

Robust Personal Audio Geometry Optimization in the SVD-based Modal Domain

Qiaoxi Zhu, Philip Coleman, Xiaojun Qiu, Ming Wu, Jun Yang, and Ian Burnett

Abstract—Personal audio generates sound zones in a shared space to provide private and personalized listening experiences with minimized interference between consumers. Regularization has been commonly used to increase the robustness of such systems against potential perturbations in the sound reproduction. However, the performance is limited by the system geometry such as the number and location of the loudspeakers and controlled zones. This paper proposes a geometry optimization method to find the most geometrically robust approach for personal audio amongst all available candidate system placements. The proposed method aims to approach the most ‘natural’ sound reproduction so that the solo control of the listening zone coincidentally accompanies the preferred quiet zone. Being formulated in the SVD-based modal domain, the method is demonstrated by applications in three typical personal audio optimizations, i.e., the acoustic contrast control, the pressure matching and the planarity control. Simulation results show that the proposed method can obtain the system geometry with better avoidance of ‘occlusion’, improved robustness to regularization, and improved broadband equalization.

Index Terms—Spatial audio, sound zones, robustness.

I. INTRODUCTION

PERSONAL audio [1] allows different listeners to hear different sound with minimized interference between each other while sharing the same space. To achieve this, the spatial distribution of a reproduced sound field is optimized by weighting the loudspeaker channels as a spatial filter. Its potential applications include loudspeaker-based audio guides in museums and galleries, private communication devices in banks and hospitals, hearing aids to compensate hearing loss [2], portable and personalized entertainment/working devices in transportation, open offices, restaurants and cafés [1], [3]–[5], and support for future audio services [6].

Since Druyvesteyn and Garas [7] first put forward the concept of personal audio in 1997, various methods have been

developed for its realization [8]–[15]. Among them, three typical methods are adopted in this paper as application examples of the proposed geometry optimization. They are the acoustic contrast control method (ACC) [8], the pressure matching method (PM) [9] and the planarity control method (PC) [11]. ACC maximizes the acoustic energy contrast between the pre-defined listening and quiet zones, PM minimizes error between the reproduced and the desired sound pressures at the controlled points, and PC maximizes correlation between the reproduced sound field and the preferred directional components of the local sound field. The performance of ACC, PM and PC has been compared and reviewed in [11], [16].

Robustness is an important aspect for practical applications. The existing methods for personal audio require knowledge of the transfer functions between the loudspeakers and the receiving points at the controlled zones. Various factors, such as inconsistencies in the sources’ sensitivity, complexities in the spatial responses, and mismatch in sources’ positions, lead to fluctuations in the transfer functions, and eventually deteriorate system performance [10], [17], [18]. Regularization [19]–[21] is the most commonly used method for robustness improvement. The literature [10], [18], [22] reviewed in detail the development of robustness control in personal audio. Our previous work [16], [22] showed that an acoustic-modelling based regularization can increase a personal audio system’s robustness with coarse prior-information, and can be further employed to system implementation with less in-situ measurement of transfer functions.

Apart from regularization, geometry optimization is also pursued to ensure robust performance. This is because the optimal performance through personal audio optimization and regularization is limited by the physical potential of the given system geometry. Though optimized geometry is desired in practical applications, limited literature has addressed this topic. Early work [23]–[25] identified an inversely proportional relationship between loudspeaker spacing and frequency for robust reproduction for crosstalk cancellation systems. In multi-zone sound reproduction, if the desired soundfield in the listening zone is due to a few virtual source directions, the loudspeaker weights can be regularised with the L1 norm, resulting in only a few loudspeakers placed closely to the virtual source directions being activated [26], [27]. Methods using objective [18] or perceptual metrics [28] have been studied for choosing loudspeaker locations with the array effort based regularization. From our previous work [22] and [16], the array effort based regularization has difficulty to determine the threshold and sometimes cannot provide proper regularization.

Q. Zhu is with Centre for Audio, Acoustics and Vibration, Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney, NSW 2007, Australia, and was with Key Laboratory of Noise and Vibration Research, Institute of Acoustics, Chinese Academy of Sciences, Beijing, 100190, China and University of Chinese Academy of Sciences, Beijing, 100049, China (e-mail: qiaoxi.zhu@gmail.com).

P. Coleman is with Institute of Sound and Recording, University of Surrey, Guildford, Surrey, GU2 7XH, UK (e-mail: p.d.coleman@surrey.ac.uk).

X. Qiu and I. Burnett are with Centre for Audio, Acoustics and Vibration, Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney, NSW 2007, Australia (e-mail: Xiaojun.Qiu@uts.edu.au, Ian.Burnett@uts.edu.au).

M. Wu and J. Yang are with Key Laboratory of Noise and Vibration Research, Institute of Acoustics, Chinese Academy of Sciences, Beijing, 100190, China and University of Chinese Academy of Sciences, Beijing, 100049, China (e-mail: mingwu@mail.ioa.ac.cn, jyang@mail.ioa.ac.cn).

Manuscript received April 19, 2005; revised August 26, 2015.

TABLE I
COST FUNCTIONS AND SOLUTIONS IN PERSONAL AUDIO OPTIMIZATION.

Cost functions
$\max_{\mathbf{w}} J_{\text{ACC}}(\mathbf{w}) = \ \mathbf{G}_L \mathbf{w}\ ^2 - \lambda_{\text{ACC}} \ \mathbf{G}_Q \mathbf{w}\ ^2$
$\min_{\mathbf{w}} J_{\text{PM}}(\mathbf{w}) = \ \mathbf{G}_L \mathbf{w} - \mathbf{p}_{\text{des}}\ ^2 + \lambda_{\text{PM}} \ \mathbf{G}_Q \mathbf{w}\ ^2$
$\max_{\mathbf{w}} J_{\text{PC}}(\mathbf{w}) = (\mathbf{G}_L \mathbf{w})^H \mathbf{\Lambda} (\mathbf{G}_L \mathbf{w}) - \lambda_{\text{PC}} \ \mathbf{G}_Q \mathbf{w}\ ^2$
Regularized solutions
$\mathbf{w}_{\text{ACC}} = \Phi[(\mathbf{G}_Q^H \mathbf{G}_Q + \delta_Q \mathbf{I})^{-1} (\mathbf{G}_L^H \mathbf{G}_L + \delta_L \mathbf{I})]$ [22]
$\mathbf{w}_{\text{PM}} = [(\mathbf{G}_L^H \mathbf{G}_L + \delta_L \mathbf{I}) + \lambda_{\text{PM}} (\mathbf{G}_Q^H \mathbf{G}_Q + \delta_Q \mathbf{I})]^{-1} \mathbf{G}_L^H \mathbf{p}_{\text{des}}$ [16]
$\mathbf{w}_{\text{PC}} = \Phi[(\mathbf{G}_Q^H \mathbf{G}_Q + \delta_Q \mathbf{I})^{-1} (\mathbf{G}_L^H \mathbf{\Lambda} \mathbf{G}_L + \zeta_L \mathbf{I})]$ [16]

The objective of this paper is to improve the robust performance of the system by optimizing both the system geometry and the regularization approach. Besides the adoption of acoustic-modelling based regularization into geometry optimization, there are three contributions in the paper. First, we formulate ACC, PC and PM in the singular value decomposition (SVD)-based modal domain, which gives new insights into the contributions of the physical modes in a personal audio system and the effects of regularization. Second, we propose a measure of realisability based on the SVD of the system geometry, which gives the system designer an indication of the efficiency (and robustness) with which sound zones can be generated by that geometry. Third, we use the proposed measure as the cost function for selecting a set of loudspeakers from a candidate set, demonstrating an improvement over the state-of-the-art placement optimization.

Section II presents the SVD-based modal domain personal audio including the ACC, PM and PC optimizations. Section III introduces the proposed measure of realisability and its application in geometry optimization. Section IV interprets the personal audio in the SVD-based modal domain and compares the proposed placement selection with the state-of-the-art methods by simulations. The conclusions are drawn in Section V.

II. PERSONAL AUDIO IN THE SVD-BASED MODAL DOMAIN

A. Personal Audio Optimizations

The ACC, PM and PC optimizations are presented in Table I, with their objectives separately targeting the listening and the quiet zones, and the corresponding regularized solution \mathbf{w} . \mathbf{G}_L and \mathbf{G}_Q are $M \times L$ transfer function matrices defining the listening and the quiet zones, respectively, where L is the number of loudspeakers for reproduction and M is the number of control points in each zone. The element $\mathbf{G}_{(m,l)}$ includes the sound propagation attenuation and delay between the l th loudspeaker and m th control point, and $(*)^H$ denotes complex conjugate matrix transpose.

The three sound zone optimizations pursue the minimized sound energy ($\|\mathbf{G}_Q \mathbf{w}\|^2$) in the quiet zone and have different objectives in the listening zone. In the listening zone, ACC maximizes the sound energy, PM aims to reproduce a pre-defined sound amplitude and phase (\mathbf{p}_{des}) at each control point, and PC maximizes the sound energy towards a certain

spatial range. In ACC and PC, $\Phi(*)$ denotes the eigenvector corresponding to the maximum eigenvalue of a matrix. As introduced in [11], $\mathbf{\Lambda} = \mathbf{Y}^H \mathbf{\Gamma} \mathbf{Y}$ in PC, where $\mathbf{\Gamma}$ is a diagonal matrix for selecting the range of acceptable angles, and the rows of \mathbf{Y} are populated by a regularized fixed-beamwidth max-SNR beamformer [29] to provide an estimate of the plane wave components in the listening zone, with better resolution than the classical delay-and-sum beamformer (equivalent to the spatial Fourier transform) [30].

The cost function of each personal audio optimization is a linear combination of certain objective(s), with non-negative parameters λ_{ACC} , λ_{PM} , and λ_{PC} weighting the relative importance between the listening zone and quiet zone objectives in the optimization. δ_Q , δ_L and ζ_L are the diagonal loading parameters in regularization.

B. Regularization

The acoustic-modelling based regularized personal audio solutions are formulated in [16], [19]. Considering the additive error existing in reproduction systems and the errors in different transfer functions with independent and identically distribution, the additive error is defined as

$$\tilde{\mathbf{G}}_{(m,l)} = \mathbf{G}_{(m,l)} + a e^{j\phi}, \quad (1)$$

where $j = \sqrt{-1}$, a and ϕ are the amplitude and phase of the additive error. Because it is difficult to estimate the error distribution, only the error bound estimate (i.e. the estimate of the maximal error amplitude a_{max}) is used to calculate the loudspeaker weights. For example, a simple error model can have amplitude a uniformly distributed at $[0, a_{\text{max}}]$ and phase ϕ uniformly distributed at $[0, 2\pi]$. In this case, the diagonal term in Table I can be expressed by

$$\delta_Q = \delta_L = M \sigma_a \quad (2)$$

and

$$\zeta_L = s \sigma_a \quad (3)$$

where $\sigma_a = a_{\text{max}}^2/3$ and s is the sum of diagonal elements in $\mathbf{\Lambda}$. The regularization with this simple error model assumption is denoted as AEQ (using the simple Additive Error model for ‘Quick’ implementation).

Besides, the array effort method [10] is widely used to determine the diagonal loads δ_Q , with $\delta_L = \zeta_L = 0$. Apart from the objectives for the ACC, PM, PC optimizations listed in Table I, the additional constraints defined by the array effort can be applied by

$$\text{AE} = 10 \log_{10} \left(\frac{\mathbf{w}^H \mathbf{w}}{|w_0|^2} \right) \leq E_0 \quad (\text{dB}), \quad (4)$$

where w_0 is the input signal required to drive a single element at the center of the array so that the mean square pressure in the listening zone is the same as that when the array is driven by \mathbf{w} [10], $|*|$ is the operator of absolute value, and E_0 is the array effort limit. Low array effort indicates reproduction with high energy efficiency and robustness.

TABLE II
PAIRS OF REGULARIZED PERSONAL AUDIO SOLUTIONS IN THE SVD-BASED MODAL DOMAIN.

Method	\mathbf{k}_Q	\mathbf{k}_L
ACC	$\Phi \left\{ \Omega_Q^{-1} \mathbf{T}^H \Omega_L \mathbf{T} \right\}$	$\Phi \left\{ \mathbf{T} \Omega_Q^{-1} \mathbf{T}^H \Omega_L \right\}$
PM	$(\mathbf{T}^H \Omega_L \mathbf{T} + \lambda_{PM} \Omega_Q)^{-1} \mathbf{T}^H \Sigma_L \mathbf{U}_L^H \mathbf{P}_{des}$	$(\Omega_L + \lambda_{PM} \mathbf{T} \Omega_Q \mathbf{T}^H)^{-1} \Sigma_L \mathbf{U}_L^H \mathbf{P}_{des}$
PC	$\Phi \left\{ \Omega_Q^{-1} \mathbf{T}^H \Pi_L \mathbf{T} \right\}$	$\Phi \left\{ \mathbf{T} \Omega_Q^{-1} \mathbf{T}^H \Pi_L \right\}$

C. SVD Formulation for Regularized Personal Audio Optimizations

Using SVD, the transfer function matrix \mathbf{G}_X can be decomposed as

$$\mathbf{G}_X = \mathbf{U}_X \Sigma_X \mathbf{V}_X^H, \quad (5)$$

where the subscript X can be L (denoting the listening zone) or Q (denoting the quiet zone), \mathbf{U}_X and \mathbf{V}_X are separately $M \times M$ and $L \times L$ unitary matrices and Σ_X is a $M \times L$ diagonal and real valued matrix. The columns of \mathbf{U}_X and \mathbf{V}_X are respectively the modes of the controlled zone space and the loudspeaker space in the system with M and L being their degrees of freedom [31].

Define $\mathbf{k}_X = \mathbf{V}_X^H \mathbf{w} / \|\mathbf{w}\|$ as the expression of the normalized solution $\mathbf{w} / \|\mathbf{w}\|$ in the SVD-based modal domain. The sound field generated by the loudspeaker weights $\mathbf{w} / \|\mathbf{w}\|$ can be represented by \mathbf{k}_X as

$$\mathbf{p}_X = \mathbf{G}_X \mathbf{w} / \|\mathbf{w}\| = \mathbf{U}_X \Sigma_X \mathbf{k}_X, \quad (6)$$

$k_{X,n}$ (the n th element in \mathbf{k}_X) shows the portion of the n th loudspeaker mode ($\mathbf{v}_{X,n}$, the n th column of \mathbf{V}_X) presented in \mathbf{w} . The sound field at the controlled zone reproduced by the n th loudspeaker mode is

$$\mathbf{p}_X^n = \mathbf{G}_X \mathbf{v}_{X,n} = \sigma_{X,n} \mathbf{u}_{X,n}, \quad (7)$$

where $\sigma_{X,n}$ is the n th diagonal element of Σ_X , and $\mathbf{u}_{X,n}$ is the n th column of \mathbf{U}_X . $\sigma_{X,n}$ represents the amount of amplification or attenuation that the mode $\mathbf{v}_{X,n}$ undergoes for the transformation \mathbf{G}_X . Usually, $\sigma_{X,1} \geq \sigma_{X,2} \geq \dots \geq \sigma_{X,L}$.

Because each loudspeaker mode $\mathbf{v}_{X,n}$ corresponds to one controlled zone mode $\mathbf{u}_{X,n}$ independently, L loudspeaker modes can generate L (out of M , $L \leq M$) independent controlled zone modes in total. The sound field generated by the L loudspeakers at the controlled zone are the superposition of L independent components \mathbf{p}_X^n , that

$$\mathbf{p}_X = \sum_{n=1}^L k_{X,n} \cdot \mathbf{p}_X^n. \quad (8)$$

The columns of \mathbf{V}_L and \mathbf{V}_Q serve as two sets of orthogonal modes of the same loudspeaker space corresponding to the listening zone modes and the quiet zone modes. Any normalized loudspeaker weighting vector $\mathbf{w} / \|\mathbf{w}\|$ can be represented separately using both sets of modes, that

$$\mathbf{w} / \|\mathbf{w}\| = \mathbf{V}_L \mathbf{k}_L = \mathbf{V}_Q \mathbf{k}_Q. \quad (9)$$

Because the two sets of orthogonal modes describe the same loudspeaker space, they are transferable, that

$$\mathbf{V}_Q = \mathbf{V}_L \mathbf{T}, \quad (10)$$

where \mathbf{T} is an $L \times L$ unitary transfer matrix between the listening and quiet zone modes. The modal domain expressions of \mathbf{w} using different loudspeaker modes are also transferable with \mathbf{T} , that

$$\mathbf{k}_L = \mathbf{T} \mathbf{k}_Q. \quad (11)$$

The pairs of regularized personal audio solutions (\mathbf{k}_L and \mathbf{k}_Q) presented in Table II is formulated by substituting \mathbf{G}_X and \mathbf{w} in the regularized solutions listed in Table I with their modal domain expressions (5), (9) and (10), as briefly introduced in the appendix. In Table II, the $L \times L$ matrices Ω_Q , Ω_L and Π_L are defined as

$$\begin{cases} \Omega_Q = \Sigma_Q^H \Sigma_Q + \delta_Q \mathbf{I}, \\ \Omega_L = \Sigma_L^H \Sigma_L + \delta_L \mathbf{I}, \\ \Pi_L = \Sigma_L^H \mathbf{U}_L^H \Lambda \mathbf{U}_L \Sigma_L + \zeta_L \mathbf{I}. \end{cases} \quad (12)$$

They are the simplified modal domain formation of the original transfer function related terms $\mathbf{G}_Q^H \mathbf{G}_Q + \delta_Q \mathbf{I}$, $\mathbf{G}_L^H \mathbf{G}_L + \delta_L \mathbf{I}$ and $\mathbf{G}_L \Lambda \mathbf{G}_L + \zeta_L \mathbf{I}$.

D. Modal Capabilities

The normalized capabilities of each quiet/listening zone mode for the reproduction of the according quiet/listening zone are represented as the elements of \mathbf{C} , that

$$\begin{cases} \mathbf{C}_Q = \frac{\hat{\Sigma}_Q}{\|\hat{\Sigma}_Q\|}, & \mathbf{C}_{L,ACC} = \frac{\hat{\Sigma}_L}{\|\hat{\Sigma}_L\|}, \\ \mathbf{C}_{L,PM} = \frac{\hat{\Sigma}_L \hat{\mathbf{U}}_L^H \mathbf{P}_{des}}{\|\hat{\Sigma}_L\| \|\mathbf{P}_{des}\|}, & \mathbf{C}_{L,PC} = \frac{\hat{\Sigma}_L \hat{\mathbf{U}}_L^H \Lambda^{\frac{1}{2}}}{\|\hat{\Sigma}_L\| \|\Lambda^{\frac{1}{2}}\|}, \end{cases} \quad (13)$$

where the matrices $\hat{\Sigma}_Q$ and $\hat{\Sigma}_L$ separately contains the square root of the elements in Ω_Q^{-1} and Ω_L with the n th diagonal element equivalent to $1/\sqrt{\sigma_{Q,n}^2 + \delta_Q}$ and $\sqrt{\sigma_{L,n}^2 + \delta_L}$, $\hat{\mathbf{U}}_L$ is the $M \times L$ matrix truncated by the first L columns of \mathbf{U}_L , and $\Lambda^{\frac{1}{2}} = \mathbf{Y}^H \mathbf{\Gamma}^{\frac{1}{2}}$ with $\mathbf{\Gamma}^{\frac{1}{2}}$ contains the square root of the elements in $\mathbf{\Gamma}$. The n th row and m th column element of \mathbf{C} represents the capability of the n th loudspeaker mode $\mathbf{v}_{X,n}$ in realizing the m th component of local sound field according to the personal audio objective. This formulation reveals the diagonal loading parameter acts to modify the assumed modal capabilities in the loudspeaker weight calculation.

In summary, this section formulates the personal audio in the SVD-based modal domain, which reformulates the problem of solving loudspeaker weights as the problem of solving the share of modes in a solution. As each mode corresponds to a certain capability in generating a dependent

sound field component in the corresponding controlled zone, this formulation directly controls the sound reproduction over the sound zones. Moreover, it formulates regularization as the modification of the assumed modal capabilities in designing sound zone reproduction. These two perspectives will be illustrated in the following Section IV-B by simulation.

III. REALISABILITY MEASURE AND ITS APPLICATION IN GEOMETRY OPTIMIZATION

A. Problem Statement

In personal audio, it is desired that under certain geometry, the solution for generating the optimal quiet zone is the same as the solution for generating the listening zone, so that

$$\mathbf{w}_{L,xx} = \mathbf{w}_Q, \quad (14)$$

where xx can be ACC, PM or PC optimizations. $\mathbf{w}_{L,xx}$ is the solution that solely controls the listening zone in the sound reproduction with the quiet zone uncontrolled, specifically,

$$\begin{cases} \mathbf{w}_{L,ACC} = \Phi(\mathbf{G}_L^H \mathbf{G}_L + \delta_L \mathbf{I}), \\ \mathbf{w}_{L,PM} = (\mathbf{G}_L^H \mathbf{G}_L + \delta_L \mathbf{I})^{-1} \mathbf{G}_L^H \mathbf{P}_{des}, \\ \mathbf{w}_{L,PC} = \Phi(\mathbf{G}_L^H \Lambda \mathbf{G}_L + \zeta_L \mathbf{I}), \end{cases} \quad (15)$$

and \mathbf{w}_Q is the solution that solely controls the quiet zone in the sound reproduction with the listening zone uncontrolled,

$$\mathbf{w}_Q = \Phi \left[(\mathbf{G}_Q^H \mathbf{G}_Q + \delta_Q \mathbf{I})^{-1} \right]. \quad (16)$$

The solution in (16) is equivalent to the eigenvector associated with the minimum eigenvalue of $(\mathbf{G}_Q^H \mathbf{G}_Q + \delta_Q \mathbf{I})$, where regularization parameter $\delta_Q \mathbf{I}$ is necessary for the robustness against potential transfer function perturbations in sound reproduction. With this geometry, $\mathbf{w} = \mathbf{w}_{L,xx} = \mathbf{w}_Q$, the minimal effort is required to control the interference between two contrast zones because no additional effort needed to generate wave superposition for sound cancellation. In this way, the sound zones are most ‘naturally’ reproduced with the most robustness to against potential perturbations in the reproduction.

Unfortunately, the solution (14) does not always exist. The challenge for the optimization problem is to find the geometry (\mathbf{G}_L and \mathbf{G}_Q) leading to the maximized similarity (correlation) of the normalized solo solutions so that

$$\frac{\|\mathbf{w}_L^H \mathbf{w}_Q\|}{\|\mathbf{w}_L\| \|\mathbf{w}_Q\|} \rightarrow 1. \quad (17)$$

In the SVD-based modal domain, define $\mathbf{k}_{L,L-xx}$ as the solution solely controlling the listening zone and $\mathbf{k}_{L,Q}$ as the solution solely controlling the quiet zone, and they are both expressed with the listening zone modes. Then (17) can be reformulated using the listening zone modes as

$$\|\mathbf{k}_{L,L-xx}^H \mathbf{k}_{L,Q}\| \rightarrow 1. \quad (18)$$

Referring to (11),

$$\|\mathbf{k}_{L,L-xx}^H \mathbf{T} \mathbf{k}_{Q,Q}\| \rightarrow 1, \quad (19)$$

where $\mathbf{k}_{Q,Q}$ represents the solution solely controls the quiet zone that is expressed with the quiet zone modes. Expressed

with the modes of the corresponding sound zone, the solutions (15) and (16) to control a solo zone can be derived as

$$\begin{cases} \mathbf{k}_{L,L-ACC} = \Phi \{ \Omega_L \}, \\ \mathbf{k}_{L,L-PM} = \Sigma_L^\dagger \mathbf{U}_L^H \mathbf{P}_{des}, \\ \mathbf{k}_{L,L-PC} = \Phi \{ \Pi_L \}, \\ \mathbf{k}_{Q,Q} = \Phi \{ \Omega_Q^{-1} \}, \end{cases} \quad (20)$$

where \dagger denotes the pseudo-inverse. According to (12), $\Phi \{ \Omega_Q^{-1} \} = [0, \dots, 0, 1]^T$. Thus,

$$\mathbf{k}_{Q,Q} = [0, \dots, 0, 1]^T, \quad (21)$$

which means $\mathbf{k}_{Q,Q}$ only enables the last quiet zone mode. Similarly, $\mathbf{k}_{L,L-ACC} = [1, 0, \dots, 0]^T$, $\mathbf{k}_{L,L-ACC}$ only enables the first listening zone mode. By substituting (21) into (19), there is

$$\|\mathbf{k}_{L,L-xx}^H \mathbf{t}_L\| \rightarrow 1, \quad (22)$$

where \mathbf{t}_L is the L th column of \mathbf{T} . It leads to

$$\mathbf{t}_L \rightarrow \mathbf{k}_{L,L-xx}. \quad (23)$$

The optimization objective (22) directly follows (17); however, it is too strict to be practical because it is not always available to obtain a geometry candidate well-satisfied to (22), which requires, for example, the 1st row and L th column element of \mathbf{T} very close to 1 and other elements nearly zero in the ACC optimization. The optimization objective assuming (21) to be the optimal solution that solely controls the quiet zone is numerically right but too strict.

B. Realisability Measure

Because of the regularization in (12) concerning practical robustness, the last few quiet zone modes perform closely in solely reproducing the quiet zone under practical scenarios with perturbations. Therefore, the optimization objective (22), only considering the last quiet zone mode, excludes much greater potential that enables some of the last few quiet zone modes as $\mathbf{k}_{Q,Q}$. Similarly, if some listening zone modes have similar performance in solely reproducing the listening zone, $\mathbf{k}_{L,L-xx}$ should be relaxed to enable some of the most competitive modes. In this section, we revise the optimization objective (19) by replacing $\mathbf{k}_{Q,Q}$ and $\mathbf{k}_{L,L-xx}$ with the normalized modal capability \mathbf{C}_Q and $\mathbf{C}_{L,xx}$ in (13), which reflect the capability of each mode to solely control the corresponding zone.

The proposed optimization objective is

$$\eta_{xx} = \|\mathbf{C}_Q^H \mathbf{T}^H \mathbf{C}_{L,xx}\| \rightarrow 1, \quad (24)$$

where η_{xx} is a measure of realisability under a certain system geometry, and specifically

$$\left\{ \begin{array}{l} \eta_{\text{ACC}} = \frac{\|\hat{\Sigma}_{\text{Q}} \mathbf{T}^{\text{H}} \hat{\Sigma}_{\text{L}}\|}{\|\hat{\Sigma}_{\text{Q}}\| \|\hat{\Sigma}_{\text{L}}\|}, \\ \eta_{\text{PM}} = \frac{\|\hat{\Sigma}_{\text{Q}} \mathbf{T}^{\text{H}} \hat{\Sigma}_{\text{L}} \hat{\mathbf{U}}_{\text{L}}^{\text{H}} \mathbf{P}_{\text{des}}\|}{\|\hat{\Sigma}_{\text{Q}}\| \|\hat{\Sigma}_{\text{L}}\| \|\mathbf{P}_{\text{des}}\|}, \\ \eta_{\text{PC}} = \frac{\|\hat{\Sigma}_{\text{Q}} \mathbf{T}^{\text{H}} \hat{\Sigma}_{\text{L}} \hat{\mathbf{U}}_{\text{L}}^{\text{H}} \mathbf{\Lambda}^{\frac{1}{2}}\|}{\|\hat{\Sigma}_{\text{Q}}\| \|\hat{\Sigma}_{\text{L}}\| \|\mathbf{\Lambda}^{\frac{1}{2}}\|}, \end{array} \right. \quad (25)$$

with $\eta_{\text{ACC}}, \eta_{\text{PM}}, \eta_{\text{PC}} \in [0, 1]$. The realisability measure evaluates the efficiency and robustness of a system geometry for personal audio optimization by calculating the compatibility of the listening and quiet zone modes in the SVD-based modal domain to achieve combined personal audio objectives.

For example, in the ACC optimization, the i th row and j th column element of $\mathbf{C}_{\text{Q}}^{\text{H}} \mathbf{T}^{\text{H}} \mathbf{C}_{\text{L,ACC}}$ is $c_{\text{Q},i} \cdot t_{j,i}^* \cdot c_{\text{L},j}$, where $c_{\text{Q},i}$ (the i th diagonal elements in \mathbf{C}_{Q}) and $c_{\text{L},j}$ (the j th diagonal elements in $\mathbf{C}_{\text{L,ACC}}$) separately represents the capability of the i th quiet zone mode and the j th listening zone mode to generate accordingly the sound energy minimized quiet zone or the sound energy maximized listening zone, and $t_{j,i}^*$ (the conjugate of the j th row and i th column element of \mathbf{T}) represents the compatibility of co-existence between the i th quiet zone mode and the j th listening zone mode. If the solution \mathbf{k}_{L} has a non-zero value in the j th mode, the solution \mathbf{k}_{Q} will have $t_{j,i}^*$ times of that value in the i th mode, by referring to (11). A large $|t_{j,i}|$ means the i th quiet zone mode and the j th listening zone mode are highly compatible to control the separate zones. Thus, $c_{\text{Q},i} \cdot t_{j,i}^* \cdot c_{\text{L},j}$ calculates the combined capability that naturally generates the i th quiet zone mode and the j th listening zone mode together to achieve certain combined personal audio objectives. Finally, the induced 2-norm operator ($\|\cdot\|$) results an evaluation of the maximal compatibility among all the potential solutions from the combination of all the available modes.

Therefore, with the assumed modal capabilities modified by AEQ regularization, the proposed realisability measure η_{xx} evaluates the compatibility between the listening zone and the quiet zone reproduction. Instead of seeking the system geometry leading to the most compatibility between the numerical solutions $\mathbf{k}_{\text{Q,Q}}$ and $\mathbf{k}_{\text{L,L-xx}}$ separately for solo goals in (19), the proposed objective seeks the system geometry that enables the most compatibility between the modal capabilities separately for solo goals. By modifying the numerical optimization into the modal-capability-based optimization, the drawbacks of the direct numerical objective (19) mentioned above can be avoided. It is a reasonable, necessary and sufficient expansion of the original optimization objective (17).

C. Geometry Optimization in the SVD-based Modal Domain

The freedom of the system geometry can be described by a set of available candidate system placements (including loudspeakers and controlled zones) with the transfer function

matrices $\mathbf{G}_{\text{Q},n}$ and $\mathbf{G}_{\text{L},n}$, where $n = 1, 2, \dots, N$, N is the number of the candidate placements. Among them, the optimized $\mathbf{G}_{\text{Q,opt}}$ and $\mathbf{G}_{\text{L,opt}}$ can be selected based on certain evaluation metric [18]. The geometry optimization using J as evaluation metric is

$$\begin{array}{l} \max_n J(\mathbf{G}_{\text{Q},n}, \mathbf{G}_{\text{L},n}), \\ \text{s.t. } (\mathbf{G}_{\text{Q},n}, \mathbf{G}_{\text{L},n}) \in \{\mathbf{G}_{\text{Q},n}, \mathbf{G}_{\text{L},n}\}_{n=1}^N, \end{array} \quad (26)$$

where $\mathbf{G}_{\text{Q},n}$ and $\mathbf{G}_{\text{L},n}$ are chosen from a set of N candidate placements. For example, the acoustic contrast metric is widely used to evaluate the sound separation between the controlled zones under each candidate placement, and is defined as

$$\text{AC (dB)} = 10 \log_{10} \left(\frac{\mathbf{w}^{\text{H}} \mathbf{G}_{\text{L}}^{\text{H}} \mathbf{G}_{\text{L}} \mathbf{w}}{\mathbf{w}^{\text{H}} \mathbf{G}_{\text{Q}}^{\text{H}} \mathbf{G}_{\text{Q}} \mathbf{w}} \right). \quad (27)$$

The metric $J = \text{AC}$ is used as the evaluation metric in the geometry optimization with the ACC optimization. Similarly, the geometry optimization with the PM and PC optimizations separately uses the level of $J = \|\mathbf{G}_{\text{L}} \mathbf{w} - \mathbf{p}_{\text{des}}\|^2 + \lambda_{\text{PM}} \|\mathbf{G}_{\text{Q}} \mathbf{w}\|^2$ and $J = (\mathbf{G}_{\text{L}} \mathbf{w})^{\text{H}} \mathbf{\Lambda} (\mathbf{G}_{\text{L}} \mathbf{w}) / \|\mathbf{G}_{\text{Q}} \mathbf{w}\|^2$ as evaluation metric. Because J is a numerical metric with limited physical concern, the obtained placement selection might not always obtain physical reliable result. This drawback can be improved by using the proposed realisability measure as evaluation metric in the geometry optimization, specifically

$$\begin{array}{l} \max_n \eta(\mathbf{G}_{\text{Q},n}, \mathbf{G}_{\text{L},n}), \\ \text{s.t. } (\mathbf{G}_{\text{Q},n}, \mathbf{G}_{\text{L},n}) \in \{\mathbf{G}_{\text{Q},n}, \mathbf{G}_{\text{L},n}\}_{n=1}^N. \end{array} \quad (28)$$

By maximizing η over candidate placements, this optimization tends to enable the loudspeaker mode(s) most compatible for both the listening zone and quiet zone reproduction.

In summary, in this section we put forward a new measure of realisability formulated in the SVD-based modal domain and proposed system geometry optimization using the new measure as evaluation metric. The improved geometric robustness by the proposed method applied to the ACC optimization will be demonstrated by simulations in the following Section IV-C. In Section IV-D, it will be discussed that the proposed method is also applicable for the PM and PC optimizations.

IV. SIMULATION AND DISCUSSION

A. Setup

The simulation is carried out on a simplified application scenario which places a loudspeaker array around a space. The optimization goal is to select 11 continuous loudspeakers (as an arc-shaped array) out of 60 candidate loudspeaker locations, evenly distributed on a circle surrounding two pre-defined sound zones. Based on the symmetric geometry setting, there are 31 unique candidate placements as illustrated in Fig. 1, where the candidate placement indexes $n = 1, 2, \dots, 31$ (from the left-end-centred to right-end-centred array placement) are denoted at the center loudspeaker of each placement. Each controlled zone is sampled by 48 control points. This system geometry is an extension of the previous work using a pre-defined 11-channel arc-shaped loudspeaker array [22] and using 60-channel circular loudspeaker array [11].

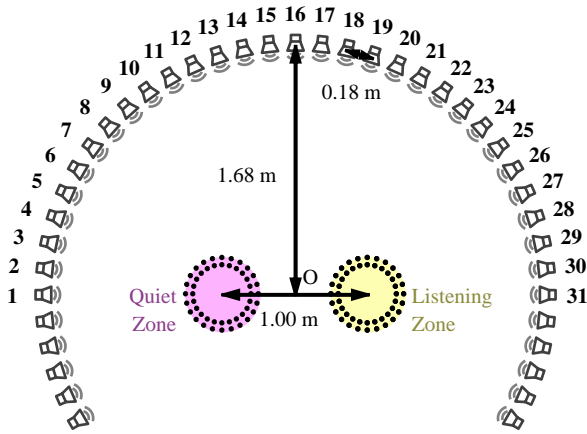


Fig. 1. System geometry for two-zone reproduction. Candidate placements as subsets of 11 continuous loudspeakers centered at positions 1–31 may be selected. Each sound zone is sampled by 48 control points.

The simulation assumes that each loudspeaker acts as a monopole, with transfer function defined by e^{jkr}/r , where k is the wave number and r is the distance to a control point. Perturbations are added to the spatial responses, assuming that the error has multiplicative form with Gaussian distribution between -3 dB and $+3$ dB in magnitude and uniform distribution between -10° and $+10^\circ$ in phase. Samples of $\tilde{\mathbf{G}}_L$ and $\tilde{\mathbf{G}}_Q$ are drawn from these distributions for Monte-Carlo trials. For the regularization, to make the assumed additive error set contain the real multiplicative error set, AEQ adopts $a_{\max} = \max\{\mathbf{G}\} \times a_{\max,ME}$, where $\max\{\mathbf{G}\}$ is the maximum among all the transfer functions in \mathbf{G} , and $a_{\max,ME} = \sqrt{\mu^2 - 2\cos(\phi_{\max})\mu + 1}$ with $\mu = 10^{\frac{3}{20}}$, $\phi_{\max} = 10^\circ$. Two array effort (AE) based regularizations, EL0 employing $E_0 = 0$ dB in (4) and ELM limiting AE towards the minimal value, are also applied in the placement selection for comparison. ELM also acts as the solution only controlling the listening zone (15). EL0, AEQ and ELM separately represent inadequate (‘under-robust’), proper and excessive (‘over-robust’) regularization [22].

The simulation results using ACC optimization are presented in Sec. IV-B and IV-C, and those using PM or PC optimization are described in Sec. IV-D.

B. Regularized Personal Audio in the SVD-based Modal Domain

The sound pressure level of the sound fields at 1 kHz solely generated by the first and last three quiet zone modes under the candidate system placement no. 16 are illustrated in Fig. 2. The upper row shows the reproduced sound fields under ideal situation without perturbations, and the lower row shows those with a trial of random perturbation. It is clear that each quiet zone mode generates a unique and orthogonal modal pattern in the quiet zone. In Fig. 2, the local sound field in the quiet zone generated by the first quiet zone mode is almost evenly distributed, while the one generated by the second quiet zone mode has a line-shaped low SPL gap across the zone, and the one generated by the third quiet zone mode has a cross-

TABLE III
ACOUSTIC CONTRAST (AC) PERFORMANCE REPRODUCED BY EACH QUIET ZONE MODE, EVALUATED UNDER IDEAL CASE WITHOUT MULTIPLICATIVE ERROR OR AVERAGED OVER 1000 MONTE-CARLO TRIALS.

Mode no.		1	2	3	...	9	10	11
AC (dB)	ideal	-23.2	-6.2	8.6	...	128.9	147.2	164.6
	aver.	-21.9	-6.0	7.5	...	21.3	16.2	7.6

shaped low SPL gap across the zone. Those generated by the last three modes have very low SPL across the zone as their modal pattern are more complicated and hard to reproduce.

Comparison of the reproduction without perturbation (upper row) and that with random perturbation (lower row) shows that different modes have different robustness, and the reproduced local sound fields of higher order modes are more sensitive to perturbation. Table III further quantifies the corresponding AC performance of the sound fields separately generated by each quiet zone mode, where the AC values in the upper row are evaluated under the ideal case without multiplicative error while those in the lower row are averaged over 1000 Monte-Carlo trials. The superposition of the local sound fields separately reproduced by all the 11 quiet zone modes covers all the potential local sound field distribution in the quiet zone; however, only the first and last three quiet zone modes are presented in Fig. 2 for conciseness and clarity.

Figure 3 presents the modal capabilities using the candidate system placement no. 16 at 1 kHz for ACC optimization. The modal capabilities are defined as the diagonal elements of \mathbf{C}_Q and $\mathbf{C}_{L,ACC}$ in (13), separately, without regularization or with AEQ regularization. It shows that the capabilities of the last few quiet zone modes without regularization can be extremely large, making the personal audio solution dominated by the modes highly sensitive to perturbations. With the AEQ regularization, the assumed capabilities of these modes are modified to a certain level, resulting in a more stable solution.

Figure 4 shows the amplitude of the elements in \mathbf{k}_L and \mathbf{k}_Q for the ACC optimization using the system placement no. 16 at 1 kHz, without regularization or with AEQ regularization. According to (9), $\mathbf{k}_L = \mathbf{V}_L^H \mathbf{w} / \|\mathbf{w}\|$ and $\mathbf{k}_Q = \mathbf{V}_Q^H \mathbf{w} / \|\mathbf{w}\|$. The presented amplitudes in Fig. 4 represent the proportion of each listening/quiet zone mode in the typical ACC solution. It shows that the ACC solution without regularization converges to the unstable last quiet zone mode, while the ACC solution with the AEQ regularization converges to the stable first listening zone mode, referring to the robustness of each mode illustrated in Fig. 2. Furthermore, the ACC solution with the AEQ regularization simultaneously converges to the ‘middle’ quiet zone modes no. 7–9, which have relatively good capability in both quiet zone generation and robustness. So, with the incorporation of AEQ regularization, the ACC solution balances the reproduction of the two contrast zones.

C. Placement Selection Results for ACC

Figure 5 illustrates the process of placement selection, including: (1) calculating the transfer function matrices $\mathbf{G}_{L,i}$ and $\mathbf{G}_{Q,i}$ for each candidate placement no. i , $i = 1, 2, \dots, N$;

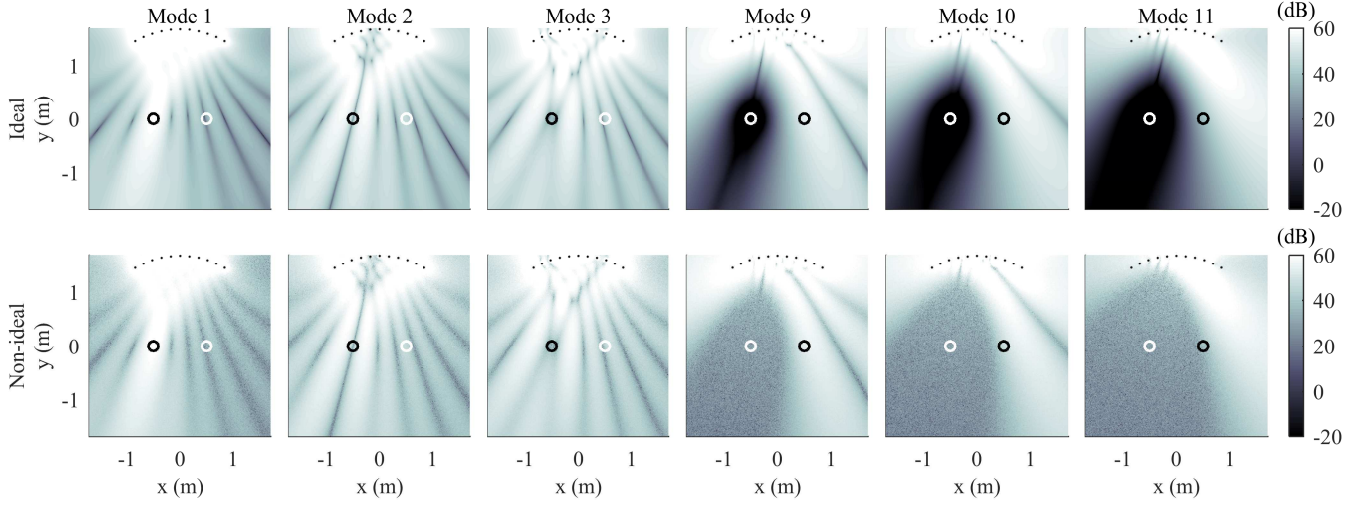


Fig. 2. Visualization of the sound pressure level (SPL) of the sound field generated by each quiet zone mode at 1 kHz, using placement no. 16 (center). The three largest and three smallest modes are presented. The upper row shows reproduction under ideal case; the lower row is reproduction with random perturbation. Based on the symmetric system geometry of placement no. 16, the listening zone modes' performance is symmetric.

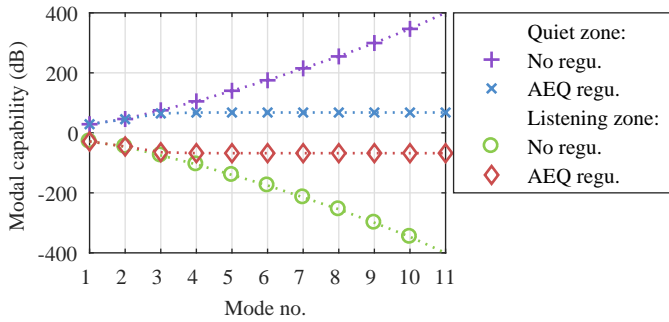


Fig. 3. The modal capability modification for acoustic contrast control using AEQ regularization, demonstrated with placement no. 16 at 1 kHz. The assumed capability of each mode for the objective in the corresponding sound zone is presented. AEQ acts to modify the assumed capabilities to their proper levels based on a coarse error estimation.

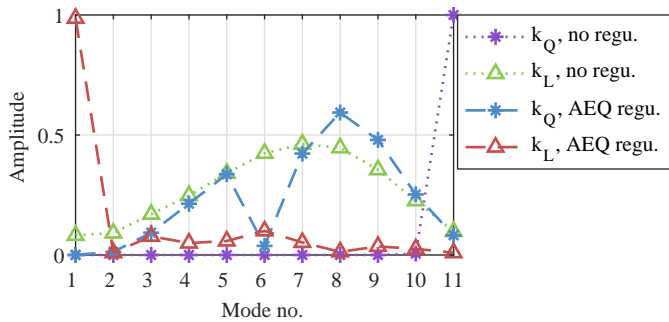


Fig. 4. The share of each mode of acoustic contrast control solution in the SVD-based modal domain with and without AEQ regularization using placement no. 16 at 1 kHz. The acoustic contrast control solution without regularization converges to the last quiet zone mode, while the solution with AEQ regularization converges to the first listening zone mode.

(2) calculating the value of the selection metric under the given geometry; (3) selecting the placement with the maximal metric value. Four loudspeaker placement selection approaches are compared in the following simulations. The approaches using AC as the selection metric are separately incorporating different regularizations, i.e., ELO, AEQ and ELM. They are

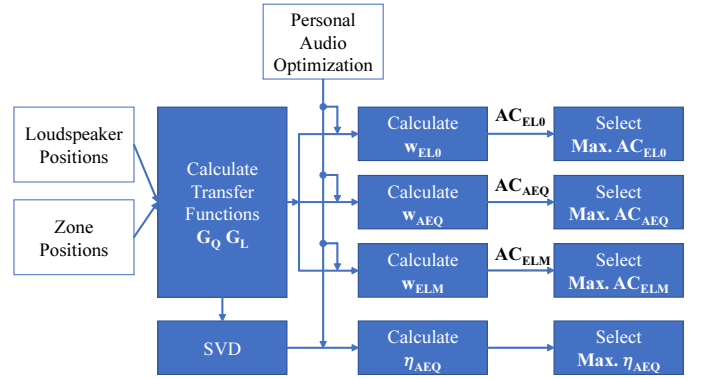


Fig. 5. The flowchart of system placement selection with four alternative approaches for acoustic contrast control, separately using acoustic contrast (AC) and realisability (η) as evaluation metric and incorporating ELO, AEQ and ELM as regularization.

separately representative for under-robust, robust and over-robust regularizations as studied in [22], and the corresponding AC metrics are denoted as AC_{ELO} , AC_{AEQ} and AC_{ELM} . The proposed approach uses η as the selection metric with the AEQ regularization to modify the assumed modal capabilities, and the metric is denoted as η_{AEQ} .

Figure 6 presents the patterns that the evaluation metrics AC_{ELO} , AC_{AEQ} , AC_{ELM} and η_{AEQ} vary with system placements and frequencies. The observation is conducted among placement no. 1–31, as gradient varying system geometries, and over 100–3600 Hz sampled by 100 Hz, covering the typical low/high/grating-lobe frequencies [22]. AC_{ELO} , AC_{AEQ} , AC_{ELM} of each placement and frequency are averaged over 1000 Monte-Carlo trials. Obviously, the ACs achieved with the ELO or ELM regularization are less satisfactory than that with the AEQ regularization. So it is important to apply proper regularization to the metric calculation for placement selection. With the same AEQ regularization, AC_{AEQ} and η_{AEQ} have quite different value patterns over geometry and frequency, as the former evaluates the numerical sound separation between

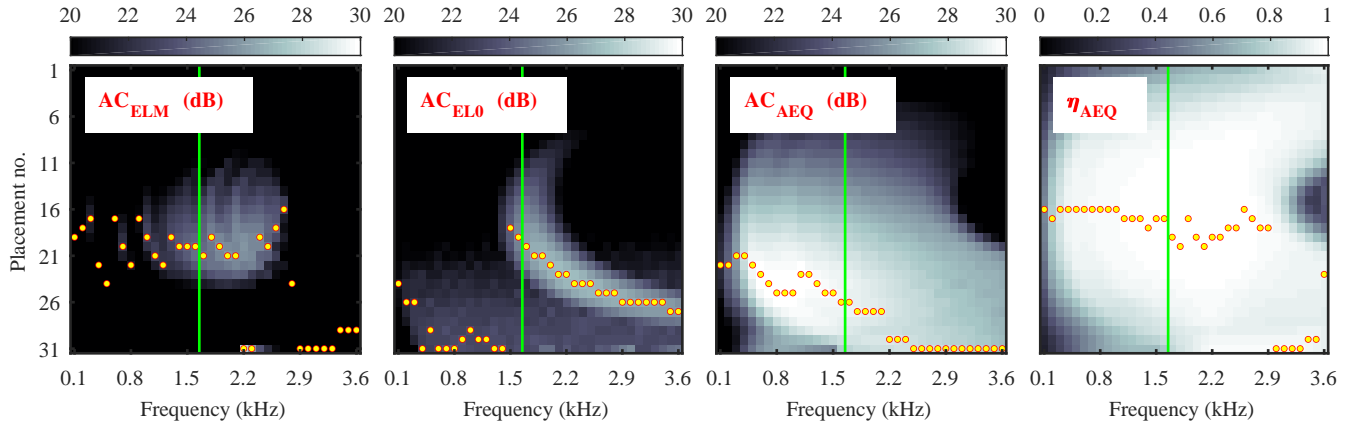


Fig. 6. The values of the mean acoustic contrast (AC) and realisability (η) metric for acoustic contrast control optimization using each candidate placement over 100–3600 Hz. AC_{EL0} , AC_{ELM} and AC_{AEQ} are mean AC separately employing EL0, ELM and AEQ as regularization. η_{AEQ} is η with AEQ regularization. The yellow dots mark the maximized metric values of each sampled frequency to indicate the selected placement indexes. The green solid lines denote $f_{zone} = 1649$ (Hz).

TABLE IV

COMPARISON OF PLACEMENT SELECTION APPROACHES AT 800 HZ, USING ACOUSTIC CONTRAST (AC) AND THE PROPOSED REALISABILITY (η) AS THE COST FUNCTIONS. THE REGULARIZATION USED IS DENOTED BY THE SUBSCRIPT, AND THE ARRAY EFFORT (AE) IS ALSO EVALUATED.

Method	AC_{EL0}	AC_{AEQ}	AC_{ELM}	η_{AEQ}
Selected no.	31	24	22	16
Mean AC (dB)	24.2	30.2	21.0	27.5
η	\	0.90	\	0.98
AE (dB)	-1.4	-9.0	-10.0	-10.1

zones and the later evaluates the physical naturalness to reproduce separate zones.

As marked by the yellow dots in Fig. 6, the index of the selected placement at each sampled frequency is determined by the position that the maximized value of the evaluation metric is achieved. The turning frequency f_{zone} is denoted by a green line, which is defined by the ratio of sound speed c and the diameter of the controlled zone D_{zone} , that

$$f_{zone} = c/D_{zone} = 343/(2 \times 0.104) \approx 1649 \text{ (Hz)}. \quad (29)$$

The selected placement optimized by η_{AEQ} is almost stable around no. 16 when $f \leq f_{zone}$, while varies and gradually inclines to no. 31 with increasing frequency when $f > f_{zone}$.

The performance of the optimized placements by the four approaches at 800 Hz is presented in Table IV. As improper regularization adopted to calculate w in AC evaluation (27), the placements optimized by AC_{EL0} and AC_{ELM} have obviously lower mean AC performance, which fail the optimization goal. With AEQ for regularization, the placement optimized by η_{AEQ} achieves satisfactory 27.5 dB in mean AC; however, there is 2.7 dB loss compared with AC_{AEQ} , as a sacrifice to pursue higher η performance for more geometric robustness. Besides, η_{AEQ} achieves the best AE performance, reflecting the high energy efficiency of the sound reproduction system. The SPL of the reproduced sound field at 800 Hz and 2500 Hz with geometry optimized by AC_{AEQ} and η_{AEQ} are presented in Fig. 7. The geometric robustness achieved by maximizing

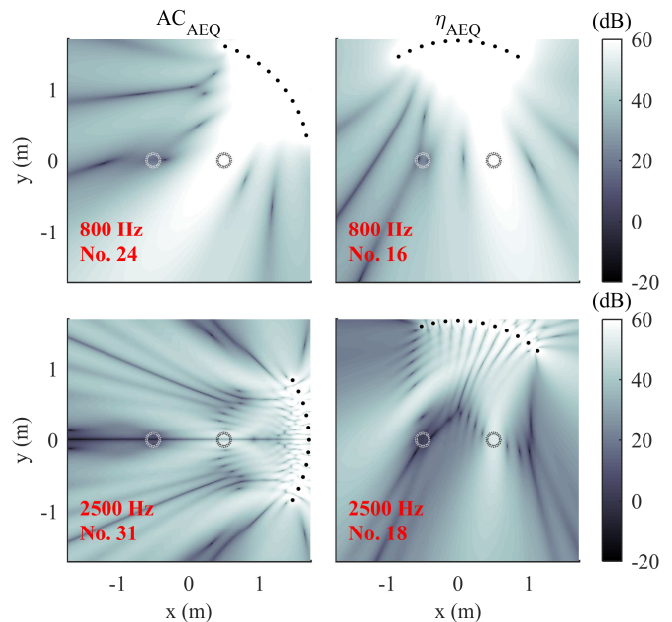


Fig. 7. Sound pressure level (SPL) of the reproduced sound fields with placement optimized by AC_{AEQ} and η_{AEQ} for acoustic contrast control. The black/white circles mark the listening/quiet zones. The frequencies and indexes of the selected placement are also marked.

η is explained by the following three major features based on comparisons among the placements optimized by η_{AEQ} and AC_{AEQ} .

Feature 1 — better avoidance of ‘occlusion’

In placement no. 31 the loudspeaker array and the controlled zones are located in a line, which is known as ‘occlusion’ [9], [32]. Among all the four approaches in Fig. 6, the placement obtained by η_{AEQ} avoids most of the ‘occlusions’ up to 3 kHz, where the grating lobe effect occurs [22]. The grating lobe effect determines the cut-off frequency of personal audio performance and is related to the spacing of loudspeakers. The ‘occlusion’ placement obtained by AC_{AEQ} starts from 2.5 kHz, above f_{zone} . The reproduced sound fields at 2.5 kHz with placements separately optimized by AC_{AEQ} and η_{AEQ} are

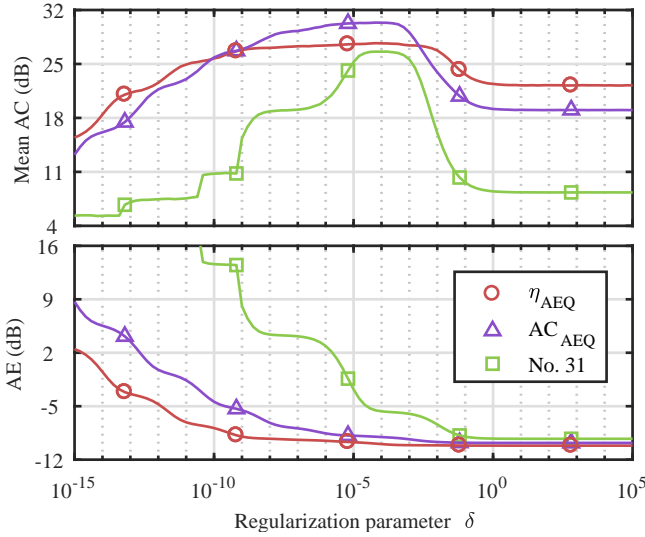


Fig. 8. The mean acoustic contrast (AC) and array effort (AE) performance of acoustic contrast control optimization applying different regularization ($\delta_L = \delta_Q = \delta$) into loudspeaker weight calculation at 800 Hz, separately using the placement optimized by AC_{AEQ} and η_{AEQ} , or the ‘occlusion’ placement.

TABLE V

BROADBAND ACOUSTIC CONTRAST (AC), REALISABILITY (η) AND ARRAY EFFORT (AE) PERFORMANCE AVERAGED OVER 100–3000 HZ.

Method	AC_{AEQ}	η_{AEQ}	$AC_{AEQ, EQ}$	$\eta_{AEQ, EQ}$
Mean AC (dB)	28.8	26.8	28.2	27.5
η	0.92	0.96	0.87	0.96
AE (dB)	-7.5	-8.6	-7.1	-8.7

visualized in Fig. 7, and their mean AC and η performance are separately 28.6 dB and 26.5 dB, 0.97 and 1.00. Despite 2.1 dB AC advantage, the ‘occlusion’ placement optimized by AC_{AEQ} suffers ‘standing wave’ that makes the listening be silenced at certain positions in the listening zone [33].

Feature 2 — improved robustness to regularization in the loudspeaker weight calculation

The η value of the placements in Fig. 8 are separately 0.62 (no. 31, ‘occlusion’), 0.90 (no. 24, optimized by AC_{AEQ}) and 0.98 (no. 16, optimized by η_{AEQ}). With the highest value of η approaching 1, the placement optimized by η_{AEQ} has the best AE performance and the least sensitivity to the selection of regularization parameter. Firstly, it enables the widest acceptable range for regularization parameter. Secondly, it has the flattest mean AC performance with the least fluctuation resulting from different regularization parameters. From Fig. 8, to achieve mean AC performance no less than 20 dB, the acceptable δ range separately using the placement optimized by η_{AEQ} , AC_{AEQ} and the placement of ‘occlusion’ are $\delta \geq 2.5 \times 10^{-14}$, $4.0 \times 10^{-13} - 1.0 \times 10^{-1}$ and $1.0 \times 10^{-6} - 4.0 \times 10^{-3}$, with mean AC variation of 7.7 dB, 6.6 dB and 10.4 dB. So the increase in geometric robustness makes it less demanding to improve the system’s robustness through accurate regularization in the loudspeaker weight calculation.

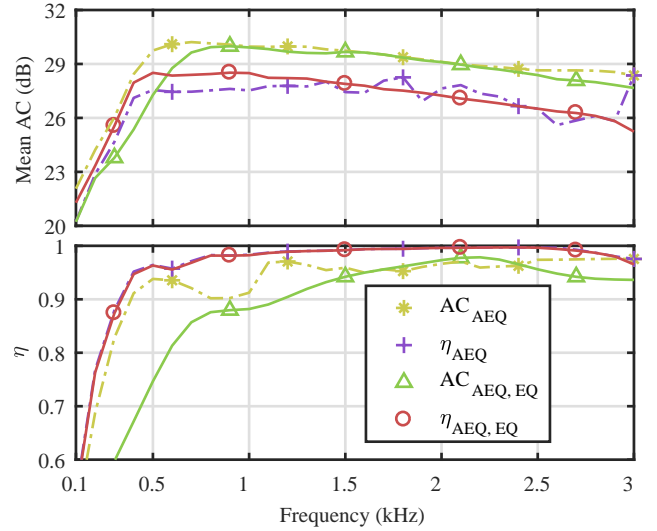


Fig. 9. The broadband performance of mean acoustic contrast (AC) and realisability (η) with placement separately optimized by AC_{AEQ} , η_{AEQ} , $AC_{AEQ, EQ}$ and $\eta_{AEQ, EQ}$. With the simple broadband equalization, $AC_{AEQ, EQ}$ and $\eta_{AEQ, EQ}$ are broadband uniform placement, separately averaged from the placements optimized by AC_{AEQ} and η_{AEQ} at 30 sampled frequencies within 100–3000 Hz.

Feature 3 — improved broadband equalization

Practical application pursues a uniform placement over a certain frequency band. In the simple broadband equalization, the placements $AC_{AEQ, EQ}$ and $\eta_{AEQ, EQ}$ in Fig. 9 are separately the average of the placements optimized by AC_{AEQ} and η_{AEQ} at 30 sampled frequencies within 100–3000 Hz. This frequency range avoids the grating lobe effect and covers main frequencies of speech and directional information delivery. The broadband performance averaged over sampled frequencies is presented in Table V. Compared to AC_{AEQ} , the placement of $AC_{AEQ, EQ}$ has reduced mean AC performance at low frequencies, and even lower than that of $\eta_{AEQ, EQ}$ below 500 Hz. On the η performance, $\eta_{AEQ, EQ}$ still achieves the maximal η (as η_{AEQ}) throughout the frequency band, together with better AE performance; while $AC_{AEQ, EQ}$ is affected by equalization and much worse than $\eta_{AEQ, EQ}$. However, the broadband mean AC of $AC_{AEQ, EQ}$ is barely 0.7 dB better than that of $\eta_{AEQ, EQ}$. The mean AC performance in Fig. 9 is evaluated with w calculation incorporating AEQ, a nearly ideal regularization. According to Fig. 8, the mean AC performance of $AC_{AEQ, EQ}$ might be worse than $\eta_{AEQ, EQ}$, when worse regularization is applied.

D. Placement Selection Results for PM and PC

Compared with ACC, PM and PC further modify the assumed capability of the listening zone modes towards an expected local sound orientation, with $\hat{U}_L^H P_{des}$ and $\hat{U}_L^H \Lambda^{\frac{1}{2}}$ in (13). With local directional optimization by PM or PC, the selected placement is more concentrated to a certain placement than that for ACC when $f \leq f_{zone}$. Thereby, ‘occlusion’ can be avoided by setting the desired local sound orientation. However, the geometry optimization with a desired direction of sound propagation over the listening zone is not always equivalent to selecting those speakers that radiate sound in the required direction. The proposed method serves to

automatically select the loudspeaker placement for optimized overall performance, balancing acoustic contrast, local sound orientation and robustness. Though not shown for conciseness and clarity, we observed that η_{AEQ} is better than J_{ELM} , J_{ELO} and J_{AEQ} as selection metric in the placement selection for PM and PC, with improved overall performance in terms of acoustic contrast, local sound orientation, array effort and robustness. The placement optimized by η_{AEQ} for PM and PC still held the features on the improved robustness to regularization in the loudspeaker weight calculation and the improved broadband equalization.

E. Discussion

The realisability measure of sound zone reproduction was initially investigated in [32, Eq. 12] and formulated in the wave-domain using spherical harmonics as the ratio of the sound leakage in the quiet zone and the acoustic energy in the listening zone. However, the proposed η metric has totally different formulation to measure the compatibility of the SVD-based modal capabilities to ‘naturally’ reproduce sound zones. Furthermore, η is further applied in geometry optimization, which results in three preferable features due to geometric robustness through simulation observations. Though not shown for conciseness and clarity, it is observed that the proposed method with ACC optimization still held the three features due to improved robustness, under the same simulated reverberant room environment with [32].

V. CONCLUSION

Considering the perturbations in real world sound reproduction, robustness is highly desirable in personal audio design. Regularization in the loudspeaker weight calculation is used to increase robustness. However, the selection of regularization parameter is not always straightforward. In this paper, we first formulated the personal audio in the SVD-based modal domain, demonstrated the effect of regularization as the modification of the assumed capabilities of SVD-based listening and quiet zone modes, and suggested to apply AEQ regularization to the state-of-the-art placement selection using AC as metric, rather than the array effort based regularization (ELO and ELM). To achieve the optimal geometric robustness, a realisability measure η was further proposed as a metric in geometry optimization. The simulation observations applied to ACC, PM and PC show that the proposed method obtains the geometry with better avoidance of ‘occlusion’, improved robustness to the value of regularization parameter, and improved broadband equalization. Future work includes investigating the effect of the proposed geometry optimization on the time-domain filter design, and its applications into other personal audio optimizations and reverberant room environments.

APPENDIX

This appendix introduces the derivation of the three personal audio optimizations in the SVD-based modal domain listed in Table II along with (12). The derivation based on the regularized personal audio solutions listed in Table I. The relevant math derivation can be found

in [16] and its supplementary https://www.researchgate.net/publication/328737949_Supplementary_Material_-_Robust_reproduction_of_sound_zones_with_local_sound_orientation.

Applying SVD decomposed transfer function matrices in (5), the basic terms of the solutions are reformulated into

$$\begin{cases} \mathbf{G}_Q^H \mathbf{G}_Q + \delta_Q \mathbf{I} = \mathbf{V}_Q \mathbf{\Omega}_Q \mathbf{V}_Q^H, \\ \mathbf{G}_L^H \mathbf{G}_L + \delta_L \mathbf{I} = \mathbf{V}_L \mathbf{\Omega}_L \mathbf{V}_L^H, \\ \mathbf{G}_L^H \mathbf{P}_{\text{des}} = \mathbf{V}_L \mathbf{\Sigma}_L \mathbf{U}_L^H \mathbf{P}_{\text{des}}, \\ \mathbf{G}_L^H \mathbf{\Lambda} \mathbf{G}_L + \zeta_L \mathbf{I} = \mathbf{V}_L \mathbf{\Pi}_L \mathbf{V}_L^H. \end{cases} \quad (30)$$

Instituting (30) into the regularized ACC solution, then

$$\mathbf{w}_{\text{ACC}} = \mathbf{\Phi}(\mathbf{V}_Q \mathbf{\Omega}_Q^{-1} \mathbf{V}_Q^H \mathbf{V}_L \mathbf{\Omega}_L \mathbf{V}_L^H). \quad (31)$$

Referring to (10), $\mathbf{V}_L = \mathbf{V}_Q \mathbf{T}^H$, then the expression without \mathbf{V}_L is

$$\mathbf{w}_{\text{ACC}} = \mathbf{\Phi}(\mathbf{V}_Q \mathbf{\Omega}_Q^{-1} \mathbf{T}^H \mathbf{\Omega}_L \mathbf{T} \mathbf{V}_Q^H). \quad (32)$$

According to the appendix in [34],

$$\mathbf{w}_{\text{ACC}} = \mathbf{V}_Q \mathbf{\Phi}(\mathbf{\Omega}_Q^{-1} \mathbf{T}^H \mathbf{\Omega}_L \mathbf{T}), \quad (33)$$

then $\mathbf{k}_Q = \mathbf{V}_Q^H \mathbf{w}_{\text{ACC}} = \mathbf{\Phi}(\mathbf{\Omega}_Q^{-1} \mathbf{T}^H \mathbf{\Omega}_L \mathbf{T})$. In the same way, \mathbf{k}_L for ACC can be derived from the expression of \mathbf{w}_{ACC} without \mathbf{V}_Q , by instituting $\mathbf{V}_Q = \mathbf{V}_L \mathbf{T}$ into (31). Similarly, \mathbf{k}_Q and \mathbf{k}_L for PM and PC can be derived.

ACKNOWLEDGEMENT

This research is currently supported under the Australian Research Council’s Linkage Project funding scheme (LP160100616), and was supported by the UCAS Joint PhD Training Program (UCAS[2015]31) funded by the University of Chinese Academy of Sciences when Q. Zhu was with Institute of Acoustics, Chinese Academy of Sciences and University of Chinese Academy of Sciences before she joined University of Technology Sydney.

REFERENCES

- [1] T. Betlehem, W. Zhang, M. Poletti, and T. D. Abhayapala. Personal sound zones: Delivering interface-free audio to multiple listeners. *IEEE Sig. Proc. Mag.*, 32(2):81–91, 2015.
- [2] M. F. Simón-Gálvez, S. J. Elliott, and J. Cheer. Personal audio loudspeaker array as a complementary tv sound system for the hard of hearing. *IEICE Trans. Fundamentals*, 97(9):1824–1831, 2014.
- [3] J.-H. Chang, C.-H. Lee, J.-Y. Park, and Y.-H. Kim. A realization of sound focused personal audio system using acoustic contrast control. *J. Acoust. Soc. Am.*, 125(4):2091–2097, 2009.
- [4] J. Cheer, S. J. Elliott, Y. Kim, and J.-W. Choi. Practical implementation of personal audio in a mobile device. *J. Audio Eng. Soc.*, 61(5):290–300, 2013.
- [5] J. Cheer, S. J. Elliott, and M. F. Simón-Gálvez. Design and implementation of a car cabin personal audio system. *J. Audio Eng. Soc.*, 61(6):412–424, 2013.
- [6] J. G. Apostolopoulos, P. A. Chou, B. Culbertson, T. Kalker, M. D. Trott, and S. Wee. The road to immersive communication. *Proc. IEEE*, 100(4):974–990, 2012.
- [7] W. F. Druyvesteyn and J. Garas. Personal sound. *J. Audio Eng. Soc.*, 45(9):685–701, 1997.
- [8] J.-W. Choi and Y.-H. Kim. Generation of an acoustically bright zone with an illuminated region using multiple sources. *J. Acoust. Soc. Am.*, 111(4):1695–1700, 2002.
- [9] M. Poletti. An investigation of 2-D multizone surround sound systems. In *125th AES Convention, San Francisco*, October Oct., 2008.

- [10] S. J. Elliott, J. Cheer, J.-W. Choi, and Y. Kim. Robustness and regularization of personal audio systems. *IEEE Trans. Audio, Speech, Language Process.*, 20(7):2123–2133, 2012.
- [11] P. Coleman, P. J. B. Jackson, M. Olik, and J. A. Pedersen. Personal audio with a planar bright zone. *J. Acoust. Soc. Am.*, 136(4):1725–1735, 2014.
- [12] S. Zhao, Y. Hu, J. Lu, X. Qiu, J. Cheng, and I. Burnett. Delivering sound energy along an arbitrary convex trajectory. *Scientific Reports*, 4, 2014.
- [13] K. Helwani, S. Spors, and H. Buchner. The synthesis of sound figures. *Multidim. System Signal Process.*, 25(2):379–403, 2014.
- [14] P.-A. Gauthier, Y. Pasco, and A. Berry. Generalized singular value decomposition for personalized audio using loudspeaker array. In *Audio Engineering Society Conference: 2016 AES International Conference on Sound Field Control*. Audio Engineering Society, 2016.
- [15] F. Olivieri, F. M. Fazi, S. Fontana, D. Menzies, and P. A. Nelson. Generation of private sound with a circular loudspeaker array and the weighted pressure matching method. *IEEE/ACM Trans. Audio, Speech, Language Process.*, 25(8):1579–1591, 2017.
- [16] Q. Zhu, P. Coleman, M. Wu, and J. Yang. Robust reproduction of sound zones with local sound orientation. *J. Acoust. Soc. Am.*, 142(1):EL118–EL122, 2017.
- [17] 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.*, 134(1):EL112–EL118, 2013.
- [18] P. Coleman. *Loudspeaker Array Processing for Personal Sound Zone Reproduction*. PhD thesis, University of Surrey, Guildford, UK, 2014.
- [19] H. Cox, R. M. Zeskind, and M. M. Owen. Robust adaptive beamforming. *IEEE Trans. Acoust., Speech, Signal Process.*, 35(10):1365–1376, 1987.
- [20] S. Doclo and M. Moonen. Design of broadband beamformers robust against gain and phase errors in the microphone array characteristics. *IEEE Trans. Signal Process.*, 51(10):2511–2526, 2003.
- [21] T. Betlehem and C. Withers. Sound field reproduction with energy constraint on loudspeaker weights. *IEEE Trans. Audio, Speech, Language Process.*, 20(8):2388–2392, 2012.
- [22] Q. Zhu, P. Coleman, M. Wu, and J. Yang. Robust acoustic contrast control with reduced in-situ measurement by acoustic modeling. *J. Audio Eng. Soc.*, 65(6):460–473, 2017.
- [23] D. B. Ward and G. W. Elko. Optimum loudspeaker spacing for robust crosstalk cancellation. In *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*, volume 6, pages 3541–3544. IEEE, 1998.
- [24] D. B. Ward and G. W. Elko. Effect of loudspeaker position on the robustness of acoustic crosstalk cancellation. *IEEE Signal Process. Lett.*, 6(5):106–108, 1999.
- [25] T. Takeuchi and P. A. Nelson. Optimal source distribution for binaural synthesis over loudspeakers. *J. Acoust. Soc. Am.*, 112(6):2786–2797, 2002.
- [26] N. Radmanesh and I. S. Burnett. Generation of isolated wideband sound fields using a combined two-stage lasso-ls algorithm. *IEEE Trans. Audio, Speech, Language Process.*, 21(2):378–387, 2013.
- [27] W. Jin and W. B. Kleijn. Theory and design of multizone soundfield reproduction using sparse methods. *IEEE/ACM Trans. Audio, Speech, Language Process.*, 23(12):2343–2355, 2015.
- [28] J. Francombe, P. Coleman, M. Olik, K. Baykaner, P. Jackson, R. Mason, M. Dewhurst, S. Bech, and J. A. Pederson. Perceptually optimized loudspeaker selection for the creation of personal sound zones. In *Audio Engineering Society Conference: 52nd International Conference: Sound Field Control-Engineering and Perception*. Audio Engineering Society, 2013.
- [29] B. D. V. Veen and K. M. Buckley. Beamforming: A versatile approach to spatial filtering. *IEEE ASSP Mag.*, 5(2):4–24, 1988.
- [30] P. J. B. Jackson, f. Jacobsen, P. Coleman, and J. A. Pedersen. Sound field planarity characterized by superdirective beamforming. *Proc. Meetings on Acoustics*, 19(1):055056, 2013.
- [31] F. M. Fazi. *Sound Field Reproduction*. PhD thesis, University of Southampton, 2010.
- [32] W. Zhang, T. D. Abhayapala, T. Betlehem, and F. M. Fazi. Analysis and control of multi-zone sound field reproduction using modal-domain approach. *J. Acoust. Soc. Am.*, 140(3):2134–2144, 2016.
- [33] 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.*, 135(4):1929–1940, 2014.
- [34] S. Shahbazpanahi, A. B. Gershman, Z.-Q. Luo, and K. W. Wong. Robust adaptive beamforming for general-rank signal models. *IEEE Trans. Signal Process.*, 51(9):2257–2269, 2003.