\overline{AE}$	13.9	−0.3*	−4.9	−5.8	−7.5
1000	$\overline{AC}$	11.1	12.7	**17.5**	17.2	17.4
	$\overline{AE}$	−8.3	−8.3*	−12.0	−13.0	−13.7
3538	$\overline{AC}$	6.3	**7.0**	6.5	5.3	5.8
	$\overline{AE}$	8.8	1.8*	7.9	8.5	−1.0

*: EL0 method picks δ closest to 0 dB from 101 logarithmically spaced samples between 10^{-20} and 10^0.
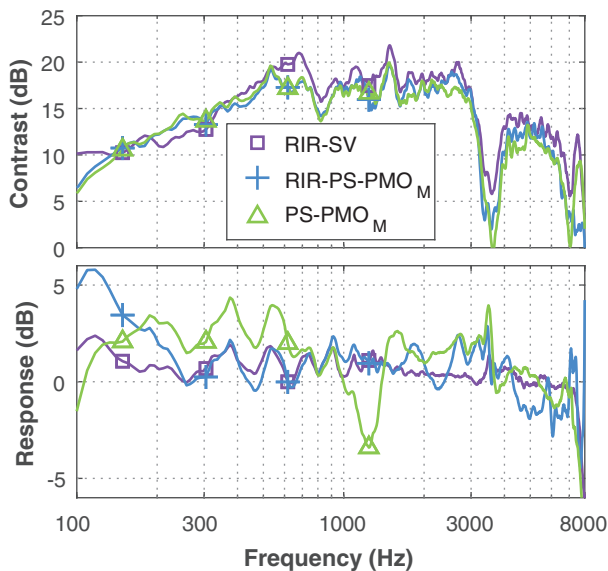

Fig. 9. Measured AC performance of RIR-SV (full in-situ measurement with SV regularization), RIR-PS-PMO$_M$ (single channel in-situ measurement with PMO$_M$), and PS-PMO$_M$ (point source model with PMO$_M$). AC and normalized listening zone FR are shown for Target A (without artificial gain errors applied).

3.4 Observation 2: Reducing in-situ Measurement

In this observation the three TF models, RIR, RIR-PS, and PS (introduced in Sec. 3.2), were individually adopted for filter calculation and their performance (with no additional inconsistency errors) is compared. Compared with RIR, RIR-PS, and PS lack accurate information about the TFs, which introduces a mismatch. We assume that the differences between RIR and RIR-PS (or PS) can be described by uniform multiplicative error within ±1 dB. The solutions found by applying PMO$_M$ to the RIR-PS and PS TFs (denoted RIR-PS-PMO$_M$ and PS-PMO$_M$), are compared with the case of fully measured TFs with maximum-singular-value dependent regularization (RIR-SV). The AC performance and the listening-zone-averaged frequency response (FR) for these three cases, reproduced for Target A, are shown in Fig. 9.

Considering AC, RIR-SV performed better than RIR-PS-PMO$_M$ and PS-PMO$_M$, except for some low frequen-

cies where it did not work well. However, the degradations between RIR-SV and RIR-PS-PMO$_M$/PS-PMO$_M$ are only 1.12/1.39 dB respectively, averaged over 100–3000 Hz.

The FR curves here are normalized by their average value over 100–8000 Hz, and (spatially) averaged over the 48 microphones in the listening zone. The flatness of FR is related to the reproduced sound quality. It can be quantified by measuring the frequency response variance, which is defined as

$$FRV = \sqrt{\frac{1}{N_f} \sum_{i=1}^{N_f} (p_i - \overline{p})^2}, \tag{18}$$

where N_f is the number of linear frequency samples, p_i is the spatially averaged listening-zone sound pressure at the ith frequency sample, and $\overline{p}$ is the sound pressure averaged over all the frequency samples. The FRV metric is included as informal listening revealed that non-flat frequency responses for some non-robust filters led to sound quality degradations. The FRVs of RIR-SV, RIR-PS-PMO$_M$, and PS-PMO$_M$ are respectively 1.22, 1.65, and 2.56 dB. The flatness of RIR-PS-PMO$_M$ is between that of RIR-SV and PS-PMO$_M$ because it contains information about the loudspeaker equalization. Taking one path measurement into account, as in RIR-PS, can therefore be expected to improve the reproduced sound quality.

The hybrid solution, using acoustic-modeling based robust control to reduce the need for in-situ measurement, could achieve satisfactory reproduction performance with a single measurement per loudspeaker, per zone. The RIR-PS-PMO$_M$ method achieved an AC performance above 15 dB over 320–3000 Hz (except a small drop at 800 Hz), and a close FRV performance with RIR-SV, especially over 250–1600 Hz, despite incorporating a simple acoustic model (point-source interpolation and quick error estimation).

Better acoustic modeling may further improve the AC and FRV performance for the reduced-measurement case RIR-PS-PMO$_M$, such as modeling the source radiation pattern. This should, for example, improve the match in FR between the RIR-SV and the reduced-measurement cases above 1600 Hz.

3.5 Observation 3: Robustness against Different Levels of Inconsistency Error

Two practical robustness problems are investigated further in this section: regularized filter performance with various levels of error and the sensitivity of the acoustic-modeling based methods to wrongly-estimated error bounds. Three different levels of loudspeaker gain inconsistency error (Sec. 3.2.2, Fig. 7) were tested. The AC performance of RIR-SV (upper) and RIR-PS-PMO$_M$ (with ±1 dB error bound estimation, lower), averaged across Target A and Target B, is shown in Fig. 10. The corresponding AC performance at three typical frequencies is listed in Table 3, which also includes RIR-PS-PMO$_M$ with ±3 dB error bound estimation. As the loudspeaker gain error increased, the overall performance of both RIR-SV and RIR-PS-PMO$_M$ dropped. However, the AC performance of both methods was above 10 dB (over 300–3000 Hz)

Table 3. Measured RIR-SV and RIR-PS-PMO$_M$ performance (AC averaged over Target A and Target B) for loudspeaker gain inconsistency with multiplicative error bounds of 0.1, 0.5, and 1.0. PMO$_M$1 dB, PMO$_M$3 dB denote PMO$_M$ with 1 dB or 3 dB error bound estimation.

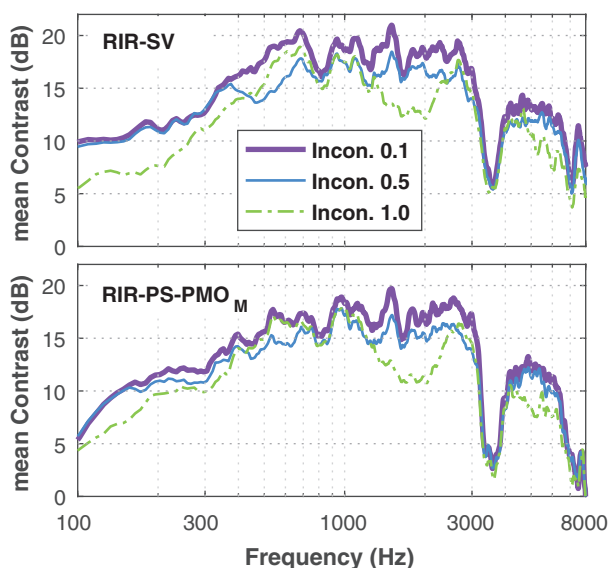| Freq. (Hz)	Incon.	RIR-SV	RIR-PS-	
			PMO$_M$ 1 dB	PMO$_M$ 3 dB
200	0.1	11.03	**11.55**	11.48
	0.5	10.79	10.97	**11.47**
	1.0	7.60	**9.72**	9.00
1000	0.1	**18.55**	18.29	18.23
	0.5	**17.43**	17.13	16.93
	1.0	**17.45**	17.39	17.32
3538	0.1	**5.94**	3.45	4.20
	0.5	**5.76**	3.56	4.26
	1.0	**5.44**	3.03	3.70

Fig. 10. Robustness against artificial gain inconsistencies with multiplicative error bounds of 0.1, 0.5, 1.0. Mean AC (Target A and Target B) is shown for RIR-SV and RIR-PS-PMO$_M$ with 1 dB error bound estimation.

for all three inconsistency levels. From Table. 3, RIR-PS-PMO$_M$ performed slightly worse than RIR-SV at 1000 Hz but slightly better at 200 Hz. At the grating lobe frequency (3538 Hz) neither approach achieved good AC, though RIR-SV was better. Overall, the reduced measurement case RIR-PS-PMO$_M$ remained competitive with RIR-SV as the reproduced error increased.

Furthermore, the two RIR-PS-PMO$_M$ filters with different assumed error bounds (± 1 dB or ± 3 dB) performed closely. That is, the control parameters selected by error-based methods (e.g., PMO$_M$) were quite robust to wrongly estimated error bounds, and the actual error in reproduction had a larger effect than the estimated error bound under the system we tested. For example, at 1000 Hz the difference in the inconsistency error bounds (± 0.1 and ± 0.5) led to a difference of more than 1.1 dB in AC performance, while the difference in estimated error bound (± 1 dB $\approx \pm 0.1$ and ± 3 dB $\approx \pm 0.4$) led to a performance difference within 0.2

dB. This experimental observation is in accordance with our previous simulations on error estimation mismatch in [23].

4 DISCUSSION

Overall, incorporating increasing amounts of prior knowledge into the system optimization allows for more robust control. With prior knowledge of the TFs, which encapsulate the system geometry and acoustic environment, we can aim to find a frequency-dependent regularization parameter. For some kinds of error, AE-based methods might be appropriate, however, the criterion for setting the AE constraint is not very clear and might make the performance "under-robust" or "over-robust" in reality. From Table 1, for example, a small difference in AE results in a quite different δ_Q. The singular-value based method (SV) also requires a threshold to be selected.

Additionally, limited knowledge about a potential error model can be incorporated by focusing on the bounds of the error, rather than the detailed error distribution (e.g., PMO$_A$, PMO$_M$, WCO). In simulation, the "quick" robust methods obtained good AC and AE performance, with reduced computational complexity compared to a Monte-Carlo simulation-based reference method [14]. More accurate acoustic modeling with WCO and PMO, might give better performance than the quick methods. However, from Figs. 3 and 4, it seems that any improvements might not be very remarkable compared with the increasing difficulty in error model estimation and complexity in calculation. In this case, the proposed "quick" methods offer a compelling means to calculate the regularization term. Furthermore, the proposed methods directly use physical acoustic modeling with a clear optimization goal (WCO or PMO) to choose the regularization parameter.

In practice, the error bound information required in the "quick" approaches can be obtained from both measurement and simulation, depending on the application scenario. For example, the bound value for the loudspeaker inconsistency in Observation 1 could be estimated by measuring a set of loudspeakers of the same model or directly provided by the loudspeaker manufacturer. Other error bounds could be obtained by simulation, for example by applying position mismatches.

With a full error analysis or extensive measurements, there is potential to increase the robustness of AC performance, mainly at low frequencies. However, the maximal performance loss (at low frequencies) resulting from "quick" parameter estimation was only 1.5 dB for simulations of the application geometry studied in our paper.

As an application of acoustic-modeling based robust control, we proposed a novel method to narrow the gap between no measurement and full measurement of TFs as input to the optimization. We measure a single TF per loudspeaker in each zone, interpolate the sound pressure to other microphone positions, and compensate the mismatch by acoustic-modeling based robust control. In this way the TF measurement was reduced from 11 × 2 × 48 RIRs to 11 × 2 RIRs. Experimental results revealed a trade-off between reducing measurements, which saves time and simplifies

the equipment requirements, and the limited sacrifice in performance. Opportunities remain to improve the accuracy of the acoustic modeling by considering loudspeaker directivity (that might be obtained directly from the manufacturer), the room effect, and more detailed estimation of the individual sources of error.

5 CONCLUSION AND FUTURE WORK

In summary we have proposed a new framework for robust sound zone reproduction design, which uses acoustic modeling to derive the entries for calculating robust control parameters. We formulated and implemented robust ACC based on WCO and PMO for multiplicative and additive error as examples of the proposed framework. We illustrated the performance improvements and comparisons by both simulation and experimental studies and demonstrated the effectiveness of acoustic-modeling based robust ACC, with "quick" error estimation. In experimental observations, incorporation of acoustic modeling gave comparable performance to the state-of-the-art methods for regularization parameter selection in robust filter design, and enabled the number of in-situ measurements to be reduced.

Future work should conduct subjective evaluation of the reproduced audio through various robust strategies, investigate the effects of different loudspeaker directivities on the robust optimization, introduce room modeling techniques alongside source radiation analysis to make better acoustic modeling of both the determined part and stochastic part for reflective room reproduction, and explore the accuracy needed in acoustic modeling for robust ACC to further reduce in-situ measurements.

6 ACKNOWLEDGMENTS

7 REFERENCES

[1] T. Betlehem, W. Zhang, M. A. Poletti and T. D. Abhayapala, "Personal Sound Zones: Delivering Interface-Free Audio to Multiple Listeners," *IEEE Sig. Proc. Mag.*, vol. 32, no. 2, pp. 81–91 (2015 Mar.). https://doi.org/10.1109/msp.2014.2360707

[2] P. Coleman, P. J. B. Jackson, M. Olik, M. Møller, M. Olsen and J. A. Pedersen, "Acoustic Contrast, Planarity and Robustness of Sound Zone Methods Using a Circular Loudspeaker Array," *J. Acoust. Soc. Am.*, vol. 135, no. 4, pp. 1929–1940 (2014 Apr.). https://doi.org/10.1121/1.4866442

[3] J.-W. Choi and Y.-H. Kim "Generation of an Acoustically Bright Zone with an Illuminated Region Using Multiple Sources," *J. Acoust. Soc. Am.*, vol. 111, no. 4, pp. 1695–1700 (2002 Apr.). https://doi.org/10.1121/1.1456926

[4] M. Poletti, "An Investigation of 2-D Multizone Surround Sound Systems," presented at the *125th Convention of the Audio Engineering Society* (2008 Oct.), convention paper 7551.

[5] M. Kolundzija, C. Faller, and M. Vetterli, " Reproducing Sound Fields Using MIMO Acoustic Channel Inversion, " *J. Audio Eng. Soc.*, vol. 59, pp. 721–734 (2011 Oct.).

[6] P. Coleman, P. J. B. Jackson, M. Olik and J. A. Pedersen, "Personal Audio with a Planar Bright Zone," *J. Acoust. Soc. Am.*, vol. 136, no. 4, pp. 1725–1735 (2014 Oct.). https://doi.org/10.1121/1.4893909

[7] Y. Cai, M. Wu and J. Yang, "Sound Reproduction in Personal Audio Systems Using the Least-Squares Approach with Acoustic Contrast Control Constraint," *J. Acoust. Soc. Am.*, vol. 135, no. 2, pp. 734–741 (2014 Feb.). https://doi.org/10.1121/1.4861341

[8] S. J. Elliott, J. Cheer, J.-W. Choi and Y. Kim, "Robustness and Regularization of Personal Audio Systems," *IEEE Trans. Audio, Speech, Language Process.*, vol. 20, no. 7, pp. 2123–2133 (2012 Sep.). https://doi.org/10.1109/tasl.2012.2197613

[9] P. Coleman, "*Loudspeaker Array Processing for Personal Sound Zone Reproduction*," Doctoral Thesis, University of Surrey, Guildford, UK (2014).

[10] D. Kim, K. Kim, S. Wang, S. Q. Lee and M. J. Crocker, "Maximization of the Directivity Ratio with the Desired Audible Gain Level for Broadband Design of Near Field Loudspeaker Arrays," *J. Sound Vibration*, vol. 330, no. 23, pp. 5517–5529 (2011 Nov.). https://doi.org/10.1016/j.jsv.2011.06.019

[11] J. Cheer, S. J. Elliott and M. F. Simón-Gálvez, "Design and Implementation of a Car Cabin Personal Audio System," *J. Audio Eng. Soc.*, vol. 61, pp. 412–424 (2013 Jul./Aug.).

[12] J. Cheer, S. J. Elliott, Y. Kim and J.-W. Choi, "Practical Implementation of Personal Audio in a Mobile Device, " *J. Audio Eng. Soc.*, vol. 61, pp. 290–300 (2013 Jun.).

[13] J.-Y. Park, J.-W. Choi and Y.-H. Kim, "Acoustic Contrast Sensitivity to Transfer Function Errors in the Design of a Personal Audio System," *J. Acoust. Soc. Am.*, vol. 134, no. 1, pp. EL112–EL118 (2013 Jun.). https://doi.org/10.1121/1.4809778

[14] M. R. Bai and C.-C. Chen, "Regularization Using Monte Carlo Simulation to Make Optimal Beamformers Robust to System Perturbations," *J. Acoust. Soc. Am.*, vol. 135, no. 5, pp. 2808–2820 (2014 May). https://doi.org/10.1121/1.4869676

[15] S. Shahbazpanahi, A. B. Gershman, Z.-Q. Luo and K. M. Wong, "Robust Adaptive Beamforming for General-Rank Signal Models," *IEEE Trans. Signal Pro-

cess., vol. 51, no. 9, pp. 2257–2269 (2003 Sep.). http://doi.org/10.1109/tsp.2003.815395

[16] S. Doclo and M. Moonen, "Design of Broadband Beamformers Robust against Gain and Phase Errors in the Microphone Array Characteristics," *IEEE Trans. Signal Process.*, vol. 51, no. 10, pp. 2511–2526 (2003 Oct.). https://doi.org/10.1109/tsp.2003.816885

[17] H. Zhang, "*Xiao chi du yang sheng qi zhen lie suan fa yan jiu [Research on Micro-Loudspeaker Array Systems]*," Master Thesis, Nanjing University, Nanjing, China (2013).

[18] Y. Cai, L. Liu, M. Wu and J. Yang, "Robust Time-Domain Acoustic Contrast Control Design under Uncertainties in the Frequency Response of the Loudspeakers," in *Proc. Inter-Noise 2014*, vol. 249, no. 1, pp. 5775–5780, (Melbourne, Australia, 2014 Nov.).

[19] M. F. Simón-Gálvez, S. J. Elliott and J. Cheer, "The Effect of Reverberation on Personal Audio Devices," *J. Acoust. Soc. Am.*, vol. 135, no. 5, pp. 2654–2663 (2014 May). https://doi.org/10.1121/1.4869681

[20] Y. Cai, "*Ji yu yang sheng qi zhen lie de ju bu sheng chong fang yan jiu [Investigation on Local Sound Reproduction Using Loudspeaker Array]*," Doctoral Thesis, Institute of Acoustics, Chinese Academy of Sciences, Beijing, China (2014).

[21] J.-H. Chang, J.-Y. Park and Y.-H. Kim, "Scattering Effect on the Sound Focused Personal Audio System," *J. Acoust. Soc. Am.*, vol. 125, no. 5, pp. 3060–3066 (2009 May). http://doi.org/10.1121/1.3101453

[22] Z. Tu, J. Lu and X. Qiu, "Robustness of a Compact Endfire Personal Audio System Against Scattering Effects (L)," *J. Acoust. Soc. Am.*, vol. 140, no. 4, pp. 2720–2724 (2016 Oct.). https://doi.org/10.1121/1.4964752

[23] Q. Zhu, P. Coleman, M. Wu and J. Yang, "Robust Personal Audio Reproduction Based on Acoustic Transfer Function Modelling," presented at the *2016 AES International Conference on Sound Field Control* (2016 Jul.), conference paper 1-4.

[24] W. Klippel and C. Bellmann, "Holographic Nearfield Measurement of Loudspeaker Directivity," presented at the *141st Convention of the Audio Engineering Society* (2016 Oct.), convention paper 9598.

[25] M. Melon, C. Langrenne, P. Herzog and A. Garcia, "Evaluation of a Method for the Measurement of Subwoofers in Usual Rooms," *J. Acoust. Soc. Am.*, vol. 127, no. 1, pp. 256–263 (2010 Jan.). https://doi.org/10.1121/1.3270392

[26] H. Cox, R. M. Zeskind and M. M. Owen, "Robust Adaptive Beamforming," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 35, no. 10, pp. 1365–1376 (1987 Oct.).

[27] M. Shin, F. M. Fazi, P. A. Nelson and F. C. Hirono, "Controlled Sound Field with a Dual Layer Loudspeaker Array," *J. Sound Vibration*, vol. 333, no. 16, pp. 3794–3817 (2014 Aug.). https://doi.org/10.1016/j.jsv.2014.03.025

[28] F. Olivieri, M. Shin, F. M. Fazi, P. A. Nelson and P. Otto, "Loudspeaker Array Processing for Multi-Zone Audio Reproduction Based on Analytical and Measured Electroacoustical Transfer Functions," presented at the *AES 52nd International Conference: Sound Field Control* (2013 Sep.), conference paper P-4.

[29] J. Francombe, R. Mason, M. Dewhurst and S. Bech, "Determining the Threshold of Acceptability for an Interfering Audio Programme," presented at the *132nd Convention of the Audio Engineering Society* (2012 Apr.), convention paper 8639.

THE AUTHORS

Qiaoxi Zhu   Philip Coleman   Ming Wu   Jun Yang

Qiaoxi Zhu received the B. Sc. in acoustics from Nanjing University, Nanjing, China, in 2012. Currently she is working towards the Ph.D. degree in acoustic signal processing at the Institute of Acoustics, Chinese Academy of Sciences, Beijing, China. From Nov. 2015 to Nov. 2016, she was a visiting Ph.D. student at Centre for Vision, Speech and Signal processing, University of Surrey, UK. Her research interests include sound field control and array signal processing.

•

Philip Coleman joined the Centre for Vision, Speech and Signal Processing, University of Surrey, UK, in 2010, earning his Ph.D. in 2014 on the topic of personal sound zones. He is currently working in the center as a research fellow on the project S3A: Future spatial audio for an immersive listening experience at home, with a focus on recording and editing object-based content. His research interests include sound field control, loudspeaker and microphone array processing, and spatial audio. Previously, he received the B.Eng. degree in electronic engineering with music technology systems in 2008 from the University of York, UK, and M.Sc. with distinction in multimedia signal processing and communication from the University of Surrey, UK, in 2010.

•

Ming Wu graduated with a degree in electronics from Nanjing University, Nanjing, China, in 2002 and received the Ph.D. degree from Nanjing University in 2007 with a dissertation on active noise control. He was with the Institute of Acoustics, Nanjing University, as a Postdoctoral Researcher from 2007 to 2008. He has been with the Institute of Acoustics, Chinese Academy of Sciences, Beijing, as an Associate Professor since 2008 and a Professor since 2016. His main research areas include noise and vibration control, electro-acoustics, and audio signal processing.

•

Jun Yang received the B. Eng. and M. Eng. degrees from Harbin Engineering University, Harbin, China, and the Ph.D. degree in acoustics from Nanjing University, Nanjing, China, in 1990, 1993, and 1996, respectively. From 1996 to 1998, he was a Postdoctoral Fellow at the Institute of Acoustics, Chinese Academy of Sciences (IACAS), Beijing, China. From October 1998 to April 1999, he was with Hong Kong Polytechnic University as a Visiting Scholar. From Jan. 1997 to May 1999, he was with IACAS as an Associate Professor. He joined the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, as a Research Fellow, a Teaching Fellow, Assistant Professor, and Associate Professor in 1999, 2001, 2003, and 2005, respectively. Since Nov. 2003, he has been a Professor at IACAS. Currently, he is the Director of the Key Laboratory of Noise and Vibration Research, Institute of Acoustics, Chinese Academy of Sciences. His main areas of research interests include communication acoustics, 3-D audio systems, acoustic signal processing, sound field control, and nonlinear acoustics. He is a Fellow of the International Institute of Acoustics and Vibration (IIAV).