FREQUENCY-BASED CUSTOMIZATION OF MULTIZONE SOUND SYSTEM DESIGN

Nasim Radmanesh, Bhaskar D. Rao

California Institute for Telecommunication and Information Technology, UC San Diego
La Jolla, CA 92093, USA, (e-mail:nradmanesh@ucsd.edu)

ABSTRACT

Methods for generating multizone soundfield have a wide range of applications. The desired dominant frequency bands to be generated and perceived, however, are not the same for different applications. Previous work proposed a low complexity algorithm for multizone wideband sound field generation using a frequency variable dictionary in a Lasso-LS optimization. This work demonstrates the flexibility of the novel algorithm in imposing variable number of active speakers to provide the desired reproduction accuracy at variable frequency bands. The deployment of this technique can lead to the customization of multizone sound system design based on both signal frequency content and listeners' perception.

***Index Terms*—** Frequency-based customization, frequency variable dictionary, Lasso-LS, multizone system, soundfield.

1. INTRODUCTION

There have been various studies of sound field generation over one area in space including wavefield synthesis [1], Ambisonics [2] and least squares (LS) approach [3], [4]. The idea of a personal sound space was first introduced in [5] and then further developed using acoustic contrast maximization (CM) [6], [7], cylindrical harmonic expansions [8], [9], beamforming technique [10] and LS pressure matching (PM) approach [11]. The multizone sound system has been employed for a number of applications such as a monitor display [12], a mobile device [13], an airplane seat head rests [14] and in an automobile cabin [15].
It has been discussed in [16] that for multizone wideband sound field generation, a large number of speakers is required. A Lasso-LS PM approach was developed in [17], [18] to minimize the number of speakers required for sound reproduction through the control of both speaker locations and weights. Effectiveness of such horizontal personal sound system was then investigated in [19] for listeners of variable heights. To reduce the complexity of speakers' subset selection in the Lasso stage, the implementation of a novel efficient harmonic nested (EHN) dictionary was then proposed in a Lasso-LS PM optimization [20]. The multizone sound reproduction techniques have been similarly employed for a range of applications, while the customization of design based on desired frequency content could lead to a more cost effective design.

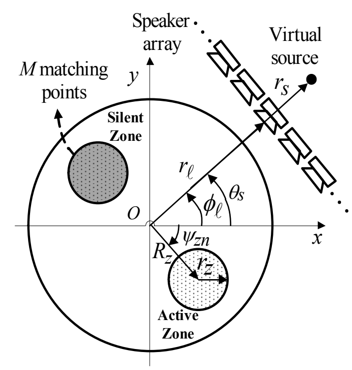


Fig. 1. Multizone sound field generation

This work investigates the capability of EHN Lasso-LS algorithm in controlling the multizone sound system performance across frequency. The flexibility of a sound reproduction algorithm to adjust the frequency contents reproduced within the active zone(s) and suppressed within the silent zone(s) makes the system design efficient for the desired application.
This paper is structured as follows: Section 2 defines the multizone problem and section 3 outlines the speaker weights estimation. Section 4 then presents EHN dictionary for Lasso subset selection and section 5 discusses a Lasso-LS algorithm using an EHN dictionary. Section 6 provides simulation results for the proposed technique, while section 7 concludes with a discussion of the results.

2. MULTIZONE SOUNDFIELD REPRODUCTION

Assuming free field conditions, a 2.5 dimensional multizone system is investigated while virtual sources and speakers are considered to be point sources, and that all zones, virtual sources and speakers are located in the same plane. In the following analysis the aim is to generate S isolated soundfields s=*1,...,S* for wideband sources (with constituent frequencies $f_q, q = 1, ..., Q$) in N zones. Fig. 1 illustrates the task scenario with the reproduction zones located at radius R_z from the origin and the nth zone's angle given by ψ_{z_n}.

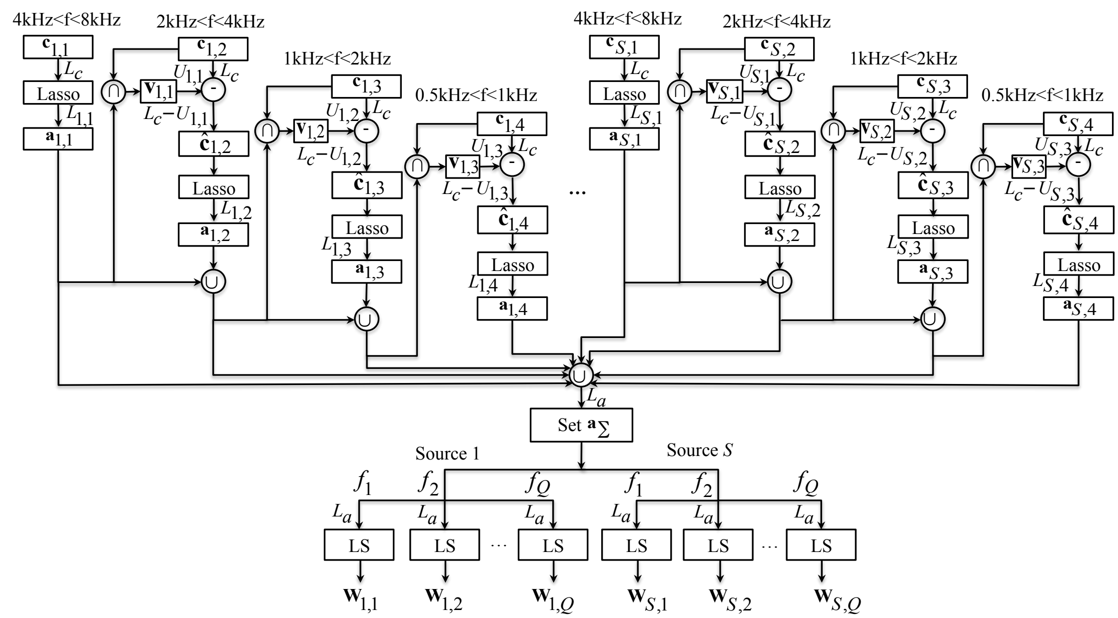


Fig. 2. A Lasso-LS Optimization algorithm using an efficient harmonic nested dictionary for subset selection

For every source angle θ_s, a linear array of L speakers is employed to generate multizone soundfield. The radius of the sth source is given by r_s and the ℓth speaker is located at radius r_ℓ and angle ϕ_ℓ. Each zone is of radius r_z with a covering of M matching points. The speaker weights $W_{s,q}(\ell)$ can be estimated using a PM approach given by [11]:

$$\mathbf{D}_{s,q} \approx \mathbf{H}_q \mathbf{W}_{s,q} \tag{1}$$

where $\mathbf{W}_{s,q}$ is the L by 1 vector of speaker weights $W_{s,q}(\ell)$, $\mathbf{D}_{s,q}$ is the MN by 1 vector of desired sound pressures at the matching points and $\mathbf{H}_q$ is the MN by L matrix of the 2-D Green's function.

3. SPEAKER WEIGHT ESTIMATION

3.1. LS Weight Estimation

The regularized LS approach provides a robust solution to (1). In this method, for generation of a frequency f_q of source s the speaker weights $W_{s,q}(\ell)$ are determined by minimizing the squared error between the desired and reproduced field with a power constraint:

$$\hat{\mathbf{W}}_{s,q} := \underset{\mathbf{W}_{s,q}}{\arg\min} \left[\|\mathbf{H}_q \mathbf{W}_{s,q} - \mathbf{D}_{s,q}\|_2^2 + \delta \|\mathbf{W}_{s,q}\|_2^2 \right] \tag{2}$$

where $\|.\|_2$ is the ℓ_2-norm, δ is the LS penalty parameter and $\|\mathbf{W}_{s,q}\|_2^2$ is the total speaker weight power.

3.2. Lasso Weight Estimation

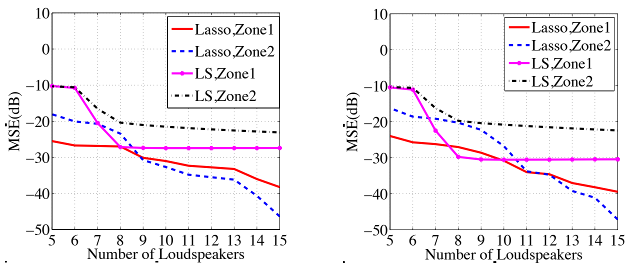
One approach to find the LS-optimal speaker locations is to derive a sparse solution to multizone PM problem. To reproduce the desired sound field for each frequency, f_q of source s, the $L_{s,q}$ speakers from L_c candidate speakers must be activated. The speaker weights are calculated from Lasso algorithm [21] as:

$$\check{\mathbf{W}}_{s,q} := \underset{\mathbf{W}_{s,q}}{\arg\min} \left[\frac{1}{2} \|\mathbf{H}_q \mathbf{W}_{s,q} - \mathbf{D}_{s,q}\|_2^2 + \lambda \|\mathbf{W}_{s,q}\|_1 \right] \tag{3}$$

where $\|.\|_1$ is the ℓ_1-norm and λ is the preselected Lasso penalty parameter. Larger values of λ produce fewer nonzero speaker weights and equation (3) can be solved using a coordinate descent method in the Frequency domain [22].

4. EHN DICTIONARY FOR LASSO SUBSET SELECTION

The major difficulty of Lasso subset selection is the computational complexity. The employment of a variable (time dependent) dictionary algorithms has been suggested in [23], [24] to reduce the computational complexity by applying the optimization only over previously unselected vectors. For wideband sound reproduction, the implementation of a novel frequency dependent dictionary termed as efficient harmonic nested (EHN) dictionary is proposed in [20] to reduce the complexity of the speaker location search. To form the EHN dictionary, a harmonic nested array is primarily employed to associate every frequency band to an optimal candidate set



(a) f=8kHz (b) f=4kHz

(c) f=2kHz (d) f=1kHz

Fig. 3. The MSE vs. the number of active speakers.

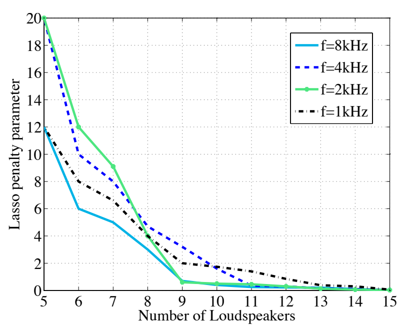


Fig. 4. Lasso Penalty Parameter vs. the number of speakers.

of speakers [25]. Removing the previously selected locations from the corresponding candidate subarray then further optimizes the dictionary.

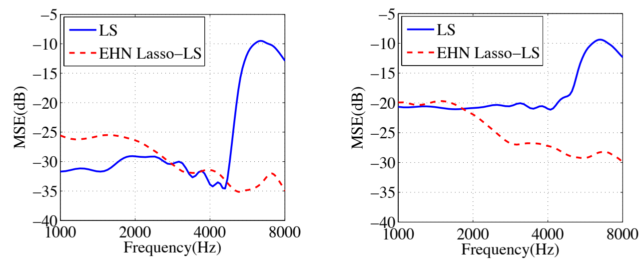
It is considered that the bth nested array candidate positions corresponding to source s are stored in set $\mathbf{c}_{s,b}$. The bth active subset, $\mathbf{a}_{s,b}$, comprises of active speakers selected from the bth modified location set, $\hat{\mathbf{c}}_{s,b}$, and its size is $|\mathbf{a}_{s,b}| = L_{s,b}$. The bth common subset, $\mathbf{v}_{s,b}$, is formed as

$$\mathbf{v}_{s,b} = (\bigcup_{i=1}^{b} a_{s,i}) \cap \mathbf{c}_{s,b+1} \tag{4}$$

The modified set of the *(b+1)*th candidate array, $\hat{\mathbf{c}}_{s,b+1}$, is then derived as:

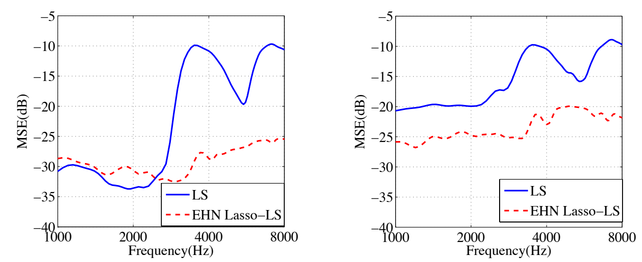
$$\hat{\mathbf{c}}_{s,b+1} = \mathbf{c}_{s,b+1} - \mathbf{v}_{s,b} \tag{5}$$

The number of candidate positions on subarray $\hat{\mathbf{c}}_{s,b+1}$ is $|\hat{\mathbf{c}}_{s,b+1}| = (L_c - U_{s,b})$ when $|\mathbf{v}_{s,b}| = U_{s,b}$. The union of



(a) Active Zone (b) Silent Zone

Fig. 5. The MSE vs. frequency. A larger number of speakers are selected for sound reproduction at frequencies over 3kHz.



(a) Active Zone (b) Silent Zone

Fig. 6. The MSE vs. frequency. A larger number of speakers selected for sound reproduction at frequencies under 3kHz.

the selected active sets across B frequency bands *b=1,B* and S sources is then stored in set $\mathbf{a}_{\Sigma}$.

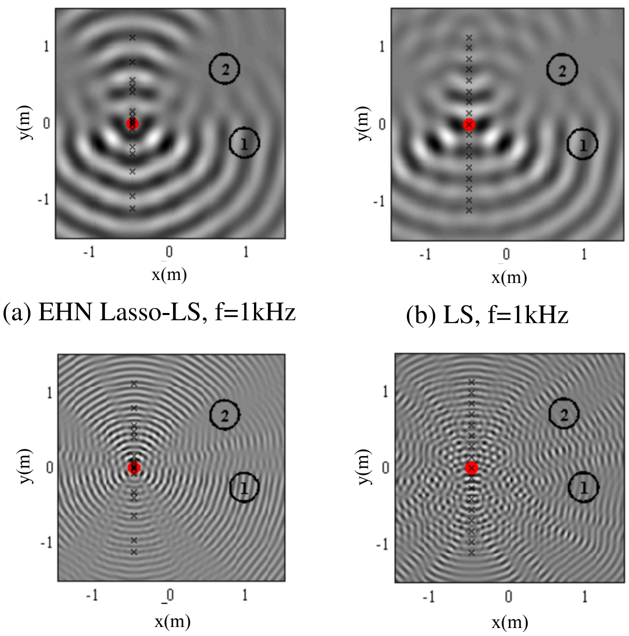
$$\mathbf{a}_{\Sigma} = \bigcup_{b,s} a_{s,b} \tag{6}$$

The number of all selected speakers is $L_a = |a_{\Sigma}|$.

5. A LASSO-LS OPTIMIZATION ALGORITHM WITH AN EHN DICTIONARY

A Lasso-LS algorithm using an EHN dictionary (Fig.2) is presented here for wideband sound field reproduction with an underlying assumption of fixed virtual sources. In the first stage, the LS-optimal speaker locations are selected from EHN dictionary across all frequency bands, b=1,..,B using Lasso algorithm. For the first nested array $(b = 1)$, Lasso minimizes the sum of the squared errors between the desired and reproduced pressures over the first harmonic nested candidate set, and penalizes it over the same set. From the second set forward $b \geq 2$, the bth set of speakers' weight corresponding to source s is calculated from the minimization of the sum of the squared errors between the desired and reproduced pressures over the bth harmonic nested candidate set, $\mathbf{c}_{s,b}$ and penalized over the bth modified candidate set, $\hat{\mathbf{c}}_{s,b}$ as:

$$\check{\mathbf{W}}_{\mathbf{c}_{s,b}} := \arg\min_{\mathbf{W}_{\mathbf{c}_{s,b}}} \left[\frac{1}{2} \|\mathbf{H}_{\mathbf{c}_{s,b}} \mathbf{W}_{\mathbf{c}_{s,b}} - \mathbf{D}_{s,b}\|_2^2 + \lambda_b \|\mathbf{W}_{\hat{\mathbf{c}}_{s,b}}\|_1 \right] \tag{7}$$



(a) EHN Lasso-LS, f=1kHz (b) LS, f=1kHz

(c) EHN Lasso-LS, f=4kHz (d) LS, f=4kHz

Fig. 7. Sound field visualization and speaker locations. Black crosses and a red circle mark respectively the speakers and the virtual source. $L = L_{a_2} = 17$

where $\mathbf{W}_{\mathbf{c}_{s,b}}$, $\mathbf{H}_{\mathbf{c}_{s,b}}$ and $\mathbf{D}_{s,b}$ are respectively the speakers' weight vector, the 2-D Green's function matrix and the desired sound pressures' vector corresponding to the *b*th harmonic set and source *s* and λ_b is the Lasso penalty parameter for the *b*th harmonic set. The Lasso penalty parameter λ_b determines the number of selected active speakers, $L_{s,b}$ for every frequency band *b*. To achieve desirable sound reproduction accuracy within different frequency bands, penalty parameters are tuned to select the required number of speakers.

A second stage regularized LS estimation is then employed for weight optimization as it is theoretically guaranteed to result in the lowest MSE for the selected set of speakers. In this stage, the non-uniformly spaced linear array of L_a active speakers is utilized for sound reproduction of all constituent frequencies, $f_q, q = 1, ..., Q$ of *S* wideband sources. The penalty parameter δ limits the power of the LS solution.

6. SIMULATION RESULTS

The simulations targeted generation of a wideband sound field within an area, zone 1 and suppressing it effectively in zone 2. The wideband source angle was $\theta_s = 180°$ with a source radii of $r_s = 0.5m$. The zones were located at $R_z = 1m$ from the origin and the number of matching points in each zone of radius $r_z = 0.2m$ was $M_z = 263$. A linear speaker array perpendicular to the *x* axis was employed with its center located at radius $r_\ell = 0.5m$ and angle $\varphi_\ell = 180°$. Zone 1 is located at $\psi_{z_1} = -15°$ and Zone 2 with a target sound attenuation of 60dB located at $\psi_{z_2} = 45°$. There are *B=4* harmonic nested subarrays of $L_c = 15$ positions at every subarray. In the Lasso stage, the centers of octave bands (from 1kHz to 8kHz) were used to select active speakers.

Fig. 3 demonstrates the performance of the described multizone system using *L=5-15* speakers. The EHN Lasso-LS lower error reduction rate with the number of speakers in comparison to the LS is due to the deployment of speakers with the major contribution in multizone sound reproduction. Fig. 4 illustrates the Lasso penalty parameter values used to impose sparsity versus the number of active speakers. Figs. 5 and 6, investigate two scenarios for EHN Lasso-LS and the LS performance assessment using different number of active speakers across optimized frequency range (1kHz-8kHz). In the first scenario (Fig. 5), the number of selected speakers in EHN Lasso-LS approach at frequencies 1kHz, 2kHz, 4kHz and 8kHz were respectively 3, 5, 10, 7 which resulted in up to 8dB lower error at frequencies over 3kHz. In the second scenario (Fig. 6) the number of selected speakers in EHN Lasso-LS approach at frequencies 1kHz, 2kHz, 4kHz and 8kHz were respectively 9, 8, 4, 3 leading to up to 5dB lower error at frequencies under 3kHz. The total number of speakers selected in both scenarios and the LS method were the same ($L = L_{a_1} = L_{a_2} = 17$). A fair comparison with similar array lengths and comparable total speakers' power in Figs. 5 and 6 illustrate up to 24dB and 22dB superior performance of EHN Lasso-LS over LS approach in scenarios 1 and 2 respectively. Fig. 7 illustrates the resulting soundfield of EHN Lasso-LS, and the LS approaches for the second scenario at frequencies $f = 1kHz$ and $4kHz$. The results demonstrated the capability of the proposed EHN Lasso-LS approach, in contrast to the LS, in adjusting the number of speakers to deliver a desired reproduction accuracy across frequency.

7. CONCLUSIONS

To provide an efficient sound system design of speakers' location and weight, frequency-based customization should be considered. An efficient harmonic nested Lasso-LS algorithm was employed in this paper for multizone wideband sound field generation. The ability of this approach in adjusting the performance of sound system across frequency was then investigated.

The results show that the EHN Lasso-LS technique enables multizone sound generation with up to 24 dB improvement in the MSE over LS optimization using e.g. 17 speakers. The flexibility of the EHN Lasso-LS approach, in contrast to the LS, in selecting the required number of speakers based on the desired reproduction accuracy at variable frequency bands is also demonstrated. Extensions of this approach to the generation of personal spaces based on applications and listeners' perception will be a topic of future work.

8. REFERENCES

[1] A J Berkhout, "A holographic approach to acoustic control," *Journal of the Audio Engineering Society*, vol. 36, no. 12, pp. 977–995, 1988.

[2] M A Gerzon, "Periphony: With-height sound reproduction," *Journal of the Audio Engineering Society*, vol. 21, no. 1, pp. 2–10, 1973.

[3] O Kirkeby and P A Nelson, "Reproduction of plane wave sound fields," *The Journal of the Acoustical Society of America*, vol. 94, pp. 2992, 1993.

[4] M Kolundzija, C Faller, and M Vetterli, "Reproducing sound fields using mimo acoustic channel inversion," *Journal of the Audio Engineering Society*, vol. 59, no. 10, pp. 721–734, 2011.

[5] W F Druyvesteyn and J Garas, "Personal sound," *Journal of the Audio Engineering Society*, vol. 45, no. 9, pp. 685–701, 1997.

[6] J-W Choi and Y-H Kim, "Generation of an acoustically bright zone with an illuminated region using multiple sources," *The Journal of the Acoustical Society of America*, vol. 111, pp. 1695, 2002.

[7] Stephen J Elliott, Jordan Cheer, Jung-Woo Choi, and Youngtae Kim, "Robustness and regularization of personal audio systems," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 20, no. 7, pp. 2123–2133, 2012.

[8] Y J Wu and T D Abhayapala, "Spatial multizone soundfield reproduction: Theory and design," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 19, no. 6, pp. 1711–1720, 2011.

[9] M A Poletti and F M Fazi, "An approach to generating two zones of silence with application to personal sound systems," *The Journal of the Acoustical Society of America*, vol. 137, no. 2, pp. 598–605, 2015.

[10] Marinus M Boone, Wan-Ho Cho, and Jeong-Guon Ih, "Design of a highly directional endfire loudspeaker array," *Journal of the Audio Engineering Society*, vol. 57, no. 5, pp. 309–325, 2009.

[11] M A Poletti, "An investigation of 2-d multizone surround sound systems," in *Audio Engineering Society Convention 125*. Audio Engineering Society, 2008.

[12] 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," *The Journal of the Acoustical Society of America*, vol. 125, pp. 2091, 2009.

[13] S J Elliott, H Murfet, and K R Holland, "Minimally radiating arrays for mobile devices," in *International Congress on Sound and Vibration 16*, 2009.

[14] S J Elliott and M Jones, "An active headrest for personal audio," *The Journal of the Acoustical Society of America*, vol. 119, pp. 2702, 2006.

[15] Jordan Cheer, *Active control of the acoustic environment in an automobile cabin*, Ph.D. thesis, University of Southampton, 2012.

[16] N Radmanesh and I S Burnett, "Reproduction of independent narrowband soundfields in a multizone surround system and its extension to speech signal sources," in *Acoustics, Speech and Signal Processing, 2011 IEEE International Conference on*. IEEE, 2011.

[17] N Radmanesh and I S Burnett, "Generation of isolated wideband sound fields using a combined two-stage lasso-ls algorithm," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 21, no. 2, 2013.

[18] N. Radmanesh and I. S. Burnett, "Wideband sound reproduction in a 2d multi-zone system using a combined two-stage lasso-ls algorithm," in *SAM Workshop, 2012 IEEE 7th*. IEEE, 2012, pp. 453–456.

[19] N Radmanesh and I S Burnett, "Effectiveness of horizontal personal sound systems for listeners of variable heights," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. IEEE, 2013, pp. 316–320.

[20] N Radmanesh, I S Burnett, and B D Rao, "A lasso-ls optimization with a frequency variable dictionary in a multizone sound system," *Audio, Speech, and Language Processing, IEEE Transactions on*, In press.

[21] R Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 58, no. 1, 1996.

[22] G N Lilis, D Angelosante, and G B Giannakis, "Sound field reproduction using the lasso," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 18, no. 8, pp. 1902–1912, 2010.

[23] B D Rao, "Signal processing with the sparseness constraint," in *IEEE International Conference on Acoustics Speech and Signal Processing*. IEEE, 1998.

[24] S F Cotter, B D Rao, K Kreutz-Delgado, and J Adler, "Forward sequential algorithms for best basis selection," *IEE Proceedings-Vision, Image and Signal Processing*, vol. 146, no. 5, pp. 235–244, 1999.

[25] N Radmanesh, *Multizone wideband sound field reproduction*, Ph.D. thesis, RMIT University, 2013.