SPARSE RECOVERY OF ACOUSTIC WAVES

Mohamed Mansour

Amazon Inc., USA

ABSTRACT

We present a general model for acoustic wave decomposition (AWD) on a rigid surface for a general microphone array configuration. The decomposition is modeled as a sparse recovery optimization problem that is independent of the shape of the rigid surface or the microphone array geometry. We describe an efficient algorithm for solving the optimization problem for broadband signals, and establish its effectiveness in practical systems.

Index Terms— Sparse Recovery, Elastic Nets, Acoustics Modeling, Plane Wave Decomposition.

1. INTRODUCTION

The mapping of microphone array observations to their constituent spatial components over the entire voiceband frequency is the main theme of microphone array processing. A canonical transform will ideally map the microphone array observations to components in the spatial domain in a way that resembles the Fourier transform mapping to frequency domain. Moreover, the canonical transform is required to be computationally tractable and invertible for sound-field synthesis applications. This canonical transform is difficult in practice as it requires a large number of microphones usually with a special geometry of the microphone array and the mounting surface [1, 2]. This work is focused on designing this invertible transform over voiceband frequency range for a microphone array with the most general configuration: arbitrary number of microphones, arbitrary microphone array geometry, and arbitrary form factor of the mounting surface.

A common approximation to the target canonical transform is through beamforming [3], where a set of beamformers are designed to cover the entire three-dimensional space. Though practically viable, this representation does not form a proper basis for the decomposition, hence, the transform is not invertible. Moreover, leakage from adjacent directions is unavoidable especially at low frequencies. The other common decomposition approach is to express the observed sound field in terms of a basis that constitutes a general solution to the acoustic wave equation [1]. For example, spherical and cylindrical harmonics were utilized to represent the sound field for spherical and cylindrical microphone arrays [4–6]. However, this representation is limited to a specific microphone array geometry and does not generalize to arbitrary microphone arrays, and it requires in general a large microphone array size for practical use cases. The plane-wave decomposition is another common approach for modeling the wave-field, and it has been used in seismic applications [7–10]. It is referred to as Radon transform [11], τ-p transform [12], or slant stacking [13]. It has also been investigated for microphone array processing for spherical and cylindrical microphone arrays [4], and for arbitrary shape microphone array [14]. In this context, spherical harmonics have been used as a proxy for plane-wave decomposition of the sound-field in microphone arrays. Another relevant body of work is in the context of sound-field reproduction, e.g., [15–17] where sparse plane-wave representation was used to approximate the sound field. For example, compressed sensing techniques were used with free-field plane waves in [17] to represent the sound field with a relatively large number of discrete microphones. The formulation is restricted to narrowband processing, where each frequency is solved independently.

In this work, a general formulation of the plane-wave decomposition, that does not require computing spherical harmonics, is developed. It extends sparse plane-wave representation to broadband signals (up to 8 kHz) and to microphones mounted on a rigid surface. The formulation works with arbitrary microphone array geometry and arbitrary form-factor of the mounting surface. It accommodates the scattering due to the device surface through the *acoustic dictionary*, which captures the total acoustic pressure on the surface due to free-field incident plane waves. With this formulation, the acoustic wave decomposition problem is reduced to a multidimensional elastic net optimization problem over the frequency range of interest; whose solution is sparse. For this optimization problem, we introduce a novel solver that leverages the physics of acoustic wave propagation to significantly reduce the complexity through a multi-stage procedure that prunes the spatial search space prior to running the solver on the entire frequency range of interest. The computational and memory complexity of the solver is relatively small and it is suited for embedded implementation. The effectiveness of the proposed procedure is established through evaluation of typical microphone arrays in smart speaker systems.

The following notation is used throughout the paper: bold small letters denote vectors, non-bold letters denote scalars, bold capital letter denote matrices. If Λ is a set, then $|\Lambda|$ denotes its cardinality.

2. FOUNDATIONS

2.1. Free Field Model

The propagation of acoustic waves in nature is governed by the acoustic wave equation, whose representation in the frequency domain (a.k.a. Helmholtz equation), in the absence of sound sources, is described as [2, 18]:

$$\nabla^2 p + k^2 p = 0 \tag{1}$$

where $p(\omega)$ is the acoustic pressure at frequency ω, and k is the wave number. Acoustic plane waves are eigenfunctions of the Helmholtz equation, and they constitute a powerful tool for analyzing the wave equation [19]. Further, an acoustic plane wave is a good approximation of the wave-field emanating from a far-field point source [1]. The acoustic pressure of a plane-wave with vector wave number $\mathbf{k}$ is defined at a point $\mathbf{r} = (x, y, z)$ in the three dimensional space as:

$$\psi(\mathbf{k}) \triangleq p_0 e^{-j\mathbf{k}^T\mathbf{r}} \tag{2}$$

where $\mathbf{k}$ is the three-dimensional wavenumber vector. For free-field propagation, $\mathbf{k}$ has the form

$$\mathbf{k}(\omega,\theta,\phi) = \frac{\omega}{c}\begin{pmatrix}\cos(\phi)\sin(\theta)\\ \sin(\phi)\sin(\theta)\\ \cos(\theta)\end{pmatrix} \tag{3}$$

where c is the speed of sound, θ and ϕ are respectively the elevation and azimuth of the vector normal to the plane wave propagation. Note that, k in (1) is $\|\mathbf{k}\|$. A local solution to the homogenous Helmholtz equation can be approximated by a linear superposition of plane waves [9, 19]:

$$p(\omega) = \sum_{l\in\Lambda} \alpha_l \, \psi\left(\mathbf{k}_l(\omega,\theta_l,\phi_l)\right) \tag{4}$$

where Λ is a set of indices that defines the directions of plane waves $\{\theta_l,\phi_l\}$, each $\psi(\mathbf{k})$ is a plane wave as in (2) with $\mathbf{k}$ is as in (3), and $\{\alpha_l\}$ are complex scaling factors that are computed to satisfy the boundary conditions [19]. Even though the expansion in (4) is derived using pure mathematical tools, it has an insightful physical interpretation, where the acoustic pressure at a point is represented as a superposition of acoustic pressures due to far-field point sources.

2.2. Rigid Surface Generalization

When an incident plane wave, $\psi(\mathbf{k})$, impinges on a rigid surface, scattering takes effect on the surface. The total acoustic pressure at a set of points on the surface, $\boldsymbol{\eta}(\mathbf{k})$, is the superposition of incident acoustic pressure (i.e., free-field plane wave) and scattered acoustic pressure [2]. The total acoustic pressure, $\boldsymbol{\eta}(\mathbf{k})$, can be either measured in anechoic room or simulated by numerically solving the Helmholtz equation with background acoustic plane wave $\psi(\mathbf{k})$ [20]. If two incident plane waves, $\psi(\mathbf{k}_1)$ and $\psi(\mathbf{k}_2)$, impinge on the surface, then from the linearity of the wave equation, the resulting total acoustic pressure is $\boldsymbol{\eta}(\mathbf{k}_1)+\boldsymbol{\eta}(\mathbf{k}_2)$. Hence, if a device with rigid surface is placed at a point whose free-field sound field is expressed as in (4), then the resulting acoustic pressure on the device surface becomes

$$\mathbf{p}(\omega) = \sum_{l\in\Lambda} \alpha_l \, \boldsymbol{\eta}\left(\mathbf{k}_l(\omega,\theta_l,\phi_l)\right) \tag{5}$$

where the free-field acoustic plane waves, $\{\psi(\mathbf{k}_l)\}$ in (4) are replaced by their *fingerprints* on the rigid surface $\{\boldsymbol{\eta}(\mathbf{k}_l)\}$; while preserving the angle directions $\{(\theta_l,\phi_l)\}$ and the corresponding weights $\{\alpha_l\}$. This preservation of incident directions on a rigid surface is a key idea that enables this work. It should be mentioned that in (5), secondary reflections, where scatterings from the surface hit other surrounding surfaces and come back to the surface, are ignored. This is a reasonable assumption when the insertion of the device does not significantly alter the sound-field in the room, as for example when the device dimensions are much smaller than the room dimensions. A similar formulation to (5) was described in [14] by using measured/simulated HRTF response, and it was used as an intermediate step for modeling through spherical harmonics. This is not needed for our formulation as the representation in (5) is utilized directly in our analysis.

Note that, the acoustic pressure $\mathbf{p}(\omega)$ in (5) could be represented by free-field plane waves, $\{\psi(\mathbf{k})\}$, where the scattered field is modeled by free-field plane waves (which are the eigenfunctions of the Helmholtz equation). However, that would distort the insightful representation in (5) as the constituent components will be merely a mathematical representation. Further, the entries of the acoustic dictionaries have both magnitude and phase information, which alleviates the spatial aliasing issues at high frequency.

2.3. Acoustic Dictionary

To enable the generalized representation in (5), the fingerprint, $\boldsymbol{\eta}(\mathbf{k})$, of each acoustic plane wave, $\psi(\mathbf{k})$, is computed at relevant points on the device surface, e.g., at the microphone array. We refer to the ensemble of all fingerprints of free-field plane waves as the *acoustic dictionary* of the device. Each entry of the device dictionary can be either measured in anechoic room with single-frequency far-field sources, or computed numerically by solving the Helmholtz equation on the device surface with background plane-wave using the device CAD model [20]. Both methods yield same result, but the numerical method has much lower cost and it is less error-prone because it does not require human labor. For the numerical method, each entry in the device dictionary is computed by solving the Helmholtz equation, using Finite Element Method (FEM) or Boundary Element Method (BEM) techniques, for the total field at the microphones with a given background plane wave, $\psi(\mathbf{k})$. The device CAD is used to specify the boundary in the simulation, and it is modeled as sound hard boundary. To have a true background plane-wave, the external boundary should be open and non-reflecting. In the simulation, the device is enclosed by a closed boundary, e.g., a cylinder or a spherical surface. To mimic open-ended boundary we use Perfectly Matched Layer (PML), which defines a special absorbing domain that eliminates reflection and refractions in the internal domain that encloses the device [21]. Standard packages for solving partial differential equations, e.g., [22] are used, and the simulation is rigorously validated with measured acoustic pressure on different form-factors. The acoustic dictionary has the form

$$\mathcal{D} \triangleq \{\boldsymbol{\eta}(\mathbf{k}_l,\omega) : \forall\, \omega, l\} \tag{6}$$

where each entry in the dictionary is a vector whose size equals the microphone array size, and each element in the vector is the total acoustic pressure at one microphone in the microphone array when a plane wave with $\mathbf{k}(\omega,\theta_l,\phi_l)$ hits the device. The dictionary also covers all frequencies of interest typically up to 8 kHz. The dictionary discretizes the azimuth and elevation angles in the three-dimensional space with angle resolution is typically less than 10°. Therefore, $|\mathcal{D}| \sim 800$ entries; which is the same order as the discretization in [17] for free-field plane waves.

2.4. Model Validation

The acoustic dictionary model was rigorously validated in [20], where the matching between simulated and analytical acoustic pressure for different form factors were presented. In the following experiment, the plane wave decomposition model with rigid surface in (5) is validated using a spherical microphone array with 32 microphones [23]. In Fig. 1, we show the reconstruction error of the acoustic pressure at the microphone array with two different source signals, white noise and multiple harmonics. The expansion is computed using the orthogonal matching pursuit method [24], and the reconstruction error (a.k.a, Goodness of Approximation, GoA) is defined as:

$$GoA \triangleq \frac{\int_\omega \|\mathbf{y}(\omega) - \sum_{l\in\Lambda} \alpha_l(\omega)\, \boldsymbol{\eta}(\mathbf{k}_l,\omega)\|^2}{\int_\omega \|\mathbf{y}(\omega)\|^2} \tag{7}$$

where $\mathbf{y}(\omega)$ is the observed sound-field at the microphone array. As noted from the figure, a small number of plane waves (~ 20) is sufficient for sound field approximation with reconstruction error less than -20 dB; which is sufficient for many practical applications. Note that, there is no overfitting in this case because the number of observations at each frequency is 32 which is larger than the number of nonzero elements in the output vector.

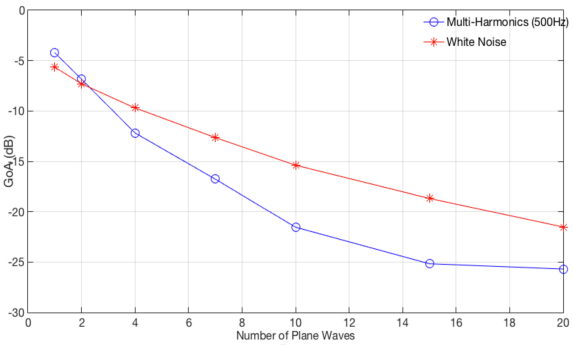

Fig. 1: Reconstruction error of the acoustic pressure at a spherical microphone array versus the number of acoustic plane waves in generalized acoustic wave decomposition in (5)

3. OPTIMIZATION PROBLEM

3.1. Optimization Model

The objective of the decomposition algorithm is to find the best representation of the observed sound field at the microphone array, $\{\mathbf{y}(\omega)\}$, using the device dictionary $\mathcal{D}$ in (6). A least-square formulation is the intuitive choice, where the objective is to minimize:

$$J(\boldsymbol{\alpha}) = \int_{\omega} \rho(\omega) \, \|\mathbf{y}(\omega) - \sum_{l \in \Lambda} \alpha_l(\omega) \boldsymbol{\eta}_l(\omega)\|_2^2 + g(\omega, \boldsymbol{\alpha}) \qquad (8)$$

where $g(.)$ is a regularization function, and $\rho(.)$ is a weighting function. An equivalent matrix form is:

$$J(\boldsymbol{\alpha}) = \int_{\omega} \rho(\omega) \, \|\mathbf{y}(\omega) - \mathbf{A}(\omega) \, . \, \boldsymbol{\alpha}(\omega)\|_2^2 + g(\omega, \boldsymbol{\alpha}) \qquad (9)$$

where the columns of $\mathbf{A}(\omega)$ are the individual entries of the acoustic dictionary at frequency ω, i.e., $\{\boldsymbol{\eta}_l(\omega)\}$. Note that, Λ in (8) refers to the nonzero indices of the dictionary entries; which represent directions in the three-dimensional space, and it is *independent* of ω. This independence stems from the fact that when a sound source emits a broadband frequency content, it is reflected by the same boundaries in its propagation path to the receiver . Therefore, all frequencies have components from the same directions but with different strengths (due to the variability of reflection index with frequency), which is manifested in the components $\{\alpha_l(\omega)\}$ [25]. Each component is a function of the source signal, the overall length of the acoustic path of its direction, and the reflectivity of the surfaces across its path. This independence between Λ and ω is a key property in characterizing the optimization problem in (9).

The typical size of an acoustic dictionary is $\sim 10^3$ entries, which corresponds to an azimuth resolution of 5°, and an elevation resolution of 10°. In a typical indoor environment, ~ 20 acoustic plane waves are sufficient for a good approximation of (5) as shown in Fig. 1. Moreover, the variability in the acoustic path of the different acoustic waves at each frequency further reduces the effective number of acoustic waves at individual frequencies. Hence, the optimization problem in (9) is a sparse recovery problem and proper regularization is needed to stimulate a sparse $\boldsymbol{\alpha}$. This requires L1-regularization as in standard LASSO optimization [26] that is encountered in numerous sparse recovery problems in statistics and signal processing. To improve the perceptual quality of the reconstructed audio, L2-regularization is added, and the regularization function $g(\boldsymbol{\alpha})$ has the general form of elastic net regularization [27]:

$$g(\omega, \boldsymbol{\alpha}) = \lambda_1(\omega) \sum_l |\alpha_l(\omega)| + \lambda_2(\omega) \sum_l \|\alpha_l(\omega)\|^2 \qquad (10)$$

3.2. Solution Strategies

The strategy for solving the elastic net optimization problem in (9) depends on the size of the microphone array. If the microphone array size is big, e.g., ≥ 20, then the observation vector is bigger than the typical number of nonzero components in $\boldsymbol{\alpha}$, and the problem is relatively simple with many efficient solutions in the literature under various conditions [28, 29]. For example, the orthogonal matching pursuit algorithm [24] with only L2-regularization was used to recover Λ and $\boldsymbol{\alpha}$ in Fig. 1 at individual frequencies. Other existing algorithms to the sparse recovery problem [28] can be used with similar performance.

The problem becomes much harder when working with a microphone array in existing consumer electronics products; which typically has less than 10 microphones [30]. In this case, the optimization problem at each ω becomes an underdetermined least-square problem because the number of observations is less than the expected number of nonzero elements in the output. For this problem, elastic net regularization in the form of (10) is necessary. Moreover, the invariance of directions (i.e., indices of nonzero elements, Λ) with frequency could be exploited to reduce the search space for a more a tractable solution, which is computed in two steps. The first step computes a pruned set of indices, Λ, that contains the nonzero coefficients at all frequencies. This effectively reduces the problem size from $|\mathcal{D}|$ to $|\Lambda|$, which is a reduction of about two orders of magnitude. The pruned set Λ is computed by a two-dimensional matched filter followed by a small scale LASSO optimization. The search procedure to compute Λ is as follows:

1. For each angle (θ_l, ϕ_l) in the device dictionary, compute

$$\Gamma(\theta_l, \phi_l) = \int_{\omega} \sigma(\omega) \, \|\langle \, \mathbf{y}(\omega), \boldsymbol{\eta}(\mathbf{k}(\omega, \theta_l, \phi_l)) \, \rangle\|^2 \qquad (11)$$

   The weighting $\sigma(\omega)$ is a function of the SNR of the corresponding time-frequency cell. This metric is computed only when the target signal is present.

2. Identify local maxima of $\Gamma(\theta_l, \phi_l)$, and discard the ones in the neighborhood of stronger maxima (e.g., within 10°). This pruning is needed to improve the numerical stability of the optimization problem.

3. Find a superset $\bar{\Lambda}$ with the indices of the strongest surviving local maxima of $\Gamma(\theta_l, \phi_l)$, with $|\bar{\Lambda}| \geq |\Lambda|$.

4. (optional) To further refine $\bar{\Lambda}$, run LASSO optimization with coordinate-descent solver [29], but with entries limited to $\bar{\Lambda}$ and choose the indices of the highest energy components in the output solution as Λ.

This search procedure runs only on a subset of high energy frequency components rather than the whole spectrum, and it does not need to run at each time frame. The LASSO optimization in the last step yields a much higher accuracy at a small complexity cost because a small number of iterations is sufficient to converge to Λ.

The second step in the solution procedure solves the elastic net optimization problem in (9) with the pruned set Λ to compute $\{\alpha_l(\omega)\}_{l \in \Lambda}$ for all ω. In our implementation, the coordinate-descent procedure [29, 31] is utilized, as it provides significant speedup as compared to its gradient-descent counterpart. Moreover, it has guaranteed convergence for the given elastic net problem [29]. The

regularization parameters, λ_1 and λ_2 , vary with frequency because dictionary vectors are more correlated at low frequencies.

4. EVALUATION

In the following experiments, an acoustic dictionary of a circular microphone array, with 7 microphones mounted atop a cylindrical rigid surface, is used. The dictionary is computed using acoustic simulation as described in section 2.3. It has an azimuth resolution of 5°, and elevation resolution of 20°.

The first experiment evaluates the accuracy of computing Λ with the pruning algorithm outlined in section 3.2. Monte Carlo simulations with speech stimuli are used with 20 random directions to evaluate the accuracy of the proposed multistage solver algorithm, as compared to full solver with standard coordinate-descent algorithm. The results are shown in Fig. 2, where the lowest point in the multistage solver performance corresponds to the performance without the LASSO optimization (i.e., using only the two-dimensional matched filter). Note that, there is a steep improvement after introducing the LASSO optimization at the cost of a small extra complexity. Nevertheless, as noted from the figure the complexity of the multistage solver is about one order of magnitude less than the full coordinate-descent solver running on the whole dictionary.

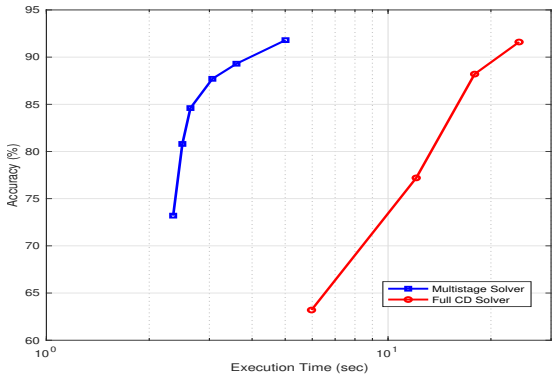

Fig. 2: Accuracy of computing the pruned set Λ with Multistage Solver and full Coordinate-Descent (CD) Solver

The second experiment evaluates signal-to-distortion ratio (SDR) [32] of the decomposition algorithm. In each trial, a source speech is drawn randomly from LibriSpeech dataset [33], and it undergoes reverberation (using a set of measured room impulse responses in different room environments) and interference from coherent noise (i.e., directional source) or incoherent (i.e., sensor) noise. In Fig. 3, the AWD output is compared with beamformer decomposition with minimum variance distortionless response (MVDR) beamformer [3] (which is optimized for coherent noise). The AWD provides a relatively uniform improvement with the two noise types. It provides a significant improvement of $\sim$ 9 dB over beamformer decomposition for the incoherent noise case, while slightly worse for the coherent case.

5. CONCLUSION

The problem of mapping microphone array observations to their constituent spatial components is at the intersection of acoustic modeling and microphone array processing. There is an existing gap in integrating acoustic modeling and microphone array processing in

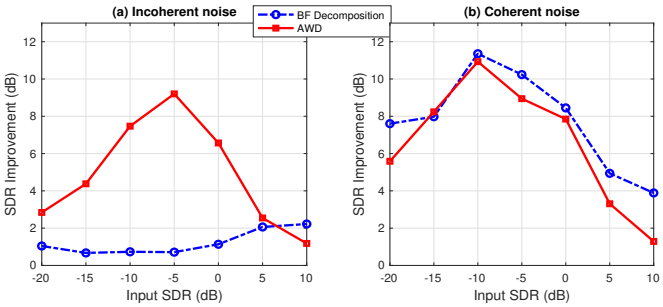

Fig. 3: Average SDR Improvement for AWD and beamformer (BF) decomposition with coherent and incoherent noise

practical real-time systems. This gap stems primarily from the imbalance between theoretical modeling that emphasizes model accuracy; and practical processing that emphasizes systems constraints. A rich relevant literature in acoustic modeling is frequently deemed inappropriate for practical systems because of the underlying system constraints. These constraints include, for example, the number of microphones, and the geometry of the microphone array and the mounting device. Moreover, computational, memory, and latency constraints in practical systems frequently becomes prohibitive for a modeling methodology.

This work provides a balanced solution that addresses the aforementioned gap for the mapping problem. The two key contributions of this work are the general formulation of the acoustic wave decomposition problem and the efficient multistage solver of the resulting elastic net optimization problem for broadband signals. Both contributions are necessary to enable real-time mapping of microphone array observations to their constituent spatial components on existing consumer electronics devices. Although the focus is on acoustic plane waves, the device dictionary might be extended to spherical acoustic waves to model near-field sources for increased efficiency. In addition to performance advantage as outlined in section 4, the proposed framework has the following merits:

1. It is an invertible map and it accommodates the scattering of the mounting surface. Therefore, the resolved directions correspond to true directions of the incident acoustic waves.

2. It accommodates any geometry of the microphone array and mounting surface, and any microphone array size. All these variations are modeled through the acoustic dictionary, which is computed using acoustic simulation of the device CAD model.

3. The proposed multistage solver is scalable with graceful degradation in the performance versus complexity. It can be customized to existing resources in the embedded system. The degrees of freedom include the angle resolution of the acoustic dictionary, and the number of iterations in the solver.

4. The magnitude information in the acoustic dictionary alleviates spatial aliasing to enable high frequency support.

This work establishes a technical foundation to enable many applications for microphone array processing; which were not discussed in this work due to space limitation, and this is a subject of future work.

6. REFERENCES

[1] H. Teutsch, *Modal array signal processing: principles and applications of acoustic wavefield decomposition*, vol. 348, Springer, 2007.

[2] E. Williams, *Fourier acoustics: sound radiation and nearfield acoustical holography*, Academic press, 1999.

[3] Michael Brandstein and Darren Ward, *Microphone arrays: signal processing techniques and applications*, Springer Science & Business Media, 2013.

[4] D. Zotkin, R. Duraiswami, and N. Gumerov, "Plane-wave decomposition of acoustical scenes via spherical and cylindrical microphone arrays," *IEEE transactions on audio, speech, and language processing*, vol. 18, no. 1, pp. 2–16, 2009.

[5] J. Meyer and G. Elko, "A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield," in *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*. IEEE, 2002, vol. 2, pp. II–1781.

[6] B. Rafaely, "Analysis and design of spherical microphone arrays," *IEEE Transactions on speech and audio processing*, vol. 13, no. 1, pp. 135–143, 2004.

[7] S. Treitel, P. Gutowski, and D. Wagner, "Plane-wave decomposition of seismograms," *Geophysics*, vol. 47, no. 10, pp. 1375–1401, 1982.

[8] O. Yilmaz and M. Taner, "Discrete plane-wave decomposition by least-mean-square-error method," *Geophysics*, vol. 59, no. 6, pp. 973–982, 1994.

[9] A. Moiola, R. Hiptmair, and I. Perugia, "Plane wave approximation of homogeneous helmholtz solutions," *Zeitschrift für angewandte Mathematik und Physik*, vol. 62, no. 5, pp. 809–837, 2011.

[10] E. Perrey-Debain, "Plane wave decomposition in the unit disc: Convergence estimates and computational aspects," *Journal of Computational and Applied Mathematics*, vol. 193, no. 1, pp. 140–156, 2006.

[11] G. Beylkin, "Discrete radon transform," *IEEE transactions on acoustics, speech, and signal processing*, vol. 35, no. 2, pp. 162–172, 1987.

[12] B. Zhou and S. Greenhalgh, "Linear and parabolic τ-p transforms revisited," *Geophysics*, vol. 59, no. 7, pp. 1133–1149, 1994.

[13] H. Brysk and D. McCowan, "A slant-stack procedure for point-source data," *Geophysics*, vol. 51, no. 7, pp. 1370–1386, 1986.

[14] D. Zotkin, N. Gumerov, and R. Duraiswami, "Incident field recovery for an arbitrary-shaped scatterer," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 451–455.

[15] W. Jin and W B. Kleijn, "Theory and design of multizone soundfield reproduction using sparse methods," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 12, pp. 2343–2355, 2015.

[16] S. Koyama, N. Murata, and H. Saruwatari, "Sparse sound field decomposition for super-resolution in recording and reproduction," *The Journal of the Acoustical Society of America*, vol. 143, no. 6, pp. 3780–3795, 2018.

[17] S. A Verburg and E. Fernandez-Grande, "Reconstruction of the sound field in a room using compressive sensing," *The Journal of the Acoustical Society of America*, vol. 143, no. 6, pp. 3770–3779, 2018.

[18] L. Kinsler, A. Frey, A. Coppens, and J. Sanders, *Fundamentals of Acoustics*, Wiley, third edition, 1982.

[19] R. Haberman, *Elementary Applied Partial Differential Equations*, vol. 987, Prentice Hall Englewood Cliffs, NJ, 1983.

[20] A. Chhetri, M. Mansour, W. Kim, and G. Pan, "On acoustic modeling for broadband beamforming," in *27th European Signal Processing Conference (EUSIPCO)*, 2019, pp. 1–5.

[21] J. Berenger, "A perfectly matched layer for the absorption of electromagnetic waves," *Journal of computational physics*, vol. 114, no. 2, pp. 185–200, 1994.

[22] COMSOL Multiphysics, "Acoustic module–user guide," 2017.

[23] MH Acoustics, "Em32 eigenmike microphone array release notes (v17. 0)," 2013.

[24] Y. Pati, R. Rezalifar, and P. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on*. IEEE, 1993, pp. 40–44.

[25] H. Kuttruff, *Room acoustics*, CRC Press, fourth edition, 2000.

[26] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society*, vol. 58, no. 1, pp. 267–288, 1996.

[27] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the royal statistical society*, vol. 67, no. 2, pp. 301–320, 2005.

[28] J. Tropp and S. Wright, "Computational methods for sparse solution of linear inverse problems," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 948–958, 2010.

[29] T. Hastie, R. Tibshirani, and M. Wainwright, *Statistical learning with sparsity: the lasso and generalizations*, Chapman and Hall/CRC, 2019.

[30] V. Kepuska and G. Bohouta, "Next-generation of virtual personal assistants (microsoft cortana, apple siri, amazon alexa and google home)," in *IEEE 8th annual computing and communication workshop and conference (CCWC)*. IEEE, 2018, pp. 99–103.

[31] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," *Journal of statistical software*, vol. 33, no. 1, pp. 1 – 22, 2010.

[32] E. Vincent, R. Gribonval, and C. Févotte, "Performance measurement in blind audio source separation," *IEEE transactions on audio, speech, and language processing*, vol. 14, no. 4, pp. 1462–1469, 2006.

[33] V. Panayotov, G. Chen, D. Povey, and S. Khudanpur, "Librispeech: an ASR corpus based on public domain audio books," in *2015 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, 2015, pp. 5206–5210.