A Multipath Sparse Beamforming Method

Afsaneh Asaei*†, Bhiksha Raj‡, Volkan Cevher†, and Hervé Bourlard*†
*Idiap Research Institute, Martigny; †École Polytechnique Fédérale de Lausanne, Switzerland
‡Machine Learning for Signal Processing group, Carnegie Mellon University, USA
Emails: afsaneh.asaei@idiap.ch, bhiksha.raj@cs.cmu.edu, volkan.cevher@epfl.ch, ,herve.bourlard@idiap.ch

***Abstract*—A novel formulation of beamforming is proposed for acquisition of the signals in reverberant acoustic clutter of interferences and noise. We derive the beamforming methods which incorporate the sparsity structure pertained to the acoustic source distribution and multipath propagation model. The quantitative assessments demonstrate that sparse beamforming enables effective beampattern steering from far fewer samples than the conventional beamformers. In addition, linear constraint on the desired channel rather than the desired direction improves the signal estimation performance in reverberant enclosures.**

I. Introduction

The signals captured by a microphone array can be represented as $\mathcal{X} = AS + N$, where S is the source signal and A is the room response from the source location to each of the microphones in the array, and N is noise. Alternately, if the room is represented as a grid, it can be represented as $\mathcal{X} = \Phi BS + N$, where ϕ is a matrix whose columns represent room response from each location in the grid to the array and B is an indicator vector identifying the locations of sources in the room. The conventional MVDR beamformer generates the output as $Y = W^H \mathcal{X}$ where the unknown filter weights $W \in \mathbb{R}^{M\times 1}$ are optimized in order to minimize the overall noise and interference power while guaranteeing the signal coming from the desired direction is received distortion-less. Alternatively, the MMSE estimator adapts the linear filtering for the reference signal and the weights are optimized to yield the minimum average power in signal recovery error [1].

II. Multipath Sparse Beamforming

The conventional beamforming requires information about the direction of the desired source to estimate the weights. In reverberant enclosure, this framework can be generalized by providing information about the desired channel denoted by ϕ_s. Hence, the multipath MVDR (m-mvdr) beamformer is obtained by

$$W^H_{\text{m-mvdr}} = \underset{W^H}{\arg\min} \left\{ W^H R_{\mathcal{X}} W \quad \text{s.t.} \quad W^H \phi_s = 1 \right\}, \tag{1}$$

The weights are optimized such that acquisition of the signal with respect to the desired channel is distortion-less. The solution using a Lagrange multiplier is expressed as

$$W^H_{\text{m-mvdr}} = \frac{\phi_s^H R_{\mathcal{X}}^{-1}}{\phi_s^H R_{\mathcal{X}}^{-1} \phi_s} \tag{2}$$

Given the forward model of multipath projections, we exploit the sparsity structure underlying the distribution of sources across the acoustic scene. The room response matrix includes reflection effects in the reverberant scenario. We can write $W = \Phi B$ and estimation of W^H amounts to estimating the sparse vector B. Hence, the optimization stated in (1) is regularized by the ℓ_1 norm of B to induce sparsity structure through

$$\hat{B} = \underset{B}{\arg\min}\{\|B^H \Phi^H \mathcal{X}\|_2 + \lambda_1 \|B\|_1 \quad \text{s.t.} \quad B^H \Phi^H \phi_s = 1\} \tag{3}$$

In a similar manner, we derive the formulation of MMSE estimator by the following optimization

$$\hat{B} = \underset{B}{\arg\min}\ E\{|B^H \Phi^H \mathcal{X} - S|^2\} \tag{4}$$

This framework requires acoustic and configuration stationarity assumption to obtain a reasonable estimate of the covariance of signal and noise. The solution can be obtained by equating the derivative with respect to B^H to zero. Thus

$$\hat{B} = \frac{\Phi^H R_{\mathcal{X}S^H}}{\Phi^H R_{\mathcal{X}} \Phi}, \quad W^H_{\text{m-mmse}} = \frac{\Phi \Phi^H R_S \phi_s^H}{\Phi^H (R_S \phi_s \phi_s^H + R_n)\Phi} \tag{5}$$

Additionally, we can incorporate the sparsity prior on B and derive the MMSE estimator as

$$\hat{S}, \hat{B} = \underset{B,S}{\arg\min}\{\|B^H \Phi^H \mathcal{X} - S\|_2 + \lambda_2 \|B\|_1 \quad \text{s.t.} \quad B^H \Phi^H \phi_s = 1\} \tag{6}$$

This formulation can be extended to incorporate the structured sparsity models pertained to the source representation [2].

III. Experiments and Conclusion

The experiments are conducted in two synthesized scenarios. (1) The reverberant scenario: 8-channel circular microphone array is used with radius 10 cm and 50 cm spatial resolution within an enclosure of dimension 3m × 3m × 3m; $RT_{60} \simeq 300$ ms. The signals are 1000 speech frames sampled at 8 kHz and analyzed using Han function of length 64 ms and the weights are estimated per 10 frames. (2) The farfield scenario: 8-channel uniform linear array is used with inter-element spacing equal to half of the wavelength and the directional resolution is 5°. The signals are 1000 trails random samples of a sinusoid at frequency 1024 Hz of length 1000 sampled at 16 kHz. However, only 5% of data samples are used to compute the statistics of the conventional beamformer (2). The sparse beamformers are computed using CVX package. The signal to noise ratio is 20 dB. We assume that the signal and noise samples are known so the only uncertainty is attributed to the number of reliable samples available for beamforming. Table I summarizes the results.

TABLE I: RMSE of signal recovery. The numbers in parenthesis show the performance of conventional beamforming formulation; in Reverb.: multipath conventional. The results of conventional beamformers using only direction of the desired source in a reverberant acoustic are omitted for being the worst.

Scenario	Beamf.	1 Source	2 Sources	3 Sources	4 Sources
Farfield	MVDR	0.05 (0.09)	0.06 (0.16)	0.12 (0.21)	0.18 (0.25)
	MMSE	0.05 (0.22)	0.05 (0.23)	0.08 (0.24)	0.08 (0.25)
Reverb.	MVDR	0.03 (0.52)	0.27 (0.54)	0.39 (0.83)	0.49 (0.84)
	MMSE	0.66 (0.83)	0.70 (0.84)	0.73 (0.96)	0.75 (0.96)

The sparse beamformer requires fewer parameters to estimate the solution than the conventional beamforming techniques hence, it enables accurate estimation when the number of reliable samples are limited and suggests a framework for missing data beamforming. Additionally, it enables better null steering and offers a more robust solution in multiparty recordings. Furthermore, we can calibrate the model parameters of an acoustic-informed beamformer using a known source signal at a given location.

References

[1] H. L. V. Trees, *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*. John Wiley and Sons, Inc., Print ISBN: 9780471093909, 2002.
[2] A. Asaei, M. Golbabaee, H. Bourlard, and V. Cevher, "Structured sparsity models for multiparty speech recovery from reverberant recordings," *arXiv:1210.6766*, 2012.