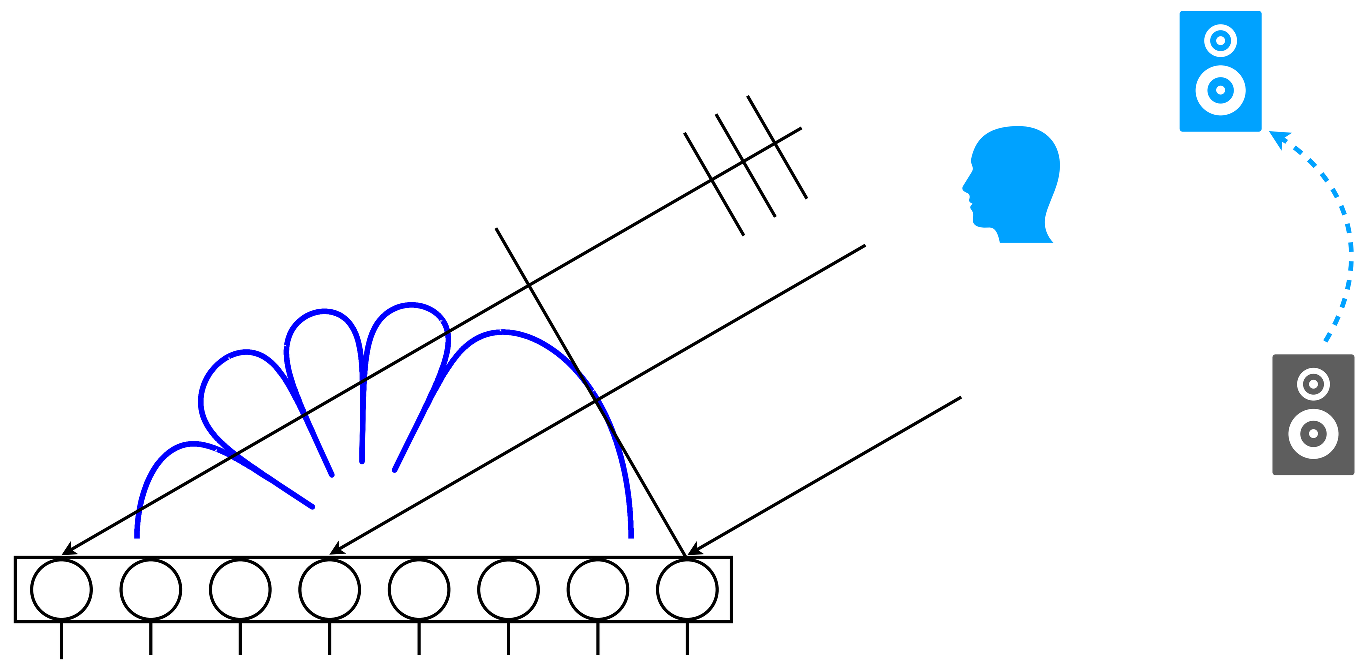
ON SPEECH ENHANCEMENT USING MICROPHONE ARRAYS IN THE PRESENCE OF CO-DIRECTIONAL INTERFERENCE

Xin Leng, Jingdong Chen, Jacob Benesty, and Israel Cohen

Introduction

- Microphone arrays have been used in a wide range of applications.
- Existing algorithms cannot deal with co-directional interference.
- A WL LCMV beamformer is presented.



Signal Model

A uniform linear microphone array of M sensors:

$$Y_m(\omega) = X_m(\omega) + J_m(\omega) + V_m(\omega)$$
$$= e^{-\jmath(m-1)\omega\tau_0\cos\theta_{\rm d}}\Big[\underbrace{X(\omega)}_{\text{Speech}} + \underbrace{J(\omega)}_{\text{Interference}}\Big] + \underbrace{V_m(\omega)}_{\text{Noise}}$$

$$\mathbf{y}(\omega) \triangleq \begin{bmatrix} Y_1(\omega) & Y_2(\omega) & \cdots & Y_M(\omega)\end{bmatrix}^T$$
$$= \mathbf{x}(\omega) + \mathbf{j}(\omega) + \mathbf{v}(\omega)$$
$$= \mathbf{d}(\omega, \cos\theta_{\rm d})\left[X(\omega) + J(\omega)\right] + \mathbf{v}(\omega)$$

Problem Formulation

Using WL estimation theory:

$$\widehat{X}(\omega) = \mathbf{h}^H(\omega)\mathbf{y}(\omega) + \mathbf{h}'^H(\omega)\mathbf{y}^*(\omega)$$
$$= \widetilde{\mathbf{h}}^H(\omega)\widetilde{\mathbf{y}}(\omega)$$
$$= \widetilde{\mathbf{h}}^H(\omega)\left[\widetilde{\mathbf{x}}(\omega) + \widetilde{\mathbf{j}}(\omega) + \widetilde{\mathbf{v}}(\omega)\right]$$

$$\widetilde{\mathbf{j}}(\omega) = \mathbf{d}_J(\omega, \cos\theta_{\rm d})J(\omega) + \widetilde{\mathbf{j}}'(\omega)$$
$$\widetilde{\mathbf{x}}(\omega) = \mathbf{d}_X(\omega, \cos\theta_{\rm d})X(\omega) + \widetilde{\mathbf{x}}'(\omega)$$
$$\widetilde{\mathbf{h}}(\omega) \triangleq \begin{bmatrix}\mathbf{h}(\omega) \\ \mathbf{h}'(\omega)\end{bmatrix}$$

$$\widehat{X}(\omega) = \widetilde{\mathbf{h}}^H(\omega)\left[\mathbf{d}_X(\omega, \cos\theta_{\rm d})X(\omega) + \widetilde{\mathbf{x}}'(\omega) + \mathbf{d}_J(\omega, \cos\theta_{\rm d})J(\omega) + \widetilde{\mathbf{j}}'(\omega) + \widetilde{\mathbf{v}}(\omega)\right]$$

Problem: finding an optimal $\widetilde{\mathbf{h}}(\omega)$ so that $\widehat{X}(\omega)$ is a good estimate of $X(\omega)$

April, 2018

WL LCMV Beamformer

$$\arg\min_{\widetilde{\mathbf{h}}(\omega)} \widetilde{\mathbf{h}}^H(\omega)\mathbf{\Phi}_{\rm in}(\omega)\widetilde{\mathbf{h}}(\omega)$$
$$\text{s.t.}\quad \widetilde{\mathbf{h}}^H(\omega)\mathbf{d}_X(\omega, \cos\theta_{\rm d}) = 1$$
$$\widetilde{\mathbf{h}}^H(\omega)\mathbf{d}_J(\omega, \cos\theta_{\rm d}) = 0$$

$$\widetilde{\mathbf{h}}_{\rm LCMV}(\omega) = \mathbf{\Phi}_{\rm in}^{-1}(\omega)\mathbf{D}(\omega)\left[\mathbf{D}^H(\omega)\mathbf{\Phi}_{\rm in}^{-1}(\omega)\mathbf{D}(\omega)\right]^{-1}\mathbf{i}_{2,1}$$

Case Study: Spatially White Noise

Spatially white, identically distributed, and circular Gaussian noise.

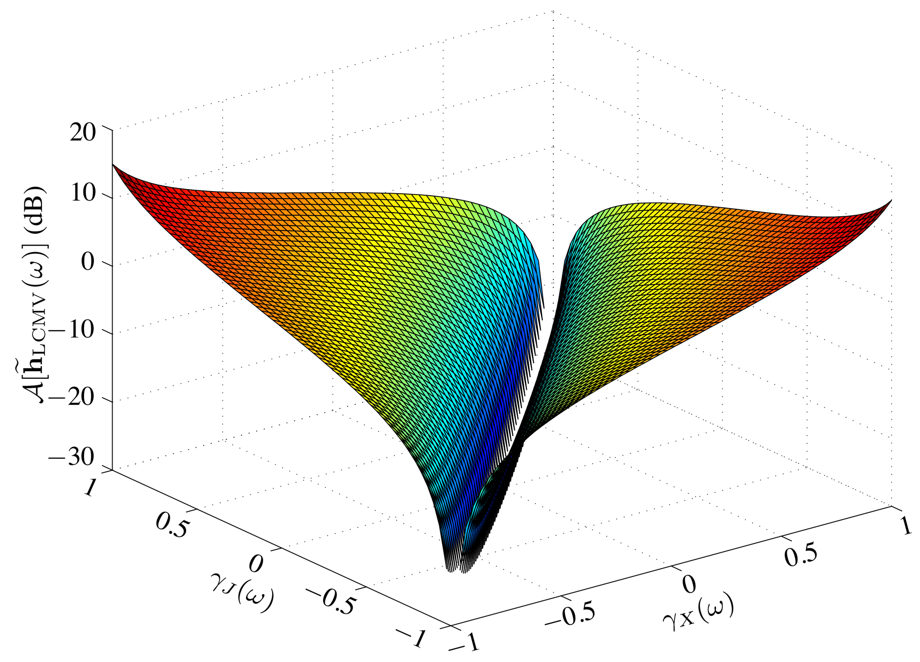


Fig. 1. Theoretical array gain.

$$\mathcal{A}[\widetilde{\mathbf{h}}_{\rm LCMV}(\omega)] = \frac{M|\gamma_X(\omega)-\gamma_J(\omega)|^2\left[\text{iSNR}^{-1}(\omega)+\text{iSIR}^{-1}(\omega)\right]}{(1+|\gamma_J(\omega)|^2)\text{iSNR}^{-1}(\omega)+M\left[1-|\gamma_X(\omega)|^2+\text{iSIR}^{-1}(\omega)(1-|\gamma_J(\omega)|^2)\right]}$$

Simulations

- Image model with $\beta = 0.8$
- ULMA with 7 sensors, broadside direction
- Speech from TIMIT database, car and white Gaussian noises

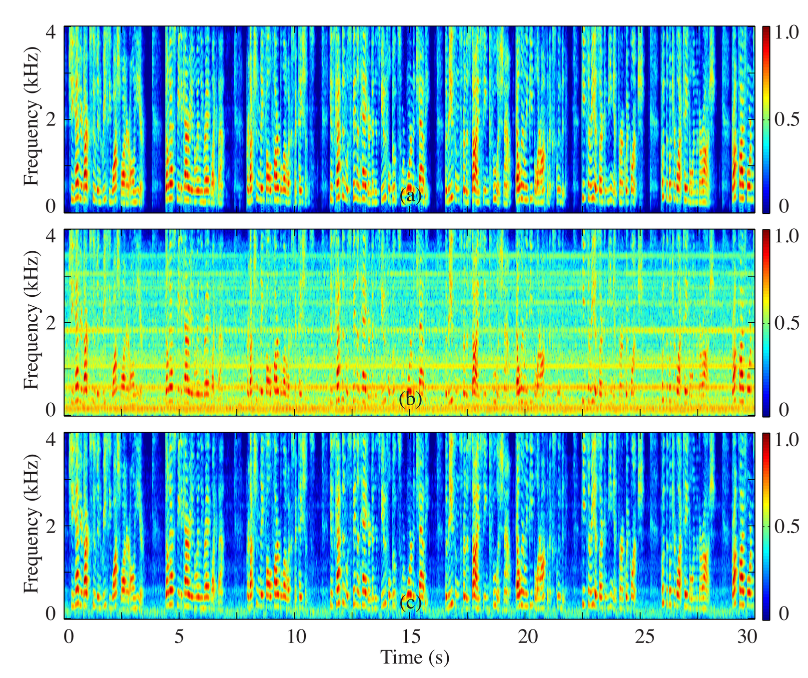


Fig. 2. Spectrograms in car noise. The input SIR is 5 dB.

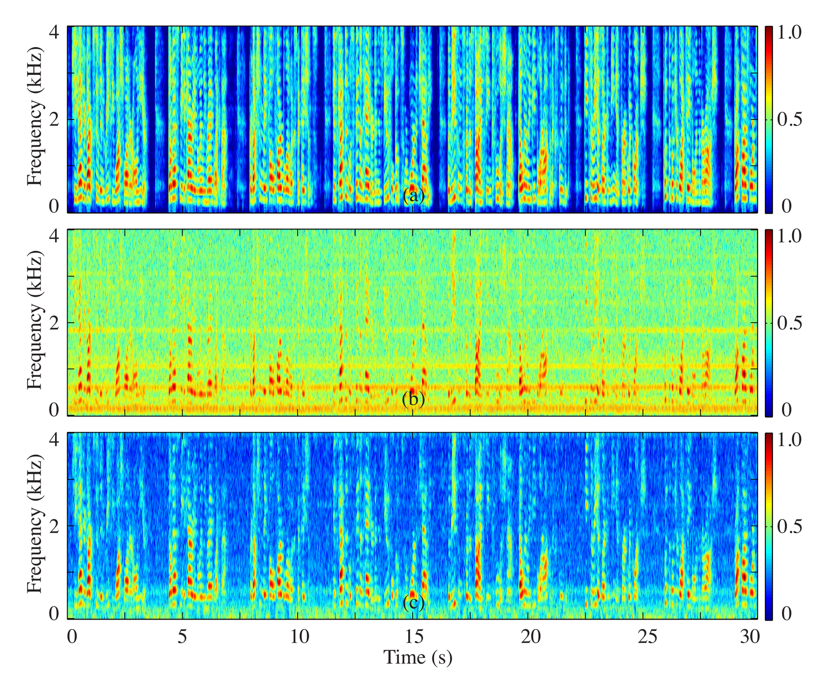


Fig. 3. Spectrograms in car-plus-Gaussian noise. The input SIR is 5 dB, and input SNR is 10 dB.

Conclusions

- Beamforming in the presence of noise and co-directional interference.
- A WL LCMV beamformer can preserve the signal of interest while reducing interference from the same direction as the desired source signal.