

Personal Sound Zones



[Delivering interface-free audio to multiple listeners]

Sound rendering is increasingly being required to extend over certain regions of space for multiple listeners, known as *personal sound zones*, with minimum interference to listeners in other regions. In this article, we present a systematic overview of the major challenges that have to be dealt with for multizone sound control in a room. Sound control over multiple zones is formulated as an optimization problem, and a unified framework is presented to compare two state-of-the-art sound control techniques. While conventional techniques have been focusing on point-to-point audio processing, we introduce a wave-domain sound field representation and active room compensation for sound pressure control over a region of space. The design of directional loudspeakers is presented and the advantages of using arrays of directional sources are illustrated for sound reproduction, such as better control of sound fields over wide areas and reduced total number of loudspeaker units, thus making it particularly suitable for establishing personal sound zones.

INTRODUCTION

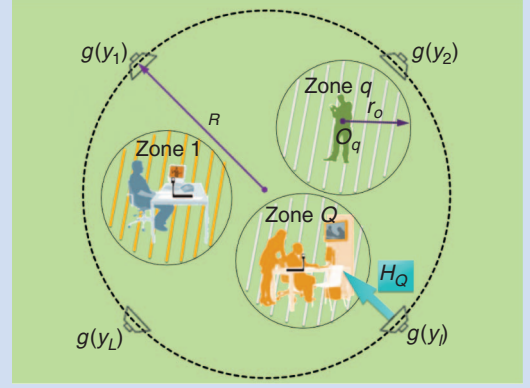
Sound recording and sound reproduction are becoming increasingly ubiquitous in our daily lives. The ultimate goal of sound

reproduction is to recreate the full richness of a sound field including not only the sound content but also the spatial properties to give the listener full knowledge about both the sound source and acoustic environment. Spatial sound reproduction technologies so far have made tremendous progress in reproducing sound fields over fairly large regions of space using an array of loudspeakers. This introduces the idea of establishing personal sound zones, whereby interface-free audio is delivered to multiple listeners in the same environment without physical isolation or the use of headphones (Figure 1). This concept has recently drawn attention due to a whole range of audio applications, from controlling sound radiation from a personal audio device to creating individual sound zones in all kinds of enclosures (such as shared offices, passenger cars, and exhibition centers) and generating quiet zones in noisy environments.

The first known demonstration of reproducing a sound field within a given region of space was conducted by Camras at the Illinois Institute of Technology in 1967, where loudspeakers were distributed on the surface enclosing the selected region to control sound radiation, and the listeners could move freely within the recreated environment [1]. The well-known ambisonics [2], wave field synthesis [3], and higher-order spherical harmonics-based techniques [4] were developed separately for more advanced spatial sound field reproduction over a large region of space. Druyvesteyn and Garas [5] first proposed the



(a)



(b)

[FIG1] (a) An illustration of personal sound zones in an office environment. (b) A loudspeaker array is used to create multiple sound zones for multiple listeners.

concept of a personal sound zone, i.e., reproducing sound within a desired region of space with a reduced sound level elsewhere. Microsoft researchers later demonstrated their “Personal Audio Space” project at Microsoft Research TechFest 2007, where a linear loudspeaker array consisting of 16 drivers was used to enhance the sound in one area while canceling sound waves in another area within the same physical space. By stepping even a few paces outside the target region, users reported that they could not hear the reproduced music. Researchers further extended this concept to develop personal audio for personal computers and televisions [6], as well as for mobile devices [7] and automobile cabins [8].

A way to create personal sound zones is to formulate a multizone sound control problem within the same physical space as illustrated in Figure 1. Here, multiple microphones and loudspeakers are used to control the reproduced sound fields. A preference is to use a single array of loudspeakers rather than separate arrays for each zone. This improves freedom and flexibility, allowing sound zones to be positioned dynamically and listeners to freely move between zones. When the system is implemented in reverberant enclosures, loudspeaker designs and audio processing are two key aspects to control sound radiation and to deal with the complexity and uncertainty associated with sound field reproduction. This article aims at reviewing these techniques to support the goal of establishing personal sound zones.

MULTIZONE SOUND CONTROL

In a general formulation, sound fields are produced over Q sound zones. Here M pressure controlling microphones are placed within each zone so that the zone sound fields are controlled by a total of QM matching points. The sound pressures measured at the microphone positions in each zone q are represented as a vector $\mathbf{p}_q = [p(x_{q,1}, \omega), \dots, p(x_{q,M}, \omega)]^T$ and given by

$$\mathbf{p}_q = \mathbf{H}_q \mathbf{g}, \quad (1)$$

where $\mathbf{g} = [g(y_1, \omega), \dots, g(y_L, \omega)]^T$ denotes the vector of loudspeaker driving signals at a given frequency ω to create personal audio sound scenes and \mathbf{H}_q represents a matrix of acoustic transfer functions (or acoustic impedances) between the loudspeaker drivers and the microphones in zone q . Sound control techniques can broadly be classified into two categories, acoustic contrast control (ACC) and pressure matching (PM), and we consider each in turn.

ACOUSTIC CONTRAST CONTROL

Choi and Kim [9] first formulated the personal audio problem by creating two kinds of sound zones: the bright zone within which we want to reproduce certain sounds with high acoustic energy, and the dark zone (or the quiet zone) within which the acoustic energy is kept at a low level. The principle of ACC is to maximize the contrast in the acoustic energy between the bright zone and the dark zone. Among the Q sound zones, we specify the first zone as the bright zone and the remaining $Q - 1$ zones as the dark zones. The acoustic energy in the bright zone is defined from the sound pressures measured at the M matching points, that is $E_b = \|\mathbf{p}_b\|^2 = \|\mathbf{H}_b \mathbf{g}\|^2$ with $\mathbf{H}_b = \mathbf{H}_1$ and $\|\cdot\|$ denoting the ℓ_2 norm. Similarly, the acoustic energy in the dark zones is represented as $E_d = \|\mathbf{p}_d\|^2 = \|\mathbf{H}_d \mathbf{g}\|^2$ with $\mathbf{H}_d = [\mathbf{H}_2^H, \dots, \mathbf{H}_Q^H]^H$ and $(\cdot)^H$ represents the Hermitian transpose.

In [9], the acoustic contrast, defined as a ratio between the average acoustic potential energy density produced in the bright zone to that in the dark zones, is maximized. The acoustic contrast maximizing method may perform well over the dark zones but may be unrobust to providing the desired maximum energy in the bright zone. To ensure the sound energy within different zones are optimized simultaneously, the problem can be reformulated as maximizing the acoustic energy in the bright zone with the constraint that the energy in the dark zone is limited to a very small value D_0 . In addition, a limit is imposed on the loudspeaker power consumption, i.e., $\|\mathbf{g}\|^2 \leq E_0$, also known as the *array effort*. These constraints ensure that sound leakage outside the Q zones is not excessive and also that realized

loudspeaker weights are chosen to ensure the implementation is robust to driver positioning errors and changes in the acoustic environment. The ACC problem can then be posed as

$$\max_{\mathbf{g}} \|\mathbf{H}_b \mathbf{g}\|^2 \quad (2a)$$

$$\text{subject to } \|\mathbf{H}_d \mathbf{g}\|^2 \leq D_0 \quad (2b)$$

$$\|\mathbf{g}\|^2 \leq E_0. \quad (2c)$$

The objective and the constraints are summarized into a single objective function represented using the Lagrangian [10],

$$\max_{\mathbf{g}} L_c(\mathbf{g}) = \|\mathbf{H}_b \mathbf{g}\|^2 - \lambda_1 (\|\mathbf{H}_d \mathbf{g}\|^2 - D_0) - \lambda_2 (\|\mathbf{g}\|^2 - E_0), \quad (3)$$

$$\lambda_1, \lambda_2 \geq 0,$$

where λ_1 and λ_2 are Lagrange multipliers to adjust the relative importance of each condition (2b) and (2c). The solution that maximizes the Lagrangian is obtained by taking the derivative of L_c with respect to \mathbf{g} and equating it to zero, and is written as

$$\lambda_1 [\mathbf{H}_d^H \mathbf{H}_d + \frac{\lambda_2}{\lambda_1} \mathbf{I}] \mathbf{g} = [\mathbf{H}_b^H \mathbf{H}_b] \mathbf{g}, \quad (4)$$

which is recognized as a generalized eigenvector problem. The optimum source strength vector \mathbf{g}_c is set as the eigenvector corresponding to the maximum eigenvalue of the matrix $[\mathbf{H}_d^H \mathbf{H}_d + (\lambda_2/\lambda_1) \mathbf{I}]^{-1} [\mathbf{H}_b^H \mathbf{H}_b]$. The ratio of Lagrange multipliers $\lambda = \lambda_2/\lambda_1$ determines the tradeoff between the performance and array effort and must be chosen iteratively for the constraint on the array effort to be satisfied. The formulation in (4) yields essentially the same answer as that in [8], or the so-called indirect formulation in [10], which diagonally loads the matrix $\mathbf{H}_d^H \mathbf{H}_d$ before inverting it to improve the matrix condition number.

The formulation adopted here leads to a straightforward way for demonstrating the connection between the ACC method and the PM method, which will be explained next.

PRESSURE MATCHING

The PM method aims to reproduce a desired sound field in the bright zone at full strength, while producing silence in other zones. The idea comes from the traditional crosstalk-cancellation problem, where small regions of personal audio are created by controlling the pressure at discrete spatial points (microphone or listener positions). Multizone sound control is an extension of the traditional approach with a sufficiently dense distribution of matching points within all the zones. Given a target sound field \mathbf{p}_{des} to be reproduced in the bright zone, the robust PM formulation can be written using an ℓ_2 PM objective along with the constraints on the sound energy in the dark zones and the array effort constraint,

$$\min_{\mathbf{g}} \|\mathbf{H}_b \mathbf{g} - \mathbf{p}_{\text{des}}\|^2 \quad (5a)$$

$$\text{subject to } \|\mathbf{H}_d \mathbf{g}\|^2 \leq D_0 \quad (5b)$$

$$\|\mathbf{g}\|^2 \leq E_0. \quad (5c)$$

The problem can then be written as a Lagrangian cost function,

$$\min_{\mathbf{g}} L_p(\mathbf{g}) = \|\mathbf{H}_b \mathbf{g} - \mathbf{p}_{\text{des}}\|^2 + \lambda_1 (\|\mathbf{H}_d \mathbf{g}\|^2 - D_0) + \lambda_2 (\|\mathbf{g}\|^2 - E_0), \quad (6)$$

$$\lambda_1, \lambda_2 \geq 0,$$

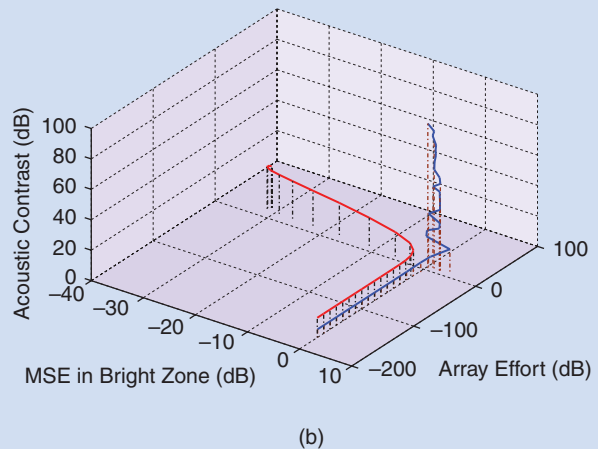
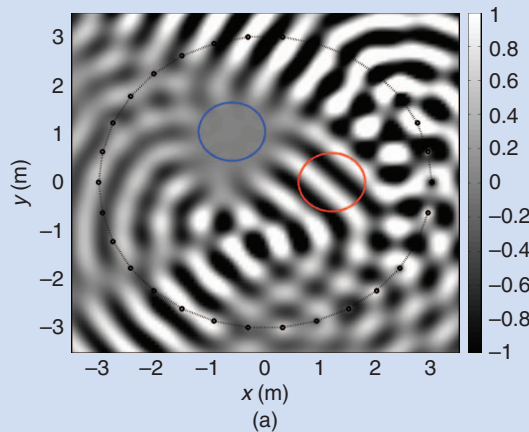
where again λ_1 and λ_2 are Lagrange multipliers. The solution that minimizes L_p is obtained by setting the derivative of L_p with respect to \mathbf{g} to zero and is written as

$$[\mathbf{H}_b^H \mathbf{H}_b + \lambda_1 \mathbf{H}_d^H \mathbf{H}_d + \lambda_2 \mathbf{I}] \mathbf{g} = \mathbf{H}_b^H \mathbf{p}_{\text{des}}. \quad (7)$$

Equation (7) may be solved using an interior point algorithm to choose appropriate values of λ_1 and λ_2 to satisfy the constraints [11]. A simpler formulation is to set the parameter $\lambda_1 = 1$, which implies applying equal effort to matching the pressure in the bright zone and minimizing the energy in the dark zone. This gives the original formulation of multizone sound control as in [12] but has an added robustness constraint on the array effort, that is $\mathbf{g}_p = [\mathbf{H}_b^H \mathbf{H}_b + \mathbf{H}_d^H \mathbf{H}_d + \lambda_2 \mathbf{I}]^{-1} \mathbf{H}_b^H \mathbf{p}_{\text{des}}$. This solution is also identical to that of the ACC method provided that 1) the choice of target pressures in the bright zone is an ACC solution, $\mathbf{p}_{\text{des}} = \mathbf{H}_b \mathbf{g}_c$ and 2) identical constraints in E_0 and D_0 are met. This demonstrates that the formulation in the PM approach to sound field reproduction subsumes the ACC problem. Chang and Jacobsen [13] investigated a combined solution of these two techniques, $\mathbf{g}_{\text{cb}} = [(1 - \kappa) \mathbf{H}_b^H \mathbf{H}_b + \kappa \mathbf{H}_d^H \mathbf{H}_d]^{-1} (1 - \kappa) \mathbf{H}_b^H \mathbf{p}_{\text{des}}$, which is actually same as the one presented in (7) with the regularization term omitted. The tuning parameter κ is equivalent to the tuning parameter λ_1 . The design has been shown effective for reproducing plane wave sound fields at frequencies even above the Nyquist frequency with good contrast control, thus providing the potential to reduced the number of loudspeakers required and increase the zone sizes and upper operating frequencies using the PM method.

The PM approach gives an explicit solution to obtain the loudspeaker driving signals and does not require solving an eigenvector problem, as is required in the case of acoustic contrast optimization. PM is especially suitable for the situation that different constraints are imposed on each sound zone when the listeners require different quality of listening experiences. However a series of Lagrange multipliers need to be determined, and a generalized eigenvalue solution is no longer possible. Instead convex-optimization methods like the interior-point method should be used [11]. The PM approach also imposes an objective on the phase of reproduced sound fields within the bright zone, and thus provides a better holographic image compared to the contrast control method. Figure 2(b) demonstrates that the ACC method always maintains a high level of contrast between the bright and dark zone using a small array effort, but a high reproduction error also indicates that the reproduced sound field may swirl around the listener in different directions. On the other hand, the PM approach achieves small reproduction error while higher contrast may be obtained by choosing an appropriate desired sound field. Preliminary perceptual tests were found to generally agree with the simulation results however the source signal itself significantly affects the system performance [14].

While the least squares solutions in the frequency domain seems to provide a great deal of simplicity and flexibility, the



[FIG2] A plane wave of 500 Hz from 45° is reproduced in the bright zone (red circle) using PM while deadening the sound in the dark zone (blue circle) using 30 loudspeakers placed on a circle of radius $R = 3$ m, and each zone is of radius $r = 0.6$ m as shown in (a). (b) The acoustic contrast versus the array effort and the mean-square reproduction error in the bright zone using the ACC method (blue line) and the PM method (red line).

positions of the loudspeakers and the matching points within sound zones must be chosen judiciously for good reproduction performance. Representing sound fields in the wave domain or mode domain as in (S1) in “Wave-Domain Sound Field Representation” can provide physical insights into these critical issues [15]. Dimensionality analysis tells us that for PM over Q sound zones, the number of loudspeakers required is determined by the upper frequency or wave number k of operation, the number of sound zones, and the size of each sound zone [15]. Here we assume that each sound zone is a circle or sphere of radius r_0 located at the origin O_q as shown in Figure 1, although without loss of generality each sound zone could be of arbitrary shape. The minimum number L is about $Q(2kr_0 + 1)$ for two-dimensional (2-D) reproduction and $Q(kr_0 + 1)^2$ for three-dimensional (3-D) reproduction, respectively [4].

DISCUSSION

Practical Implementation

When a small number of loudspeakers are used, for example, three speakers used in a mobile device, current personal audio systems can only achieve limited performance, i.e., ~ 10 dB contrast level between bright and dark zones [7]. An array of nine sources has been implemented for personal audio systems in televisions and personal computers, in an anechoic chamber achieving over 19 dB contrast under ideal conditions [6]. However, in terms of practical implementation in a car cabin, Cheer et al. [8] demonstrated that the optimized level of acoustic contrast obtained from the ACC method may not be achieved because of errors and uncertainties and the least-squares-based PM approach may provide a more robust solution. In addition, multizone reproduction is fundamentally constrained whenever attempting to reproduce a sound field in the bright zone that is directed to or obscured by another zone. This is known as the *occlusion problem* [11], [12].

Loudspeaker Positions

Using the compressive sensing idea, the formulation of multizone sound field reproduction can be regularized with the ℓ_1 norm of

WAVE-DOMAIN SOUND FIELD REPRESENTATION

The Helmholtz wave equation can be solved to express any sound field as a weighted sum of basis functions,

$$p(\mathbf{x}, \omega) = \sum_{n=1}^{\infty} \alpha_n(\omega) \beta_n(\mathbf{x}, \omega), \quad (\text{S1})$$

where $\alpha_n(\omega)$ are sound field coefficients corresponding to mode index n , $\beta_n(\mathbf{x}, \omega)$ are basis functions with the orthogonality property

$$\langle \beta_n, \beta_m \rangle \triangleq \int_{\mathbb{C}} \beta_n^*(\mathbf{x}, \omega) \beta_m(\mathbf{x}, \omega) d\mathbf{x} = \xi_n(\omega) \delta_{nm}.$$

The sound field within a control region \mathbb{C} can be represented using a finite number of basis functions, i.e., $n \in [1, N]$ and $\xi_n(\omega) = \langle \beta_n, \beta_n \rangle$ is the strength of each mode over the control zone.

The modal basis functions for source distributions and sound fields expressed in cylindrical coordinates and spherical coordinates can be written as [17]

$$p_{2D}(\mathbf{x}, \omega) = \sum_{v=-N}^N \alpha_v(\omega) \mathcal{J}_v^{(2D)}(kr) \exp(iv\phi) \quad (\text{S2a})$$

$$p_{3D}(\mathbf{x}, \omega) = \sum_{v=0}^N \sum_{\mu=-v}^v \alpha_{v\mu}^{\mu}(\omega) \mathcal{Y}_v^{(3D)}(kr) Y_v^{\mu}(\theta, \phi), \quad (\text{S2b})$$

where $\exp(\cdot)$ and $Y_v^{\mu}(\cdot)$ are complex exponentials and spherical harmonics, respectively and $\mathcal{J}_v^{(2D)}(kr)$ and $\mathcal{Y}_v^{(3D)}(kr)$ are functions representing the 2-D and 3-D mode amplitudes at radius r , wave number k , and the truncation number $N \approx kr_0$ [4], we have the following dimensionality results: $N_{2D} = 2kr_0 + 1$ and $N_{3D} = (kr_0 + 1)^2$. This gives the Nyquist sampling condition for a uniform circular array ($M \geq N_{2D}$) and a spherical array ($M \geq N_{3D}$), respectively.

the loudspeaker weights and solved using the least-absolute shrinkage and selection operator [16]. The assumption here is that the desired sound field can be reproduced by a few loudspeakers, which are placed close to the direction of the virtual source and are sparsely distributed in space. This can produce low sound levels outside the bright zones and hence can reduce the interference to the dark zone.

Further Remarks

While the reproduction error has been widely used to quantify the performance of sound field rendering methods, a planar wavefront may be reproduced whose direction of propagation does not coincide with the desired direction, which may give a high reproduction error. In [18], the cost function of the ACC method is refined to optimize the extent to which a sound field resembles a plane wave. A constraint is imposed on the plane-wave energy within the bright zone over a range of incoming directions, thus optimizing the spatial aspects of the sound field for ACC. Simulation results demonstrate that a circular array of 48 equally spaced loudspeakers produces consistently high contrast and a planar target sound zone of radius 0.15 m for frequencies up to 7 kHz.

ACTIVE ROOM COMPENSATION

One challenge in the personal audio problem is room reverberation. A strong wall reflection may ruin the personal audio listening experience [14]. Room reverberation can be corrected by using active room compensation, provided the acoustic transfer function (ATF) matrices are determined. For static room environments these matrices may be premeasured but for time-varying environments they must be determined adaptively. In this section, methods for determining and correcting for these matrices to compensate for room responses over spatial regions are described.

The room compensation approaches described here are more robust at low frequencies. At high frequencies, a reverberant sound field is diffuse. Compensation is extremely sensitive to small changes within the room and cannot be practically compensated for without very fast filter adaptation. Personal sound systems may not be able to compensate for these variations. Instead, diffuse components may be treated as noise and the system made robust to them.

We summarize the advances made for the case of a single zone with the ATF matrix, $H \equiv H_1$, using wave-domain or modal-space processing. These approaches demonstrate the challenges inherent in applying room compensation to the multizone problem. We also review a crosstalk-cancellation approach to the multizone case that utilizes impulse response reshaping.

MODAL-SPACE PROCESSING

Based on the wave-domain sound field representation (S1), the sound field $p(x, \omega)$ can be expressed as in (3). The ATF $H_\ell(x, \omega)$ from each loudspeaker ℓ to a point x inside the sound control zone can also be parameterized as

$$H_\ell^{(2D)}(x, \omega) = \sum_{\nu=-N}^N \gamma_{\nu\ell}(\omega) \mathcal{J}_\nu^{(2D)}(kr) \exp(i\nu\phi), \quad (8a)$$

$$H_\ell^{(3D)}(x, \omega) = \sum_{\nu=0}^N \sum_{\mu=-\nu}^{\nu} \gamma_{\nu\ell}^\mu(\omega) \mathcal{J}_\nu^{(3D)}(kr) Y_\nu^\mu(\theta, \phi), \quad (8b)$$

where $\gamma_{n\ell}(\omega)$ and $\gamma_{\nu\ell}^\mu(\omega)$ are ATF coefficients. The sound pressure vector p and ATF matrix H can then be written in matrix form

$$p = B\alpha, \quad (9a)$$

$$H = B\Gamma, \quad (9b)$$

where B is the $M \times N$ matrix of basis functions evaluated at each of the M microphone positions defined $[B]_{mn} = \beta_n(x_m, \omega)$, α is an M -long vector of sound field coefficients, Γ is the $N \times L$ matrix of the ATF coefficients defined $[\Gamma]_{n\ell} = \gamma_{n\ell}$, and N is either N_{2D} or N_{3D} . The PM problem of (5a) in the mode domain becomes $\Gamma g = \alpha_{\text{des}}$, where α_{des} is the N -long vector of coefficients for the desired sound field. The compensation problem can then be solved in offline manner by determining the least-squares solution [19].

An adaptive mode-domain approach was devised in [20]. The ATF matrix can be further parameterized

$$H = U\mathbf{\Gamma}, \quad (10)$$

where U is a tall Vandermonde matrix (2-D) or spherical harmonic matrix (3-D) with the property that $U^H U = I$ and J is a diagonal matrix of the mode amplitudes at the microphone positions. The vector of microphones' signals $p = Hg$ are hence transformed into mode-domain coefficients through $\alpha = J^{-1} U^H p$. For modest levels of room reverberation, Γ can be expressed as the sum of an anechoic room component and a small reverberant component. By approximating the reverberation as small, a simple iterative procedure for choosing g to drive α to α_{des} can be formulated. Reverberant compensation methods [19], [20] may have difficulties in practice with preringing artefacts, but these artefacts may be reduced by using more advanced multiple-input, multiple-output polynomial filter designs [21].

ACTIVE LISTENING ROOM COMPENSATION WITH WAVE-DOMAIN ADAPTIVE FILTERING

Active listening room compensation can be used to make a reverberant room problem look like an anechoic room problem [22]. By applying a compensation filter matrix to the input loudspeaker signals, the uncompensated anechoic-room driving signals can then be used. The essence of the problem as depicted in Figure 3 is to minimize the error energy $e^H e$, where

$$e = H_0 g - H C g,$$

H_0 is the anechoic-room ATF matrix, and C is an $L \times L$ compensation filter matrix. This effectively chooses the filter matrix C to drive the net transfer function matrix HC to the anechoic-room ATF matrix H_0 .

In massive multichannel problems for which the number of loudspeakers L and microphones M are large, the resultant

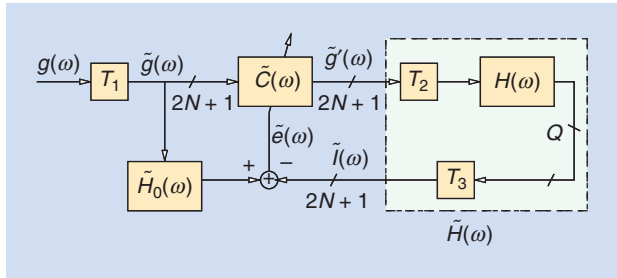
matrices are large and may have issues with computational requirements (for filtered x-RLS) and convergence rates (for filtered x-LMS). Poor convergence can be solved using eigenspace adaptive filtering [22] by performing a generalized singular value decomposition (SVD) to diagonalize the system. Unfortunately the SVD still incurs a high computational cost.

Fortunately, the problem can be solved computationally and efficiently by using a wave-domain approach. If the microphones are arranged over a uniform circular array of radius r and the sources are arranged over a concentric uniform circular array, then the anechoic-room ATF matrix may be parameterized

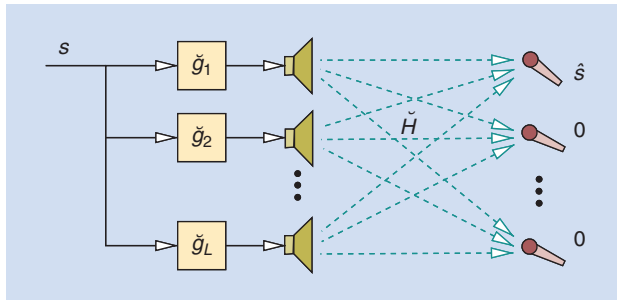
$$H_0 = UJ \underbrace{K^H V^H}_{\Gamma_0}, \quad (11)$$

where Γ_0 is a matrix of ATF coefficients corresponding to the anechoic room, K is a diagonal matrix of Hankel functions, and V is a tall Vandermonde matrix (2-D) or a spherical harmonic matrix (3-D). Matrix V possesses the property $V^H V = I$, provided that at least one loudspeaker is present for each mode to be controlled, i.e., $L \geq N_{2D}$ or $L \geq N_{3D}$.

The wave-domain adaptive filtering (WDAF) approach transforms the signals at the microphones and the loudspeaker signals into the wave domain through the transform \mathcal{T}_1 and \mathcal{T}_3 , then adaptively calculates the mode-domain compensation signals $\tilde{C}(w)$, and transforms the compensated loudspeaker signals back using the inverse transform \mathcal{T}_2 as depicted in Figure 3. If the compensation filter matrix $\tilde{C}(w)$ is forced to be diagonal, then



[FIG3] The listening room compensation using WDAF. The free-field transformed loudspeaker excitation signals \tilde{g} are used in a reverberant room with the filter matrix \tilde{C} to compensate for the ATFs in matrix H .



[FIG4] Crosstalk cancellation for delivering a time-domain signal s to the top microphone while canceling the signals at the remaining $Q - 1$ microphones.

each of its diagonal entries can be determined from decoupled adaptive filters. This would explicitly solve the problems of computational complexity that appeared in multipoint compensation techniques. While it is straightforward to choose \mathcal{T}_1 and \mathcal{T}_3 to do so, in reality \mathcal{T}_2 cannot always be chosen without a priori knowledge of the ATF matrix. However, [22] and [23] show that the system can be partially diagonalized by choosing $\mathcal{T}_1 = V^H$, $\mathcal{T}_2 = V$, and $\mathcal{T}_3 = U^H$.

SYSTEM IDENTIFICATION OF THE ATF MATRIX

The ATFs change in a room as people move about and as temperature changes. Since active room compensation in particular is sensitive to this phenomenon, it is better if the ATFs are determined adaptively. Similar to active listening room compensation, this task can be performed efficiently in the wave domain where transforms are used to part-diagonalize the reverberant-room ATF matrix [23].

The advantages of WDAF and the mode-domain approaches are that 1) sound pressure is controlled over the entire control region and not just at specific points and 2) they represent the problem with a reduced number $N_{2D} < M$ (or $N_{3D} < M$) of parameters, which reduces the complexity and reduces the correlation in the elements of the ATF matrix since the filters are part decoupled. This helps speed the convergence of adaptive filtering.

Since many more microphones and loudspeakers are required for a 3-D control zone, active room compensation is more practically deployed in 2-D scenarios. However, 2-D compensation cannot satisfactorily correct for roof and floor reflections, so sound absorbers must be employed to eliminate these effects.

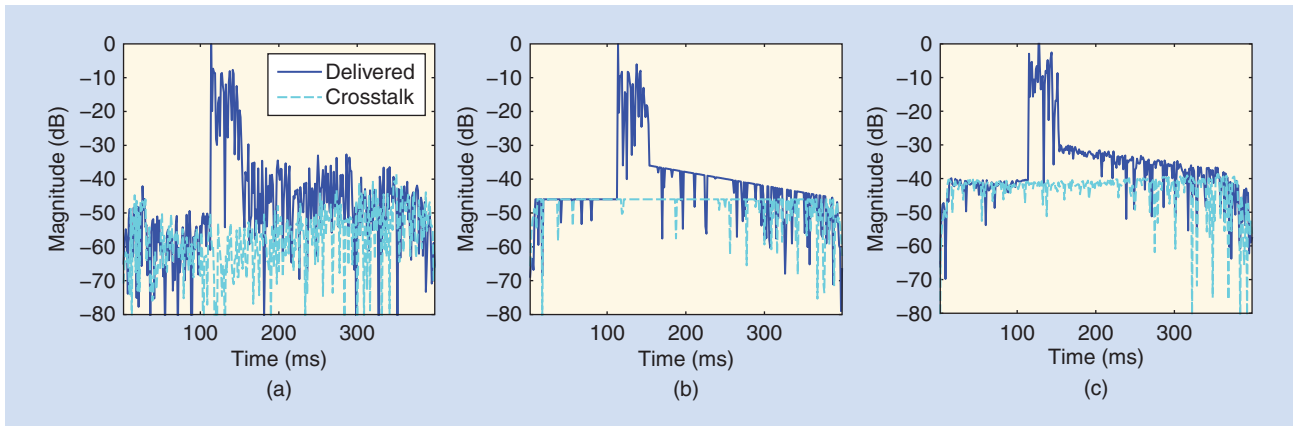
IMPULSE RESPONSE RESHAPING

Multiple listening zones may also be achieved by using crosstalk cancellation. Here, each of Q signal is delivered to a listening position while canceling the crosstalk paths to the remaining $Q - 1$ positions using L loudspeakers and, for monaural signals, $M = 1$ microphone in each zone. As shown in Figure 4, this problem is solved by implementing crosstalk-cancellation filters. The basic idea of the impulse response reshaping approach is that fully equalizing the delivered paths is unnecessary. It is more robust and efficient to reshape these impulse responses.

Using impulse response reshaping, the early reflections of the delivered paths are reinforced while late reverberation and crosstalk are minimized [25]. Here, by defining windows on these desired and undesired ATF components $w_q^{(d)}$ and $w_q^{(u)}$ respectively in each zone q , the ratio of undesired-to-desired components is minimized

$$\min_{\tilde{g}} \log \frac{\|W_u \tilde{r}\|_{p_u}}{\|W_d \tilde{r}\|_{p_d}}, \quad (12)$$

where \tilde{r} represents the global impulse response given a concatenated vector of crosstalk cancellation filters $\tilde{g} \triangleq [\tilde{g}_1^T, \dots, \tilde{g}_L^T]^T$ and a block-Toeplitz matrix \tilde{H} representing the room impulse responses, i.e., $\tilde{r} = \tilde{H}\tilde{g}$, $W_u \triangleq \text{Diag}(w_1^{(u)}, \dots, w_Q^{(u)})$, and $W_d \triangleq \text{Diag}(w_1^{(d)}, \dots, w_Q^{(d)})$. Different p_d - and p_u -norms may be chosen for the desired and undesired components, but it has been shown to be perceptually favorable to choose norms that approximate the



[FIG5] The shortening of impulse responses to 50 ms in a room of reverberation time 250 ms using (a) relaxed multichannel least squares, (b) the relaxed minimax approach in [24], and (c) the ratio optimization approach of [25].

infinite norm. Equation (12) can be solved analytically for the $p_u = p_d = 2$ case where it reduces to a generalized Rayleigh quotient. In general, (12) is solved using the steepest descent methods [25]. A relaxed multichannel approach using least squares [26] and minimax metrics [24] may include regularizations to reduce the array effort below that of the ratio-based approach in [25]. These approaches are compared in Figure 5 for simulation with $L = 3$ and $Q = 2$ in a room with a reverberation time of 250 ms using only 150 ms-long reshaping filters. The ratio-based approach shown is for $p_u = 10$, $p_d = 10$, and 1,000 steepest descent iterations.

Impulse response reshaping, in principle, can be applied to the PM- and modal-space approaches of creating personal sound zones. More robust and efficient filters can be obtained than equalization by canceling the undesirable late reverberation while leaving in some beneficial early reflections. Unfortunately this problem must be formulated in the time domain, which results in a computationally intractable massive multichannel problem. The future development of lower-complexity convex optimization algorithms may permit practical solutions to these large problems.

DIRECTIONAL SOURCES

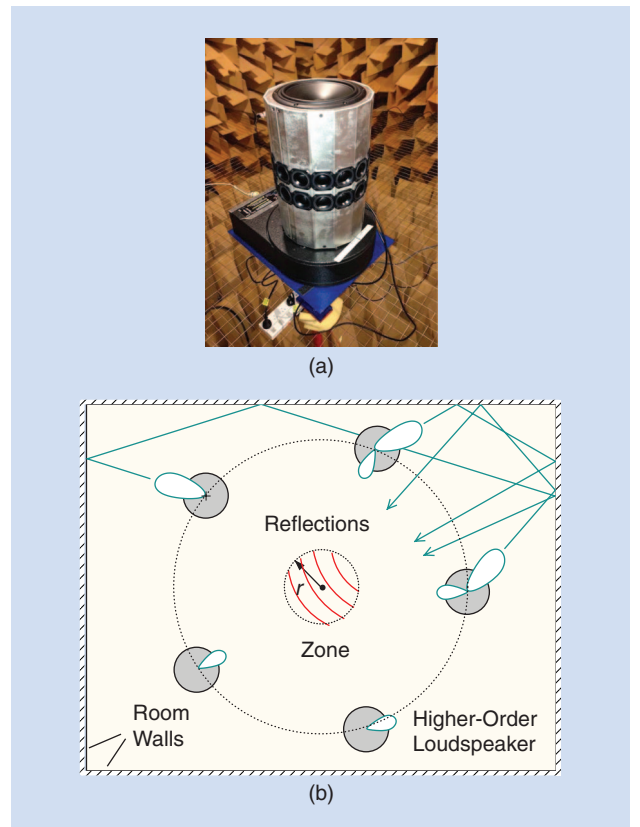
The use of directional sources can provide advantages over conventional loudspeakers, whose directivity is omnidirectional at low frequencies and is not typically controllable. Directional sources that provide multiple modes of sound radiation can be used with active compensation to produce sound arriving from angles where there are no sources by reflecting sound from room surfaces and can also cancel unwanted reverberation (Figure 6).

In a multilistener situation, a single directional loudspeaker can reduce unwanted radiation of sound to other listeners by maximizing the direct sound to the intended recipient relative to the reverberant field. A loudspeaker with directivity D and radiating acoustic power W in an ideal Sabinian space produces a direct sound intensity $I_{\text{dir}} = WD/(4\pi r^2)$ and a reverberant sound intensity of $I_{\text{rev}} = 4W/R'$, where $R' = S\epsilon/(1 - \epsilon)$ is the room constant, S the room surface area, and ϵ the mean absorption

coefficient of the room surfaces. The direct to reverberant intensity ratio is thus

$$DRR = \frac{DR'}{4\pi r^2}. \quad (13)$$

Increasing the directivity then allows the direct sound at the listener to be increased relative to the reverberant sound. Equivalently, the reverberant field is reduced by $1/DRR$.



[FIG6] A demonstration of the higher-order loudspeaker in (a) a cylindrical baffle and (b) the schematic plot of its behavior.

Standard loudspeakers typically have insufficient directivity to provide a significant enhancement of direct sound in a reverberant space. High directivity can be achieved using traditional array techniques such as delay and sum beamforming, but the array size must be large at low frequencies to achieve significant directivity. For practical use, superdirectional arrays are required, which achieve higher directivities than an array with uniform amplitude weightings [27]. Superdirectivity can be achieved using linear differential arrays, where the transducer weights have alternating signs, or by using circular and spherical arrays, where the weights are obtained from trigonometric or spherical harmonic functions, respectively. Such loudspeakers are termed *higher-order sources* (HOSs) and can produce multiple radiation patterns that are described by cylindrical or spherical harmonics.

Because superdirectional arrays are compact relative to their directivity, they may be built into a single unit, and we therefore assume here that a directional source is a single unit, typically of similar dimension to a standard loudspeaker. This section considers the design of directional loudspeakers and their application to maximum directivity and then focuses on the advantages of using arrays of sources, which allow greater control of sound fields over wide areas and are particular suitable for establishing personal sound zones.

SPHERICAL ARRAYS

The sound field produced by an arbitrary source of maximum radius a positioned at the origin and radiating a complex frequency $\exp(i\omega t)$ is represented in the wave domain as in (S2b) [17]

$$p(r, \theta, \phi, w) = \sum_{\nu=0}^N \sum_{\mu=-\nu}^{\nu} \alpha_{\nu}^{\mu}(w) h_{\nu}^{(2)}(kr) Y_{\nu}^{\mu}(\theta, \phi), \quad r \geq a, \quad (14)$$

where $h_{\nu}^{(2)}(kr)$ is the spherical Hankel function of the second kind, i.e., the radial function to represent the mode amplitude at r and $\alpha_{\nu}^{\mu}(w)$ are sound field coefficients. Similar to the dimensionality analysis in the wave domain, we will assume that the directivity of the source can be described by a maximum order N so that $\nu \in [0, N]$.

The most direct method for constructing a loudspeaker that can produce a controllable directivity is to mount a number of drivers in a spherical baffle of radius a [28]. The general behavior of such a source is most simply explained by deriving the sound field due to a sphere with arbitrary surface velocity

$$v(\theta_s, \phi_s, t, w) = e^{i\omega t} \sum_{\nu=0}^N \sum_{\mu=-\nu}^{\nu} \zeta_{\nu}^{\mu}(w) Y_{\nu}^{\mu}(\theta_s, \phi_s), \quad (15)$$

where (θ_s, ϕ_s) is the driver position on the sphere. The exterior field has the general form of (14). The expansion coefficients are found by calculating the radial velocity for the general case, and requiring that they equal (15), i.e.,

$$\alpha_{\nu}^{\mu}(w) = -i\rho c \frac{\zeta_{\nu}^{\mu}(w)}{h_{\nu}^{(2)}(ka)}$$

and the sound field, including the effect of mass-controlled drivers, is

$$p(r, \theta, \phi, t, w) = -\frac{i\rho c e^{i\omega t}}{k} \sum_{\nu=0}^N \sum_{\mu=-\nu}^{\nu} \zeta_{\nu}^{\mu}(w) \frac{h_{\nu}^{(2)}(kr)}{h_{\nu}^{(2)}(ka)} \times Y_{\nu}^{\mu}(\theta, \phi), \quad r \geq a.$$

Hence, each coefficient of the surface velocity produces a corresponding mode of radiation whose polar response is governed by a spherical harmonic.

The normalized magnitude of the mode responses for orders 0–5 are shown in Figure 7(a). For all modes greater than order $\nu = 0$, the response reduces at low frequencies. All modes of order ν become active at a frequency approximately given by $ka = \nu$ or

$$f = \frac{\nu c}{2\pi a}. \quad (16)$$

This means that it is not possible to create high-order directivities at low frequencies. The spherical loudspeaker is omnidirectional at low frequencies and can produce increasing directivities as more modes become active above frequencies given by (16).

In practice, the surface velocity in (15) must be approximated using a discrete array of L_0 drivers positioned on the sphere. Ideally the drivers are positioned so that they are spaced equally from each other which produces the most robust approximation to the integration over the sphere required to approximate each spherical harmonic. This is possible if the drivers are placed in the center of the faces of platonic solids, allowing up to 20 drivers (for the icosahedron). Higher numbers of drivers can be used using numerically optimized integration nodes for the sphere.

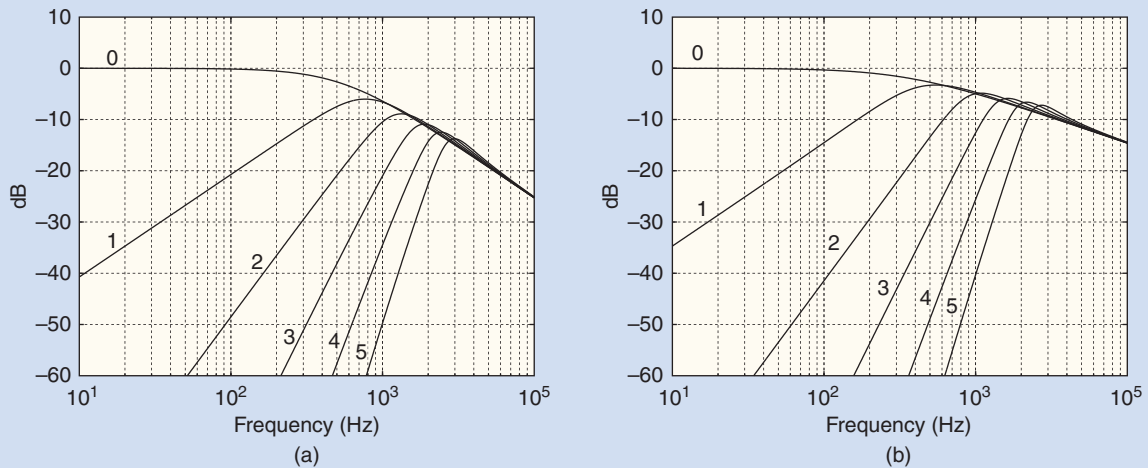
A simple way to model the discrete approximation is to assume each driver is a point source. The sound field due to a point source on a sphere then models a single driver, and the sound fields due to L_0 point sources allows the calculation of the total field. However, this approach ignores the directivity of each driver, which becomes significant at high frequencies. A more accurate model of the drivers that is mathematically tractable is to model each one as a spherical cap vibrating radially [28].

The sampling of the sphere means that the spherical loudspeaker is unable to generate spherical harmonic terms above the spatial Nyquist frequency of the array. This may be derived by noting that there are a total of $N = (N+1)^2$ spherical harmonics up to order N . Controlling this number of modes using L_0 loudspeakers is possible for $L_0 \geq N$. At a given frequency, the maximum-mode order that can be radiated is $N = ka$. Hence, the spatial Nyquist frequency is

$$f_{\text{Nyq},3D} = \frac{c(\sqrt{L_0} - 1)}{2\pi a}. \quad (17)$$

The number of drivers required for a sphere of radius a to produce N th-order directional responses up to a frequency f is given by

$$L_{3D} = \left(\frac{2\pi a f}{c} + 1 \right)^2.$$



[FIG7] The normalized magnitude of the mode responses of (a) a spherical source and (b) a cylindrical source for orders 0–5.

For example, a third-order speaker with radius $a = 0.1$ m and a Nyquist frequency of 4 kHz would require 70 drivers. This is a large number of drivers, and motivates the investigation of simpler approaches such as cylindrical and line arrays.

CYLINDRICAL ARRAYS

A simpler approach may be taken if the directional loudspeaker is only required to produce directivity in a 2-D plane. This is commonly the case for sound reproduction in the home, where stereo and 5.1-surround formats are ubiquitous. A circular array requires fewer drivers than a spherical array for the same spatial Nyquist frequency. To see this, consider a sphere where L_0 drivers are placed on the equator instead of equally spaced around the sphere. This arrangement allows for the generation of sectorial spherical harmonics, where $\nu = |\mu|$, which produce radiation with lobes only in the (x, y) plane. The driver spacing is now $2\pi a/L_0$ and the spatial Nyquist frequency is

$$f_{\text{Nyq},2D} = \frac{c(L_0 - 1)}{4\pi a}. \quad (18)$$

The number of drivers for a given 2-D spatial frequency is

$$L_{2D} = \frac{4\pi a f}{c} + 1.$$

Comparing (18) with (17), the 2-D Nyquist frequency can be much higher than the 3-D Nyquist frequency for the same number of drivers. The limitation of the circular array is that the transducer layout does not provide sufficient vertical directivity at high frequencies, and the source begins to produce unwanted radiation lobes in elevation. To reduce these lobes, the transducers must either have greater aperture in elevation or a line array must be used to control the vertical directivity. Since a line array is more effective when mounted on a cylinder than on a sphere, a practical alternative to the spherical array for the 2-D case is a cylindrical baffle in which multiple

circular arrays are mounted (Figure 6). Such a geometry can still use fewer transducers than the spherical case, for the same spatial Nyquist frequency.

The radiation of sound for the cylindrical case can be approximated by assuming that the cylinder is infinite and that each driver is represented as a surface velocity distribution in height z and azimuth angle ϕ [29]. Its produced mode responses are shown in Figure 7(b). The responses are similar to those for the spherical source, and the activation frequencies are the same. The limitation of this analysis is that, in practice, a truncated cylinder must be used leading to variations of the mode response magnitude around the infinite cylinder values due to diffraction from the ends of the cylinder.

LINE ARRAYS

The simplest array for providing high directivity is a line array, which produces an axisymmetric polar response. While this does not provide the full control of 3-D or 2-D radiation that the spherical and cylindrical arrays do, it is sufficient for maximizing the direct to reverberant ratio. It has the same limitation as the circular and spherical arrays, that is difficult to create high-order responses at low frequencies. However, the line array allows an order N response to be produced using $L_0 = N + 1$ transducers as opposed to $(N + 1)^2$ using a spherical array or $2N + 1$ for a circular array (assuming no vertical directivity control). The maximum directivity produced in 3-D is [30]

$$D = (N + 1)^2.$$

An order N loudspeaker with this directivity will produce the maximum direct to reverberant ratio for an on-axis listener. The simplest case, $N = 1$, results in a polar response $p(\theta) = 0.25 + 0.75 \cos(\theta)$, which has a directivity of four [7]. The first-order response can be implemented using $N = 2$ coupled or uncoupled drivers, or more simply, using a single driver and controlling the

radiation from the rear of the driver, although the directivity results can be less accurate with frequency [7].

ARRAYS OF DIRECTIONAL SOURCES

If multiple directional loudspeakers are available, then it becomes possible to create multiple zones of sound. Multizone reproduction requires a large number of monopole loudspeakers. The use of directional sources allows the production of multizone fields using significantly fewer loudspeaker units. In effect, a large number of drivers are grouped into a small number of physical devices to allow the creation of complex sound fields.

It has been shown that an array of L N th order sources operating in free-field conditions has a spatial Nyquist frequency of approximately $2N$ times that of the same geometry monopole array [31]. Results better than free-field can be achieved in a reverberant room by using the techniques discussed in [32]. In this case, the directional sources are able to exploit room reflections to provide directions of arrival other than those directly from the sources. The use of L HOSs, each of which can produce up to order N responses, can produce a similar accuracy of a reconstructed field to $L(2N+1)$ monopole loudspeakers in the 2-D case, and $L(N+1)^2$ loudspeakers in the 3-D case. For example, Figure 8 shows the sound field reproduction error achieved using a circular array of five higher-order loudspeakers in comparison with an array of 45 monopole sources. For a virtual source angle of 72° (the desired source position is equal to the first loudspeaker position), the error is similar to that produced by the monopole sources. At the angle of 36° (the desired source halfway between two loudspeakers), the error is about 10 dB higher than the monopole case but still reasonably accurate, particularly at low frequencies. Reproduction has been achieved over a 1-m diameter using only five loudspeaker units with room

dereverberation. The simulation is limited to 2-kHz bandwidth for computational complexity reasons. The worst-case reproduction error will be below -10 dB up to around 3 kHz. The bandwidth and reproduction radius of accurate reproduction can be extended by using more sources and higher orders, creating sufficient space for multiple listeners listening to independent sound fields.

The use of HOSs can be viewed as an optimization problem with a constraint on the total number of loudspeaker units in the room. The only way to improve reproduction in such a case is to add capability to the existing loudspeakers. HOSs offer a practical approach to providing the control of the high-spatial-dimension sound fields that are required for creating multiple personal sound zones. For example, the reproduction of sound in Q zones of radius r_0 , up to a spatial frequency k_{\max} , using L HOSs requires a maximum order per source of

$$N = \left\lceil \frac{Q(k_{\max}r_0 + 0.5)}{L} - 0.5 \right\rceil. \quad (19)$$

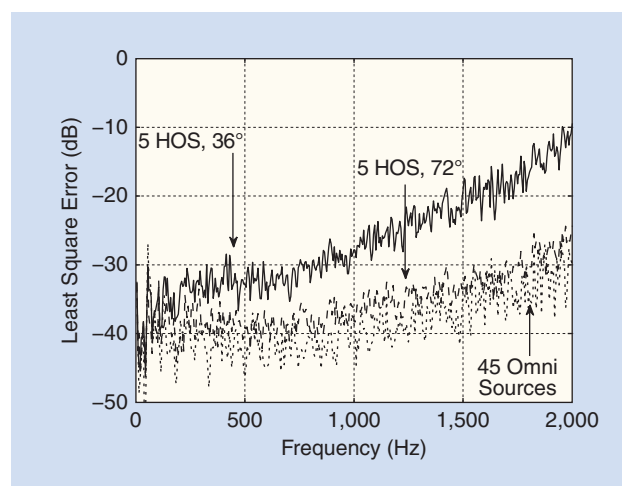
For 8 kHz reproduction over regions of radius 0.2 m, the order is $N = 10$ for $L = 10$ sources and $N = 6$ for $L = 15$ sources. Such numbers are achievable in moderate- to large-sized rooms.

SUMMARY AND FUTURE OPPORTUNITIES

In this article, we presented, according to our involvement and insights, the audio processing and loudspeaker design aspects that support the goal of establishing personal sound zones. The problems that have been explored include multizone sound control, wave-domain active room compensation, and directional loudspeaker design, which allow for sound control over spatial regions. A high-performance personal audio system would likely address many of these aspects in its design. In sound field control, interference mitigation and room compensation robust to changes and uncertainty in the acoustic environment remain as challenging problems. Yet future opportunities exist in 1) higher-order surround sound using an array of directional sources and wave-domain active room compensation to perform multizone sound control in reverberant enclosures and 2) personal audio devices using multiple sensors to establish personal sound zones by efficiently canceling crosstalk and using distributed beamforming.

AUTHORS

Terence Betlehem (Terence.Betlehem@callaghaninnovation.govt.nz) received the B.S., B.E., and Ph.D. degrees in telecommunications engineering from the Australian National University (ANU) in 1998, 2000, and 2005 respectively. From 2005 to 2006, he was a research fellow at the ANU Research School of Information Sciences and Engineering and a visiting researcher at National Information and Communication Technology Australia working in the areas of spatial signal processing and wireless channel modeling. Since 2007, he has worked at Callaghan Innovation (formerly Industrial Research Limited) in Lower Hutt, New Zealand, in the areas of spatial audio and wireless communications, where he is currently a senior research engineer. His research interests



[FIG8] The least squares error of reproduction as a function of frequency for an array of five fourth-order sources at 36° exactly between a pair of loudspeakers (dashed) and 72° coinciding with a loudspeaker (dashed), and a circular array of 45 omnidirectional line sources (unbroken) in a 2-D rectangular room of dimensions 6.4×5 m and with wall reflection coefficients of 0.7.

include active room compensation, microphone array processing, and room acoustic modeling.

Wen Zhang (wen.zhang@anu.edu.au) received the M.E. and Ph.D. degrees in electrical engineering from the Australian National University in 2005 and 2010, respectively. She worked as a Commonwealth Scientific and Industrial Research Organization (CSIRO) Office of the Chief Executive postdoctoral fellow at the CSIRO Process Science and Engineering Division in Sydney from 2010 to 2012. She is currently a research fellow at the College of Engineering and Computer Science at the Australian National University. Her research interests include spatial audio, source separation and localization, and active noise cancellation. She is currently an affiliate member of the Audio and Acoustics Signal Processing Technical Committee of the IEEE Signal Processing Society.

Mark A. Poletti (Mark.Poletti@callaghaninnovation.govt.nz) received an M.S. degree in physics at the University of Auckland in 1984, then worked at the Acoustics Research Centre at Auckland University for five years, where he was involved in acoustics testing and signal processing research. In 1989, he joined the Department of Scientific and Industrial Research communications group in Lower Hutt, New Zealand. This group became a part of Industrial Research Limited in 1992, which became Callaghan Innovation in 2013, where he is currently employed. In the 1990s, he developed the Variable Room Acoustic System (now called the Constellation System) for the electroacoustic enhancement of room acoustics. This work was the topic of his Ph.D. dissertation. His current research interests include electronic enhancement of room acoustics, holographic sound recording and reproduction systems using higher-order loudspeakers, and virtual acoustics systems.

Thushara D. Abhayapala (thushara.abhayapala@anu.edu.au) received the B.E. degree in interdisciplinary systems engineering and the Ph.D. degree in telecommunications engineering from the Australian National University (ANU), Canberra, in 1994 and 1999, respectively. Since December 1999, he has been a faculty member at ANU. His research interests are in the areas of spatial audio and acoustics signal processing and array signal processing. He is an associate editor of *IEEE/ACM Transactions in Audio, Speech, and Language Processing*. He is a member of the Audio and Acoustic Signal Processing Technical Committee (2011–2015) of the IEEE Signal Processing Society.

REFERENCES

- [1] M. Camras, "Approach to recreating a sound field," *J. Acoust. Soc. Amer.*, vol. 43, no. 6, pp. 1425–1431, 1967.
- [2] M. A. Gerzon, "Ambisonics in multichannel broadcasting and video," *J. Audio Eng. Soc.*, vol. 33, no. 11, pp. 859–871, 1985.
- [3] A. J. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis," *J. Acoust. Soc. Amer.*, vol. 93, no. 5, pp. 2764–2778, 1993.
- [4] D. B. Ward and T. D. Abhayapala, "Reproduction of a plane-wave sound field using an array of loudspeakers," *IEEE Trans. Speech Audio Processing*, vol. 9, no. 6, pp. 697–707, 2001.
- [5] W. F. Druyvesteyn and J. Garas, "Personal sound," *J. Audio Eng. Soc.*, vol. 45, no. 4, pp. 685–701, 1997.
- [6] 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. Amer.*, vol. 125, no. 4, pp. 2091–2097, 2009.
- [7] S. J. Elliott, J. Cheer, H. Murfet, and K. R. Holland, "Minimally radiating sources for personal audio," *J. Acoust. Soc. Amer.*, vol. 128, no. 4, pp. 1721–1728, 2010.
- [8] J. Cheer, S. J. Elliott, and M. F. S. Gálvez, "Design and implementation of a car cabin personal audio system," *J. Audio Eng. Soc.*, vol. 61, no. 6, pp. 414–424, 2013.
- [9] J.-W. Choi and Y.-H. Kim, "Generation of an acoustically bright zone with an illuminated region using multiple sources," *J. Acoust. Soc. Amer.*, vol. 111, no. 4, pp. 1695–1700, 2002.
- [10] S. J. Elliott, J. Cheer, J.-W. Choi, and Y.-H. Kim, "Robustness and regularization of personal audio systems," *IEEE Trans. Audio Speech Lang. Processing*, vol. 20, no. 7, pp. 2123–2133, 2012.
- [11] T. Betlehem and P. D. Teal, "A constrained optimization approach for multizone surround sound," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Prague, Czech Republic, May 2011, pp. 437–440.
- [12] M. A. Poletti, "An investigation of 2D multizone surround sound systems," in *Proc. 125th Audio Engineering Soc. Convention*, San Francisco, CA, Oct. 2008, pp. 1–9.
- [13] J.-H. Chang and F. Jacobsen, "Sound field control with a circular double-layer array of loudspeaker," *J. Acoust. Soc. Amer.*, vol. 131, no. 6, pp. 4518–4525, 2012.
- [14] M. Olik, J. Francombe, P. Coleman, P. J. Jackson, M. Olsen, M. Møller, R. Mason, and S. Bech, "A comparative performance study of sound zoning methods in a reflective environment," in *Proc. 52th AES Conf. Sound Field Control*, Guildford, U.K., Sept. 2013, pp. 1–10.
- [15] Y. Wu and T. D. Abhayapala, "Spatial multizone soundfield reproduction: Theory and design," *IEEE Trans. Audio Speech Lang. Processing*, vol. 19, no. 6, pp. 1711–1720, 2011.
- [16] N. Radmanesh and I. S. Burnett, "Generation of isolated wideband sound field using a combined two-stage Lasso-LS algorithm," *IEEE Trans. Audio Speech Lang. Processing*, vol. 21, no. 2, pp. 378–387, 2013.
- [17] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*. San Diego, CA: Academic, 1999.
- [18] P. Coleman, P. Jackson, M. Olik, and J. A. Pederson, "Optimizing the planarity of sound zones," in *Proc. 52nd Audio Engineering Society Int. Conf.*, Guildford, U.K., Sept. 2013, pp. 1–10.
- [19] T. Betlehem and T. D. Abhayapala, "Theory and design of sound field reproduction in reverberant rooms," *J. Acoust. Soc. Amer.*, vol. 117, no. 4, pp. 2100–2111, 2005.
- [20] D. S. Talagala, W. Zhang, and T. D. Abhayapala, "Multi-channel adaptive room equalization and echo suppression in sound field reproduction," *IEEE/ACM Trans. Audio Speech Lang. Processing*, vol. 22, no. 10, pp. 1522–1532, 2014.
- [21] L.-J. Brannmark, A. Bahne, and A. Ahlen, "Compensation of loudspeaker-room response in a robust MIMO control framework," *IEEE Trans. Audio Speech Lang. Processing*, vol. 21, no. 6, pp. 1201–1216, 2013.
- [22] S. Spors, H. Buchner, R. Rabenstein, and W. Herboldt, "Active listening room compensation for massive multichannel sound reproduction systems," *J. Acoust. Soc. Amer.*, vol. 122, no. 1, pp. 354–369, 2007.
- [23] M. Schneider and W. Kellermann, "Adaptive listening room equalization using a scalable filtering structure in the wave domain," in *IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Kyoto, Japan, May 2012, pp. 13–16.
- [24] T. Betlehem, P. Teal, and Y. Hioka, "Efficient crosstalk canceler design with impulse response shortening filters," in *IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Kyoto, Japan, Mar. 2012, pp. 393–396.
- [25] J. O. Jungmann, R. Mazur, M. Kallinger, T. Mei, and A. Mertins, "Combined acoustic MIMO channel crosstalk cancellation and room impulse response reshaping," *IEEE Trans. Audio Speech Lang. Processing*, vol. 20, no. 6, pp. 1829–1842, 2012.
- [26] W. Zhang, E. A. P. Habets, and P. A. Naylor, "On the use of channel shortening in multichannel acoustic system equalization," in *Proc. Int. Workshop on Acoustic Signal Enhancement*, Tel Aviv, Israel, Aug. 2010.
- [27] G. W. Elko, "Differential microphone arrays," in *Audio Signal Processing for Next-Generation Multimedia*. Norwell, MA: Kluwer, 2004, pp. 11–65.
- [28] B. Rafaely and D. Khaykin, "Optimal model-based beamforming and independent steering for spherical loudspeaker arrays," *IEEE Trans. Audio Speech Lang. Process.*, vol. 19, no. 7, pp. 2234–2238, 2011.
- [29] M. Poletti and T. Betlehem, "Design of a prototype variable directivity loudspeaker," in *Proc. 52th AES Conf. Sound Field Control*, Guildford, U.K., Sept. 2013, pp. 1–10.
- [30] A. T. Parsons, "Maximum directivity proof for three-dimensional arrays," *J. Acoust. Soc. Amer.*, vol. 82, no. 1, pp. 179–182, 1987.
- [31] M. A. Poletti and T. D. Abhayapala, "Spatial sound reproduction systems using higher order loudspeakers," in *IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Prague, Czech Republic, May 2011, pp. 57–60.
- [32] T. Betlehem and M. A. Poletti, "Two dimensional sound field reproduction using higher-order sources to exploit room reflections," *J. Acoust. Soc. Amer.*, vol. 135, no. 4, pp. 1820–1833, 2014.