TECHNION
Israel Institute of Technology

The Andrew and Erna Viterbi Faculty of
ELECTRICAL & COMPUTER ENGINEERING

Beamforming-Based Multichannel Acoustic Echo Cancellation

Yuval Konforti
M.Sc. Thesis Seminar

Supervisors: Prof. Israel Cohen and Dr. Baruch Berdugo

Outline

Background and motivation

Acoustic Echo Cancellation

Region-of-Interest Beamforming

Multichannel Acoustic Echo Cancellation with Beamforming in Dynamic Environments

Array Geometry Optimization for Region-of-Interest Broadband Beamforming

Conclusions

Research Contributions

Future research

Publication Info

- Y. Konforti, I. Cohen, and B. Berdugo, ”Array geometry optimization for region-of-interest broadband beamforming”, in Proc. IEEE International Workshop on Acoustic Signal Enhancement (IWAENC), pp. 1–5, 2022.
- Y. Konforti, I. Cohen, and B. Berdugo, ”Multichannel acoustic echo cancellation with beamforming in dynamic environments”, submitted to IEEE Open Journal of Signal Processing, 2023.

TECHNION
Israel Institute
of Technology

The Andrew and Erna Viterbi Faculty of
ELECTRICAL & COMPUTER
ENGINEERING

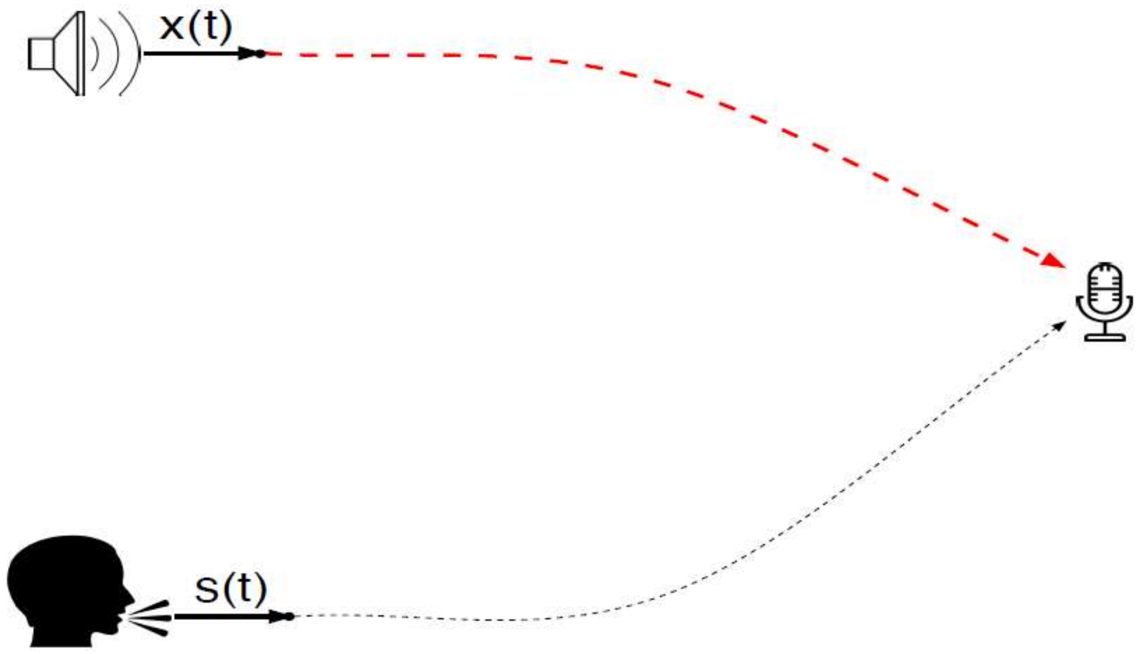
Background and Motivation

TECHNION
Israel Institute
of Technology

The Andrew and Erna Viterbi Faculty of
ELECTRICAL & COMPUTER
ENGINEERING

Acoustic Echo Cancellation

- Echo can interfere and disrupt conversations.
- Stems from the acoustic coupling between a loudspeaker and a microphone.
- Echo is eliminated with an acoustic echo canceller (AEC).
  - Usually solved by a normalized least-mean-square filter (NLMS) approach.
  - Involves a double-talk detector.



Applications

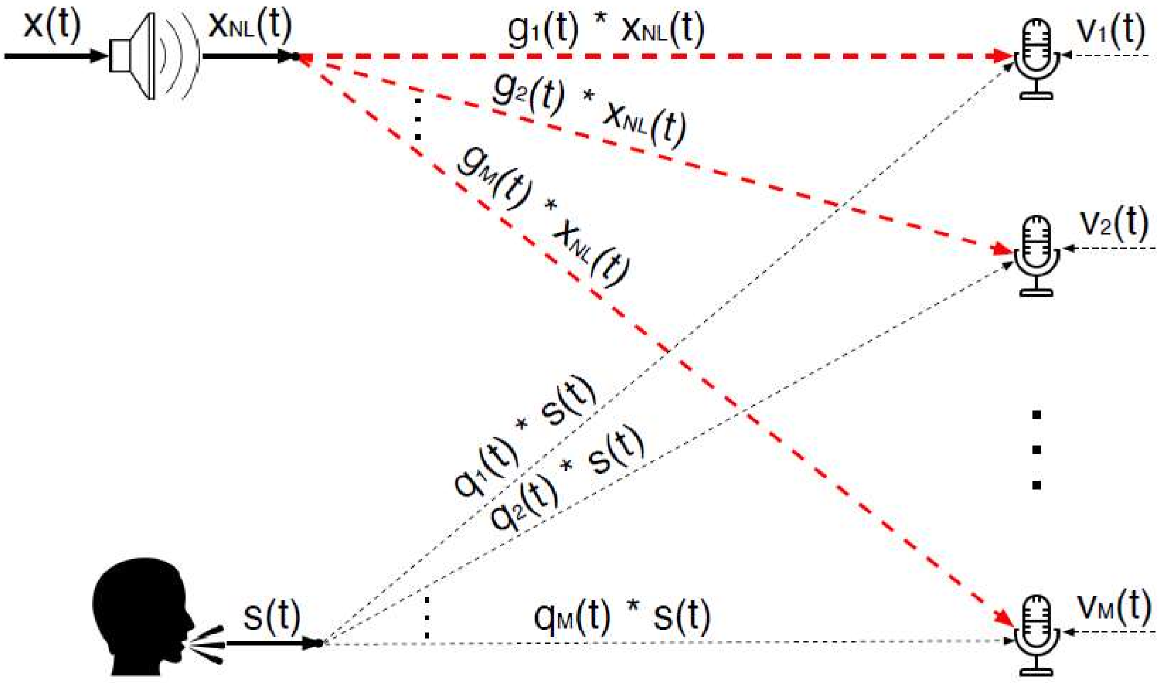
APPLE INC.

zoom

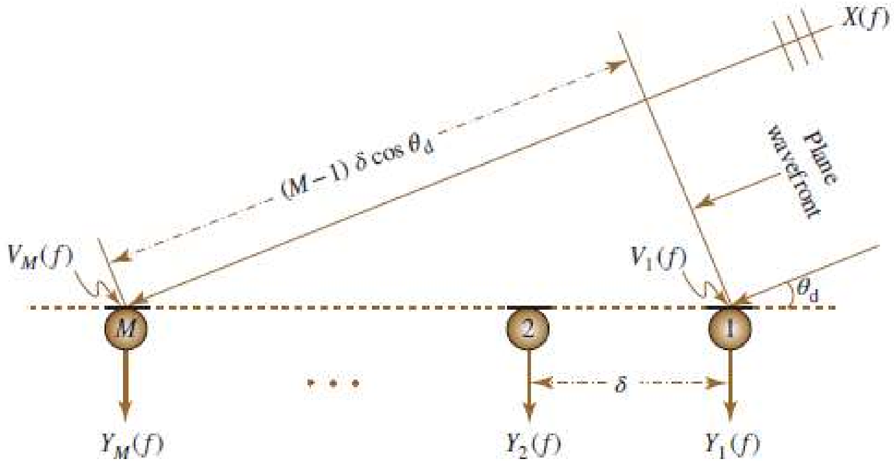
webex
by CISCO

Multichannel Acoustic Echo Cancellation

- In real environments two phenomena worsen performance:
  - Nonlinear loudspeaker distortion
  - Background noise
- Multichannel AEC can be used to mitigate these effects.



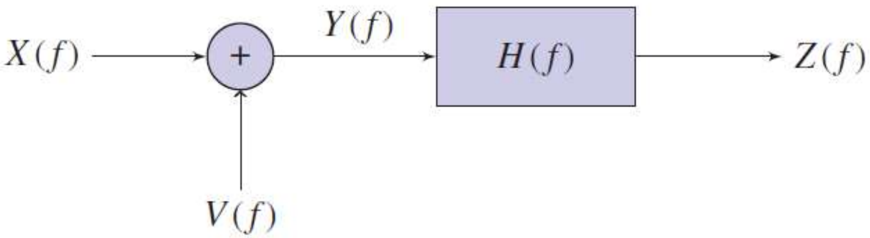
Spatial Filtering



- A signal is propagated in space.
- The signal arrives from a certain direction.
- A sensor array is utilized. The time difference of arrival (TDOA) between the sensors is in accordance with the direction.

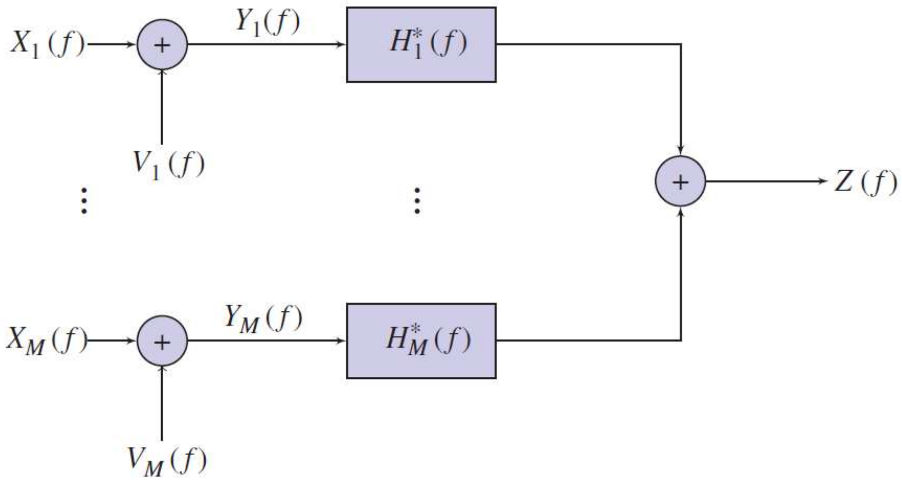
Beamforming – a Sum of Digital Filters

- With one sensor:



$$Z(f) = H(f)Y(f)$$

- With multiple sensors:



$$Z(f) = \sum_{m=1}^{M} H_m^*(f)Y_m(f)$$
$$= \mathbf{h}^H(f)\mathbf{y}(f)$$

Sensor Array Beamforming

- We define the steering vector

$$\mathbf{d}\left(f, \cos\theta\right) = \begin{bmatrix} 1 & e^{-\jmath 2\pi f \tau_0 \cos\theta} & \cdots & e^{-\jmath (M-1) 2\pi f \tau_0 \cos\theta} \end{bmatrix}^T$$

- Each sensor utilizes a digital filter $H_m(f)$.
- The response is measured by the beampattern:
  - Angle dependent.
  - Frequency dependent.

$$\mathcal{B}\left[\mathbf{h}(f), \cos\theta\right] = \mathbf{d}^H\left(f, \cos\theta\right)\mathbf{h}(f) = \sum_{m=1}^{M} H_m(f) e^{\jmath (m-1) 2\pi f \tau_0 \cos\theta}$$

TECHNION Israel Institute of Technology

The Andrew and Erna Viterbi Faculty of ELECTRICAL & COMPUTER ENGINEERING

Multiframe Filtering

- In the short-time fourier transform (STFT) domain, speech signals are correlative between frames.
- The inter-frame correlations must be updated, as they highly vary in time.
- Multiframe and Multichannel filters were used for speech separation and noise reduction(*), yet not for AEC.

(*) E. A. P. Habets, J. Benesty, and J. Chen, "Multi-microphone noise reduction using interchannel and interframe correlations," in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 305–308, 2012.
(*) Z. Zhang, Y. Xu, M. Yu, S. X. Zhang, L. Chen, D. S. Williamson, and D. Yu, "Multi-channel multi-frame ADL-MVDR for target speech separation," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 29, pp. 3526–3540, 2021.
(*) M. Tammen and S. Doclo, "Deep multi-frame MVDR filtering for binaural noise reduction," in Proc. International Workshop on Acoustic Signal Enhancement (IWAENC), 2022.

AEC Performance Measures

- Echo-return loss enhancement (ERLE):
  - $y_1(t)$ - Echo component at reference microphone.
  - $y_{re}(t)$- Residual echo component.

$$\xi(t) = \frac{\mathrm{LPF}\left\{y_1^2(t)\right\}}{\mathrm{LPF}\left\{y_{\mathrm{re}}^2(t)\right\}}$$

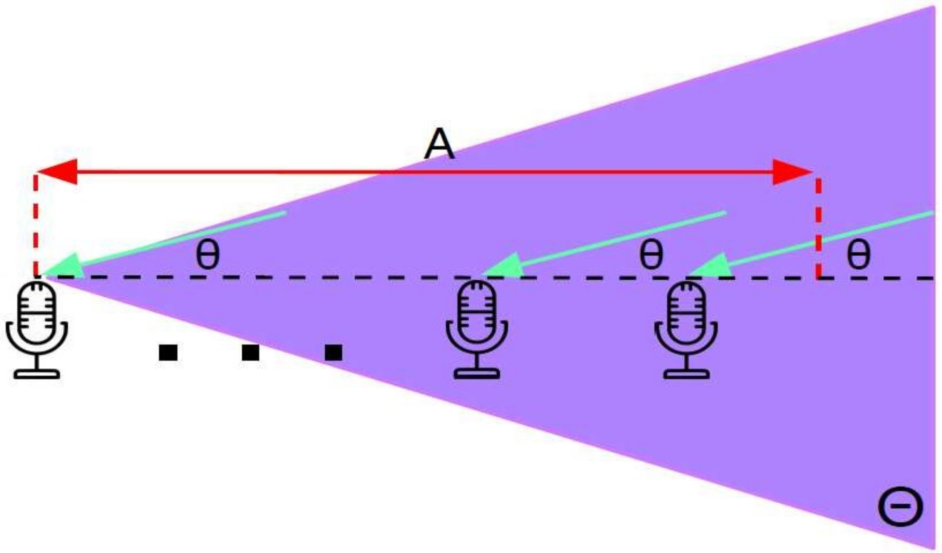
- Distortion Index (DI):
  - $u_1(t)$ - Desired component at reference microphone.
  - $u_f(t)$ - Filtered desired component.

$$\nu(t) = \frac{\mathrm{LPF}\left\{\left[u_1(t) - u_{\mathrm{f}}(t)\right]^2\right\}}{\mathrm{LPF}\left\{u_1^2(t)\right\}}$$

TECHNION
Israel Institute of Technology

The Andrew and Erna Viterbi Faculty of
ELECTRICAL & COMPUTER ENGINEERING

Region-of-Interest Beamforming



- In many applications, the source location is unknown, but it can be assumed it is in a region-of-interest (ROI).
- Source localization / Direction-of-arrival (DOA) estimation may be employed to track a moving source.
- While the coefficients change, the array geometry is preserved.

TECHNION
Israel Institute
of Technology

The Andrew and Erna Viterbi Faculty of
ELECTRICAL & COMPUTER
ENGINEERING

Applications

Array Geometry Optimization

- Typically, the beamforming coefficients are found per given geometry. Usually symmetric geometries are considered:
  - Uniform Linear Arrays (ULAs).
  - Rectangular Arrays.
  - Uniform Circular Arrays (UCAs).
  - Uniform Concentric Circular Arrays (UCCAs).
- Differential Microphone Arrays (DMAs) produce an approximate frequency invariant (FI) response.
- The geometry of a microphone array has an important impact on beamforming performance.

TECHNION
Israel Institute
of Technology

The Andrew and Erna Viterbi Faculty of
ELECTRICAL & COMPUTER
ENGINEERING

Optimization Methods

- Two common methods for array geometry optimization:
  - Greedy-based approaches.
  - Genetic Algorithms.
- Greedy-based approaches find the best position for the placed microphone in each step, but the overall result may not be optimal.
- Genetic algorithms are unstable and may also converge to nonoptimal results.
- The study in (*) optimize the geometry for a ROI, but only for narrowband signals.

(*) X. Chen, C. Pan, J. Chen, and J. Benesty, "Planar array geometry optimization for region sound acquisition," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 756–760, 2021.

Beamforming Performance Measures

- White Noise Gain: $$\mathcal{W}\left[\mathbf{h}(\mathrm{x},\omega,\theta)\right] \triangleq \frac{\left|\mathbf{d}^H(\mathrm{x},\omega,\theta)\,\mathbf{h}(\mathrm{x},\omega,\theta)\right|^2}{\mathbf{h}^H(\mathrm{x},\omega,\theta)\,\mathbf{h}(\mathrm{x},\omega,\theta)}$$
- Directivity Factor (DF): $$\mathcal{D}\left[\mathbf{h}(\mathrm{x},\omega,\theta)\right] \triangleq \frac{\left|\mathbf{d}^H(\mathrm{x},\omega,\theta)\,\mathbf{h}(\mathrm{x},\omega,\theta)\right|^2}{\mathbf{h}^H(\mathrm{x},\omega,\theta)\,\mathbf{\Gamma}(\mathrm{x},\omega)\,\mathbf{h}(\mathrm{x},\omega,\theta)}$$

  where $$\Gamma_{i,j}(\mathrm{x},\omega) = \frac{\sin\left(\omega\left(x_i - x_j\right)/c\right)}{\omega\left(x_i - x_j\right)/c}, \quad 1 \le i,j \le M$$
- Directivity Index: $$\mathcal{DI}_{[\omega_L,\omega_H]}\left[\mathbf{h}(\mathrm{x},\omega,\theta)\right] \triangleq \frac{\int_{\omega_L}^{\omega_H}\left|\mathbf{d}^H(\mathrm{x},\omega,\theta)\,\mathbf{h}(\mathrm{x},\omega,\theta)\right|^2 d\omega}{\int_{\omega_L}^{\omega_H}\mathbf{h}^H(\mathrm{x},\omega,\theta)\,\mathbf{\Gamma}(\mathrm{x},\omega)\,\mathbf{h}(\mathrm{x},\omega,\theta)\,d\omega}$$

TECHNION
Israel Institute
of Technology

The Andrew and Erna Viterbi Faculty of
ELECTRICAL & COMPUTER
ENGINEERING

Multichannel Acoustic Echo Cancellation with Beamforming in Dynamic Environments

A Beamforming-Based Approach

- $\boldsymbol{y}(k,n)$- Echo component received by the array.
- $\boldsymbol{u}(k,n)$- Desired component received signal by the array.
- $\boldsymbol{v}(k,n)$- Background noise received signal by the array.
- $\boldsymbol{d}(k,n)$- total received signal by the array.

$$\hat{U}(k,n) = \mathbf{h}^H(k,n)\,\mathbf{d}(k,n)$$

$$\mathbf{d}(k,n) = \mathbf{y}(k,n) + \mathbf{u}(k,n) + \mathbf{v}(k,n)$$

Linear-Constraint-Minimum-Variance Beamforming

- Theoretically enforces:
  - Complete echo cancellation.
  - No distortion.
  - Minimum residual noise.
- Ideal ERLE and DI.
- Requires accurate estimates of $\boldsymbol{g}(k)$ and $\boldsymbol{q}(k)$.

$$\mathbf{h}^*(k,n) = \underset{\mathbf{h}(k,n)}{\arg\min} \|\mathbf{h}(k,n)\|^2$$

$$\text{s.t.} \quad \mathcal{C}_1[\mathbf{h}(k,n)] : \mathbf{h}^H(k,n)\,\mathbf{g}(k) = 0$$

$$\mathcal{C}_2[\mathbf{h}(k,n)] : \mathbf{h}^H(k,n)\,\mathbf{q}(k) = 1$$

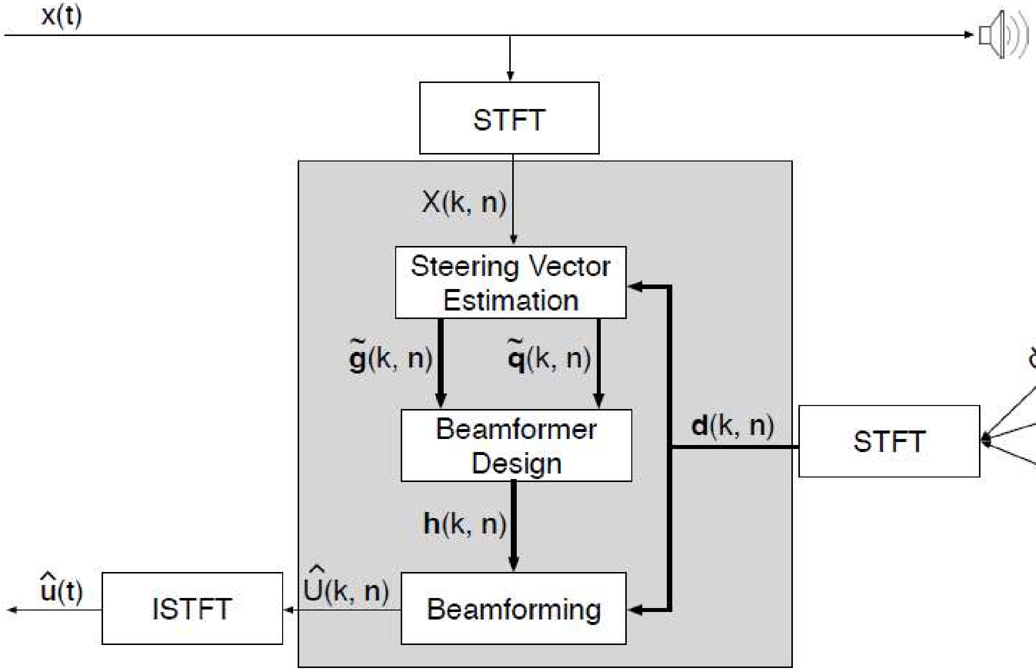
$$\mathbf{h}^*(k,n) = \mathbf{C}(k)\left[\mathbf{C}^H(k)\,\mathbf{C}(k)\right]^{-1}\mathbf{i}_c$$

where $\mathbf{C}(k,n) = [\mathbf{q}(k), \mathbf{g}(k)]$

$$\mathbf{i}_c = [1,0]^T$$

AEC Scheme

- Procedure:
  1. Steering vectors $\boldsymbol{g}(k)$ and $\boldsymbol{q}(k)$ are estimated.
  2. Beamformer $\boldsymbol{h}(k,n)$ is designed.
  3. Estimate $\widehat{U}(k,n)$ is produced.
- No double talk detection used.



Utilizing Multiple Sensors – A Frame Invariant Expression

- Neglecting nonlinear loudspeaker distortion.
- Neglecting background noise.
- Assuming a static environment in the last L frames.
- Utilizing the Multiplicative Transfer Function (MTF) approximation (*).

$$D_m(k, n-l+1) = G_m(k)\,X(k, n-l+1) + Q_m(k)\,S(k, n-l+1)$$

- For any two microphones m_1 and m_2:

$$\frac{Q_{m_1}(k)}{Q_{m_2}(k)} = \frac{D_{m_1}(k, n-l+1) - G_{m_1}(k)\,X(k, n-l+1)}{D_{m_2}(k, n-l+1) - G_{m_2}(k)\,X(k, n-l+1)}$$

(*) Y. Avargel and I. Cohen, "On multiplicative transfer function approximation in the short-time fourier transform domain," *IEEE Signal Processing Letters*, vol. 14, pp. 337–340, 2007.

Utilizing Multiple Frames

- For any two recent frames l_1 and l_2:

$$\frac{D_{m_1}(k,n-l_1+1)-G_{m_1}(k)X(k,n-l_1+1)}{D_{m_2}(k,n-l_1+1)-G_{m_2}(k)X(k,n-l_1+1)}=\frac{D_{m_1}(k,n-l_2+1)-G_{m_1}(k)X(k,n-l_2+1)}{D_{m_2}(k,n-l_2+1)-G_{m_2}(k)X(k,n-l_2+1)}$$

- Quadratic elements of $G_{m_1}(k)G_{m_2}(k)$ are reduced.
- A linear equation with respect to $G_{m_1}(k)$ and $G_{m_2}(k)$ is obtained:

$$\begin{aligned}&G_{m_1}(k)\left[X(k,n-l_1+1)D_{m_2}(k,n-l_2+1)-X(k,n-l_2+1)D_{m_2}(k,n-l_1+1)\right]+\\&G_{m_2}(k)\left[X(k,n-l_2+1)D_{m_1}(k,n-l_1+1)-X(k,n-l_1+1)D_{m_1}(k,n-l_2+1)\right]=\\&D_{m_1}(k,n-l_1+1)D_{m_2}(k,n-l_2+1)-D_{m_1}(k,n-l_2+1)D_{m_2}(k,n-l_1+1)\end{aligned}$$

A Linear Set of Equations

- Overall, for any pick of $1 \leq m_1, m_2 \leq M$ and $1 \leq l_1, l_2 \leq L$, we get such an equation.
- Uninformative equations:
  - $m_1 = m_2$
  - $l_1 = l_2$
  - Swapping $m_1 \Leftrightarrow m_2$
  - Swapping $l_1 \Leftrightarrow l_2$
- Must have more equations than variables: $M \leq \binom{M}{2}\binom{L}{2} \Rightarrow L(L-1)(M-1) \geq 4$
- Must have multiple frames, multiple sensors, and $M \geq 3$ or $\mathrm{L} \geq 3$.

A Least-Mean-Squares Approach

- Equations may contradict due to assumptions.
- A least-mean-squares (LMS) solution provides a good estimate.
- As L grows:
  - More equations are added.
  - The environment is assumed to be static for longer periods.
- As M grows:
  - More equations are added.
  - More variables are added.

Matrix Formulation

Inputs: M, L,

$X(k, n-l+1)$, $D_m(k, n-l+1)$ $\quad 1 \le l \le L \quad 1 \le m \le M$

Outputs: $\tilde{G}_m(k,n)$ $\quad 1 \le m \le M$

Create list M_{list} of $\binom{M}{2}$ non-repetitive pairs (m_1, m_2)

Create list L_{list} of $\binom{L}{2}$ non-repetitive pairs (l_1, l_2)

$\mathbf{A}(k,n) \leftarrow \mathbf{0}_{\binom{M}{2}\binom{L}{2}\times M}$

for $i = 1, 2, \ldots, \binom{M}{2}$ **do**

$(m_1, m_2) \leftarrow M_{\text{list}}(i)$

for $j = 1, 2, \ldots, \binom{L}{2}$ **do**

$(l_1, l_2) \leftarrow L_{\text{list}}(j)$

$$\mathbf{A}_{\left[(i-1)\binom{L}{2}+j, m_1\right]}(k,n) \leftarrow X(k, n-l_1+1) D_{m_2}(k, n-l_2+1) - X(k, n-l_2+1) D_{m_2}(k, n-l_1+1)$$

$$\mathbf{A}_{\left[(i-1)\binom{L}{2}+j, m_2\right]}(k,n) \leftarrow X(k, n-l_2+1) D_{m_1}(k, n-l_1+1) - X(k, n-l_1+1) D_{m_1}(k, n-l_2+1)$$

$$\mathbf{b}_{\left[(i-1)\binom{L}{2}+j\right]}(k,n) \leftarrow D_{m_1}(k, n-l_1+1) D_{m_2}(k, n-l_2+1) - D_{m_1}(k, n-l_2+1) D_{m_2}(k, n-l_1+1)$$

end for

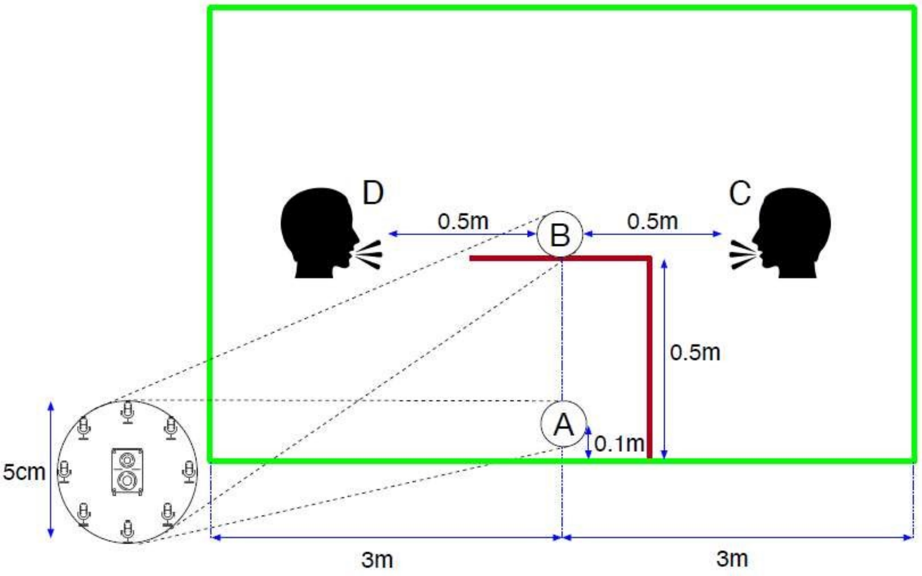
end for

$$\left[\tilde{G}_1(k,n), \tilde{G}_2(k,n), \ldots, \tilde{G}_M(k,n)\right]^T \leftarrow \mathbf{A}^{\dagger}(k,n)\,\mathbf{b}(k,n)$$

$$\tilde{\mathbf{g}}(k,n) = \left[1, \frac{\tilde{G}_2(k)}{\tilde{G}_1(k)}, \ldots, \frac{\tilde{G}_M(k)}{\tilde{G}_1(k)}\right]^T$$

$$\tilde{\mathbf{q}}(k,n) = \left[1, \frac{D_2(k,n) - \tilde{G}_2(k,n) X(k,n)}{D_1(k,n) - \tilde{G}_1(k,n) X(k,n)}, \ldots, \frac{D_M(k,n) - \tilde{G}_M(k,n) X(k,n)}{D_1(k,n) - \tilde{G}_1(k,n) X(k,n)}\right]^T.$$

Simulation Configuration



- Dynamic environment.
- Double-talk.
- Low Signal-to-Echo Ratio in 4 scenarios:
  1. Speakerphone at A, talker at C.
     -17.89 dB.
  2. Speakerphone at A, talker at D.
     -18.7 dB.
  3. Speakerphone at B, talker at C.
     -15.27 dB.
  4. Speakerphone at B, talker at D.
     -16.89 dB.

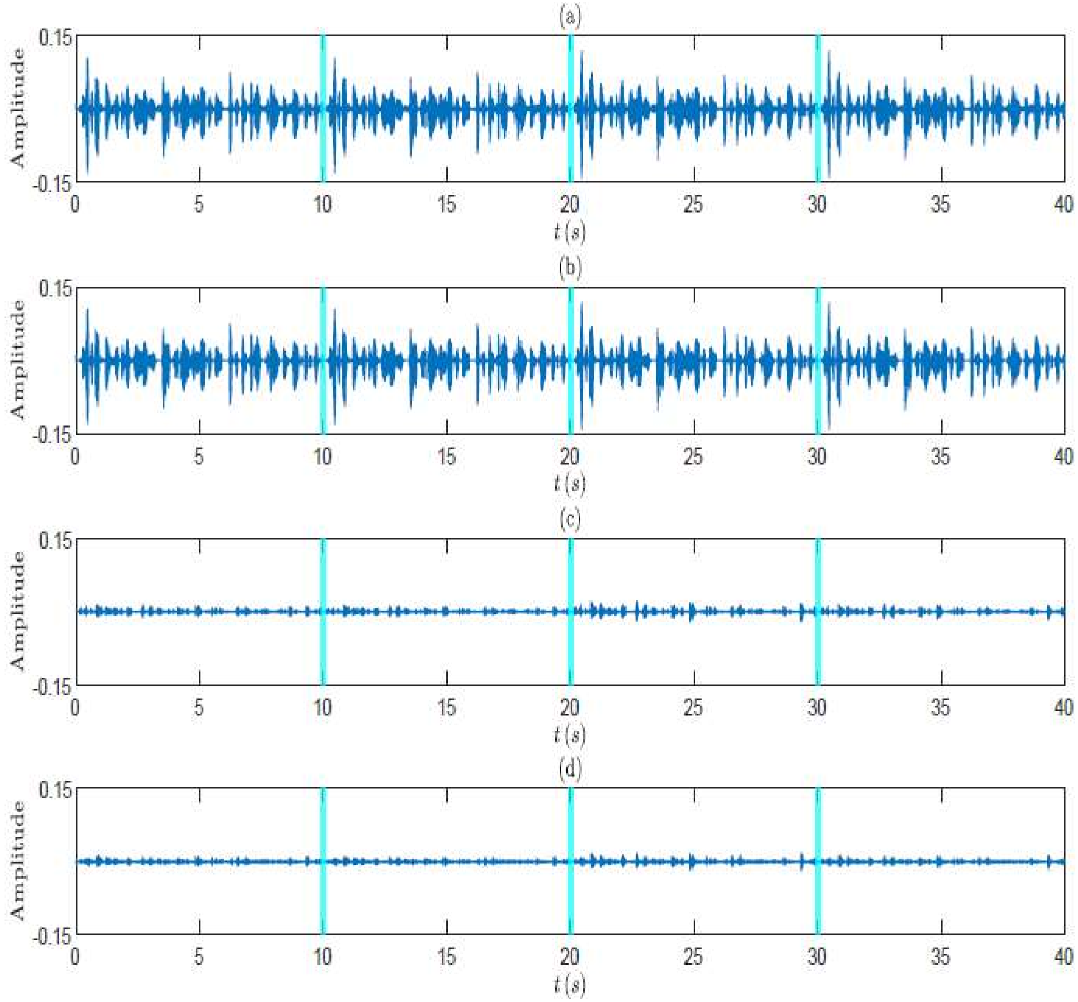
TECHNION Israel Institute of Technology

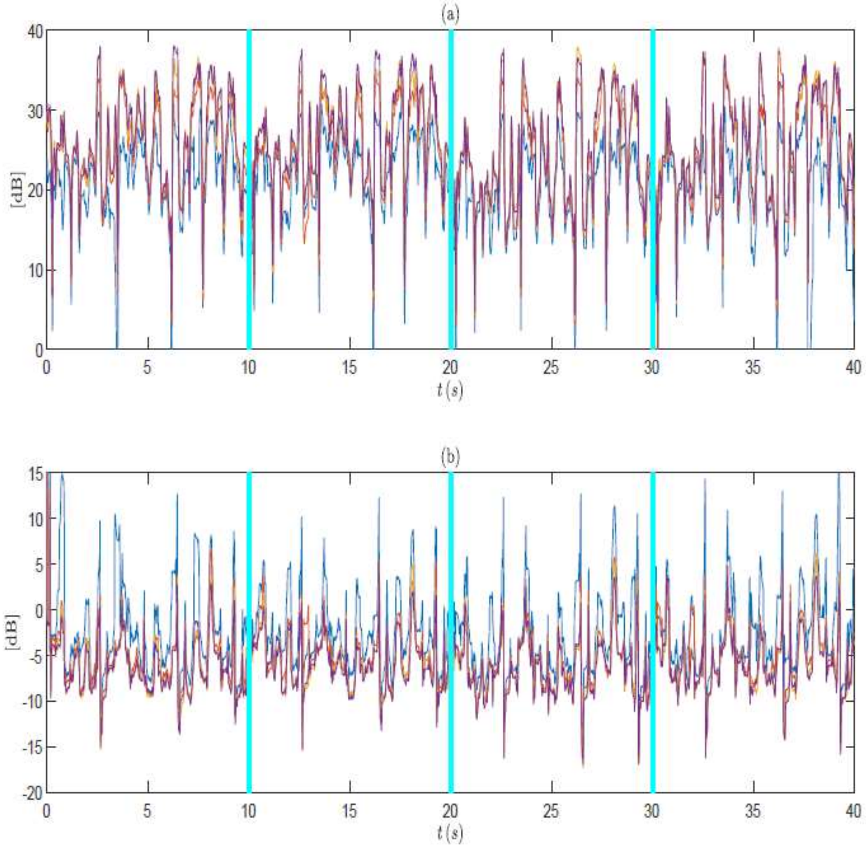
The Andrew and Erna Viterbi Faculty of ELECTRICAL & COMPUTER ENGINEERING

Results - Echo Cancellation

a. The total received signal in the reference microphone $d_1(t)$
b. The echo component signal in the reference microphone $y_1(t)$
c. The desired component signal in the reference microphone $u_1(t)$
d. The beamformer output signal $\hat{u}(t)$.

- Good echo cancellation despite significant acoustic coupling.





Results - Performance as Function of Microphones

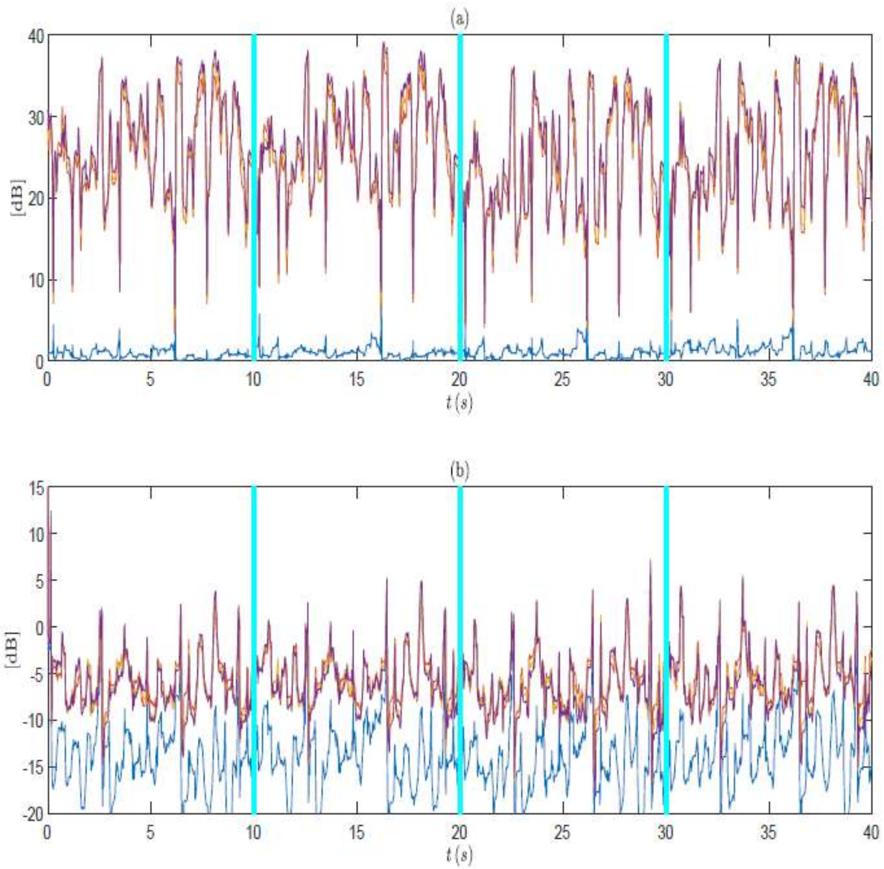
a. ERLE.

b. DI.

- Blue - $M = 2$, Red - $M = 3$, Yellow - $M = 4$, Purple - $M = 5$. $\mathrm{L} = 4$.
- Significant change only between $M = 2$ and $M = 3$.
- Performance limit. Increasing M also increases number of variables in estimation process.

TECHNION Israel Institute of Technology

The Andrew and Erna Viterbi Faculty of ELECTRICAL & COMPUTER ENGINEERING



Results - Performance as Function of Frames

a. ERLE.
b. DI.

- Blue - $L = 2$. Red - $L = 3$.
  Yellow - $L = 4$. Purple - $L = 5$. $M = 4$.
- $L = 2$ are insufficient for the proposed approach.
- Performance limit. Increasing L also relies more strongly on a longer period where the environment is static.

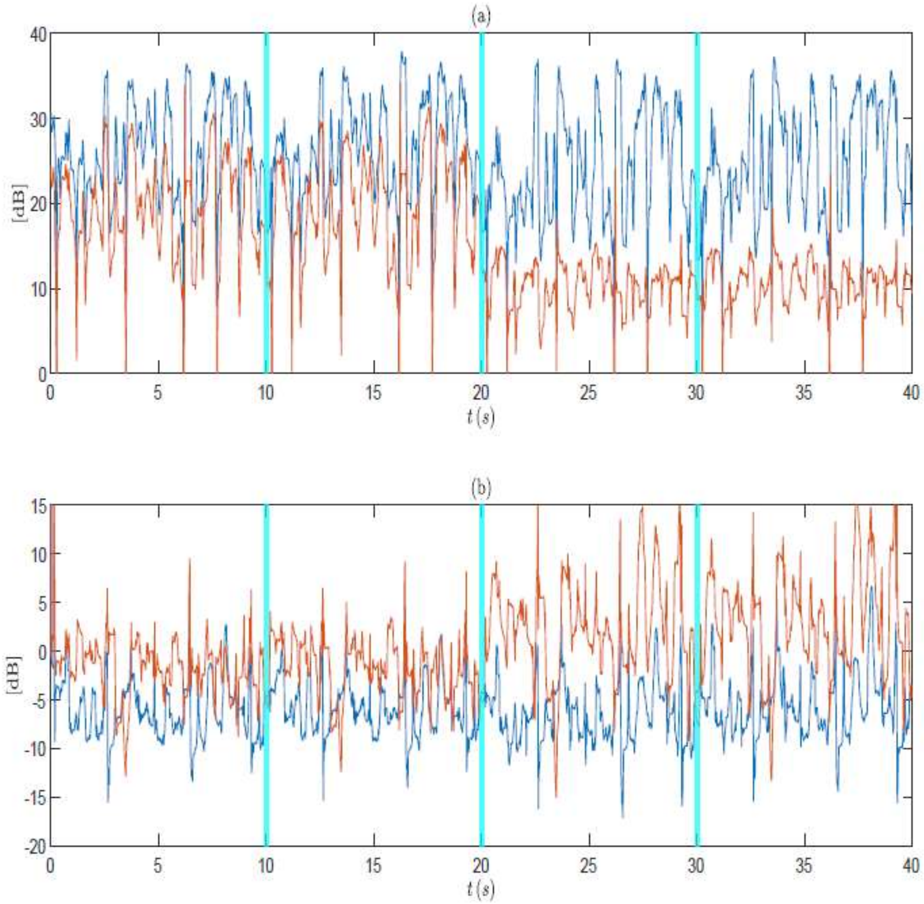
Results - Method Comparison (*)

a. ERLE.

b. DI.

- The NLMS filter in (*) was trained in a period like the first segment, but with no near-end speech.
- Blue – proposed. Red – competing.

- Improvement in both ERLE and DI.
- Significant improvement after the speakerphone moves due to changing reflections. The NLMS filter is irrelevant once the reflections have changed.



(*) H. Huang, C. Hofmann, W. Kellermann, J. Chen, and J. Benesty, "Multiframe echo suppression based on orthogonal signal decompositions," in *Proc. Speech Communication; ITG Symposium*, pp. 287–291, 2016.

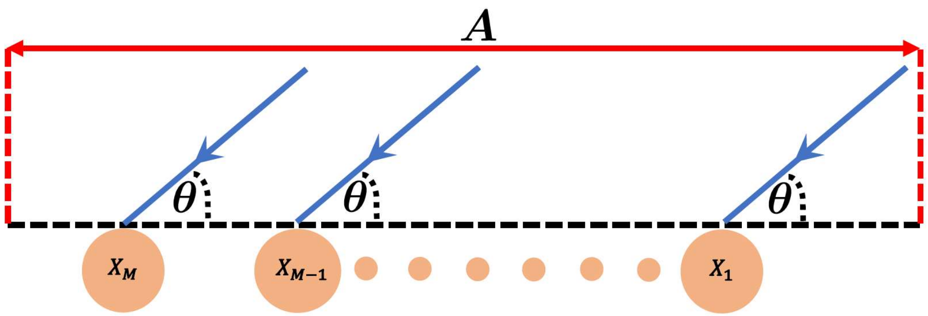
TECHNION
Israel Institute
of Technology

The Andrew and Erna Viterbi Faculty of
ELECTRICAL & COMPUTER
ENGINEERING

Array Geometry Optimization for Region-of-Interest Broadband Beamforming

A Variable Geometry

- M omnidirectional microphones are placed nonuniformly across a linear aperture A in placements x_m, $m = 1, 2, \dots, M$.



- The steering vector is

$$\mathbf{d}(\mathbf{x}, \omega, \theta) = \left[e^{-\jmath \frac{\omega}{c} x_1 \cos\theta}, e^{-\jmath \frac{\omega}{c} x_2 \cos\theta}, \dots, e^{-\jmath \frac{\omega}{c} x_M \cos\theta}\right]^T$$

TECHNION Israel Institute of Technology

The Andrew and Erna Viterbi Faculty of ELECTRICAL & COMPUTER ENGINEERING

Coefficients Dependent on Geometry

- For different geometries, the coefficients are designed differently. For a geometry x and look direction θ, we define the coefficients vector as

$$\mathbf{h}(\mathbf{x},\omega,\theta) \triangleq [H_1(\mathbf{x},\omega,\theta), H_2(\mathbf{x},\omega,\theta), \ldots, H_M(\mathbf{x},\omega,\theta)]^T$$

- With geometry x, the beampattern directed toward $\tilde{\theta}$ has a response at angle θ of

$$\mathcal{B}\left[\mathbf{h}\left(\mathbf{x},\omega,\tilde{\theta}\right),\theta\right] = \mathbf{d}^H(\mathbf{x},\omega,\theta)\,\mathbf{h}\left(\mathbf{x},\omega,\tilde{\theta}\right)$$

Problem Formulation

- Our objective is to find the optimal array geometry x, that maximizes the worst-case directivity index, in an ROI Θ. Each beamformer, directed toward θ, must admit to the distortionless constraint, have sufficient WNG, and maintain minimal distances.

$$
\begin{aligned}
\mathbf{x}^* = \arg\max_{\mathbf{x}} \quad & \min_{\theta\in\Theta} \mathcal{DI}_{[\omega_L,\omega_H]}\left[\mathbf{h}\left(\mathbf{x},\omega,\theta\right)\right] \\
\text{s.t.} \quad & \mathcal{B}\left[\mathbf{h}\left(\mathbf{x},\omega,\theta\right),\theta\right] = 1 && \forall\theta\in\Theta, \forall\omega\in\Omega \\
& \mathcal{W}\left[\mathbf{h}\left(\mathbf{x},\omega,\theta\right)\right] \geq \delta && \forall\theta\in\Theta, \forall\omega\in\Omega \\
& |x_i - x_j| \geq d_c && \forall i,j\in[1,M], i\neq j \\
& 0 \leq x_m \leq A && \forall m\in[1,M]
\end{aligned}
$$

- Not convex. Cannot be solved by convex optimization algorithms. 👎

Formalizing a Solvable Problem - Constraints

- Consider N candidate microphone locations, Q frequencies, and P look directions.
- To guarantee number of microphones:

$$\mathcal{C}_1\left[\mathbf{s}\right]:\mathbf{s}^H\mathbf{i}_N=M$$

- To guarantee minimal spacing:

$$\mathcal{C}_2\left[\mathbf{s}\right]:\mathbf{s}^H U\leq\mathbf{i}_G^T$$

$$\mathbf{s}=[1,0,1,0,1]^{\mathrm{T}}$$

- To guarantee the distortionless constraint:

$$\mathcal{C}_3\left[\mathbf{h}_{\text{tot}}\left(\omega,\theta\right)\right]:\mathbf{d}_{\text{tot}}^H\left(\omega_q,\theta_p\right)\mathbf{h}_{\text{tot}}\left(\omega_q,\theta_p\right)=1 \qquad \forall p\in\left[1,P\right],\quad\forall q\in\left[1,Q\right]$$

- To guarantee the desired WNG:

$$\mathcal{C}_4\left[\mathbf{h}_{\text{tot}}\left(\omega,\theta\right)\right]:\mathbf{h}_{\text{tot}}^H\left(\omega_q,\theta_p\right)\mathbf{h}_{\text{tot}}\left(\omega_q,\theta_p\right)\leq\frac{1}{\delta}$$
$$\forall p\in\left[1,P\right],\quad\forall q\in\left[1,Q\right]$$

- To guarantee beamformer use of selected placements:

$$\mathcal{C}_5\left[\mathbf{s},\mathbf{h}_{\text{tot}}\left(\omega,\theta\right)\right]:\left|H_{\text{tot},i}\left(\omega_q,\theta_p\right)\right|^2\leq\frac{S_i}{\delta}$$
$$\forall i\in\left[1,N\right],\quad\forall p\in\left[1,P\right],\quad\forall q\in\left[1,Q\right]$$

TECHNION Israel Institute of Technology

The Andrew and Erna Viterbi Faculty of ELECTRICAL & COMPUTER ENGINEERING

Formalizing a Solvable Problem - Target

- When the distortionless constraint is met, the directivity index is determined only by the denominator. Therefore, to maximize the worst-case directivity index we should minimize:

$$R\left[\mathbf{h}_{\text{tot}}(\omega,\theta)\right] = \max_{p\in[1,P]} \sum_{q=1}^{Q} \mathbf{h}_{\text{tot}}^{H}(\omega_q,\theta_p)\,\boldsymbol{\Gamma}_{\text{tot}}(\omega_q)\,\mathbf{h}_{\text{tot}}(\omega_q,\theta_p)$$

TECHNION Israel Institute of Technology

The Andrew and Erna Viterbi Faculty of ELECTRICAL & COMPUTER ENGINEERING

Optimal Array Design

- The optimal design is found by solving a mixed-integer convex optimization problem.
- The non-zero elements of the optimal binary vector $\boldsymbol{s}^*$ yield the optimal microphone locations $\boldsymbol{x}^*$. The non-zero elements of the optimal coefficients $\boldsymbol{h}^*_{tot}(\omega,\theta)$ yield the optimal coefficients $\boldsymbol{h}^*(\boldsymbol{x}^*,\omega,\theta)$.

$$\begin{aligned} \min_{\mathbf{s},\mathbf{h}_{\mathrm{tot}}(\omega,\theta)} \quad & R\left[\mathbf{h}_{\mathrm{tot}}(\omega,\theta)\right] \\ \text{s.t.} \quad & \mathcal{C}_1[\mathbf{s}], \mathcal{C}_2[\mathbf{s}], \mathcal{C}_3[\mathbf{h}_{\mathrm{tot}}(\omega,\theta)], \\ & \mathcal{C}_4[\mathbf{h}_{\mathrm{tot}}(\omega,\theta)], \mathcal{C}_5[\mathbf{s}, \mathbf{h}_{\mathrm{tot}}(\omega,\theta)] \end{aligned}$$

Coefficient Post-Processing

- Since the worst-case look direction is considered, beamformers directed toward other directions may not yield the best possible directivity. To circumvent this, we introduce a post-processing scheme.
- To maximize directivity while maintaining sufficient WNG, the coefficients are found by the robust superdirective beamformer:

$$\mathbf{h}_\epsilon\left(\mathbf{x}^*,\omega,\theta\right)=\frac{\mathbf{\Gamma}_\epsilon^{-1}\left(\mathbf{x}^*,\omega\right)\mathbf{d}\left(\mathbf{x}^*,\omega,\theta\right)}{\mathbf{d}^H\left(\mathbf{x}^*,\omega,\theta\right)\mathbf{\Gamma}_\epsilon^{-1}\left(\mathbf{x}^*,\omega\right)\mathbf{d}\left(\mathbf{x}^*,\omega,\theta\right)}$$

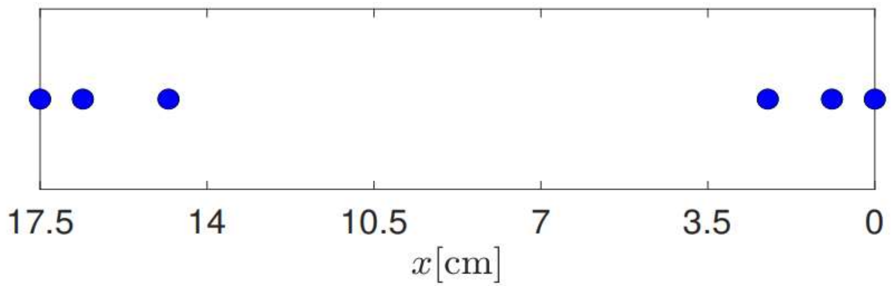
- For every frequency and look direction, ϵ is found by a bisection search:

$$0\leq\epsilon\leq\frac{\lambda_1-\sqrt{M/\delta}\lambda_M}{\sqrt{M/\delta}-1}$$

where $\mathbf{\Gamma}\left(\mathbf{x}^*,\omega\right)=\mathbf{Q}\left(\mathbf{x}^*,\omega\right)\mathbf{\Lambda}\left(\mathbf{x}^*,\omega\right)\mathbf{Q}^T\left(\mathbf{x}^*,\omega\right)$ is the eigenvalue decomposition such that that $\mathbf{\Lambda}=\operatorname{diag}\left[\lambda_1,\lambda_2,...,\lambda_M\right]$, $\lambda_1\geq\lambda_2\geq...\geq\lambda_M$.

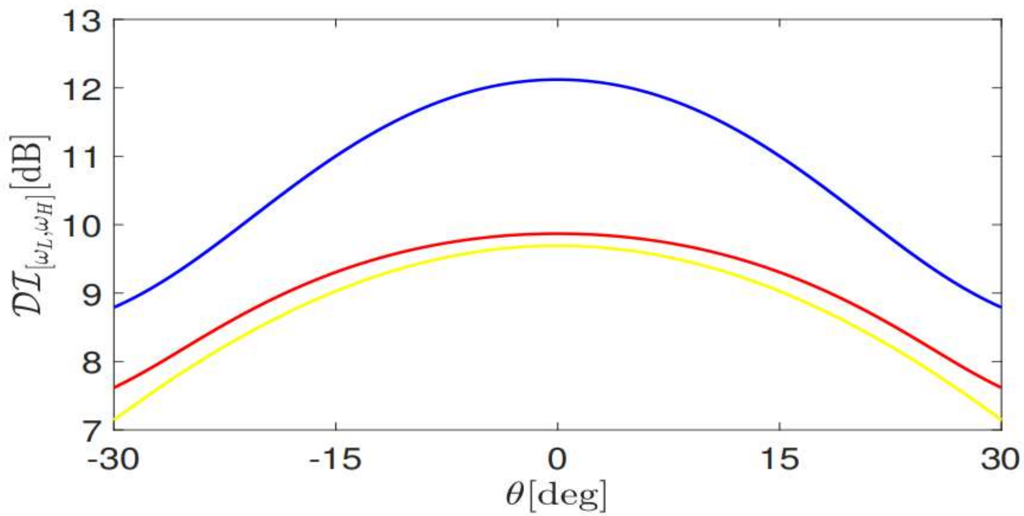
Results - Geometry

- $M = 6$ microphones, aperture length $A = 17.5[cm]$, minimal distance $d_c = 0.5[cm]$. Frequencies from $f_L = 2[kHz]$ to $f_H = 6[kHz]$, look directions up to $\theta_H = 30°$. Minimum WNG is $\delta = -10[dB]$. $N = 40$, $Q = 15$, $P = 15$.



- Some sensors close together yield high directivity, like DMAs.
- Some sensors further apart, due to WNG constraint.

Results – Directivity Index



- Blue – proposed. Red – ULA. Yellow – dense geometry.
- Better directivity in the worst-case direction. 👍
- Better directivity in **all** ROI directions. 👍

TECHNION Israel Institute of Technology

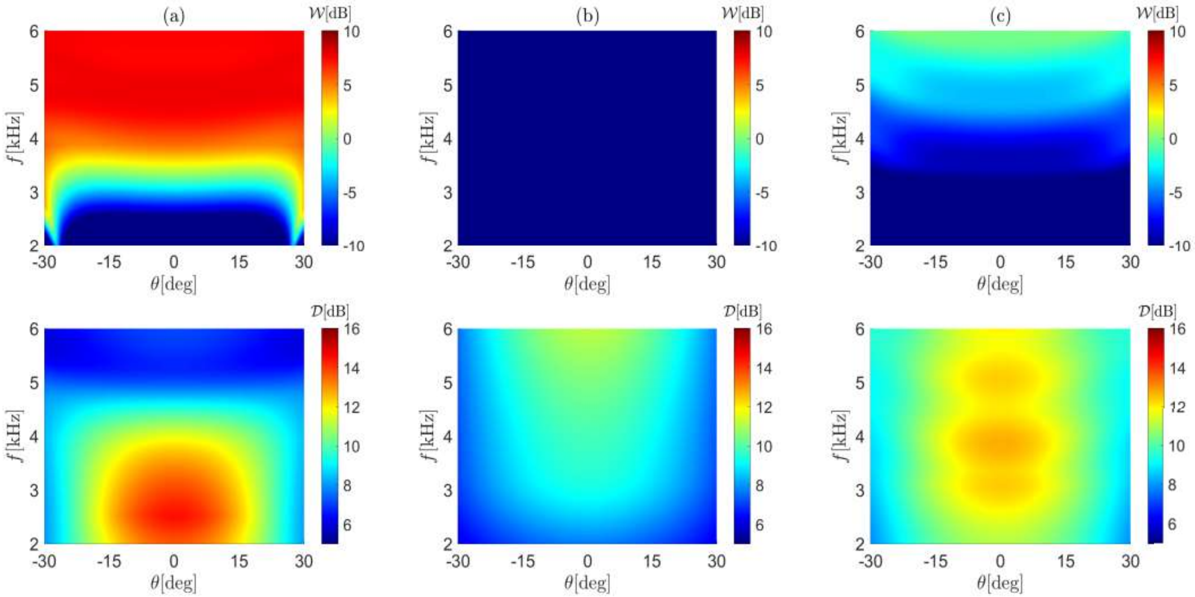
The Andrew and Erna Viterbi Faculty of ELECTRICAL & COMPUTER ENGINEERING

Results – WNG and Directivity Factor

- (a) ULA – Excellent WNG, but spatial aliasing damages DF at higher frequencies.
- (b) Dense geometry – Barely sufficient WNG, cannot produce high directivity due to WNG constraint.
- (c) Proposed – Good WNG, good broadband directivity. 👍

WNG:

DF:



Conclusions

Technion Israel Institute of Technology

The Andrew and Erna Viterbi Faculty of ELECTRICAL & COMPUTER ENGINEERING

Research Contributions

- A novel beamforming-based multichannel AEC method was proposed.
  - Capable of operating in dynamic environments.
  - Does not require double-talk detection.
  - Robust to extremely low SER.
  - Achieves higher ERLE and lower DI compared to an existing method.
- The array geometry for broadband ROI was optimized.
  - A convex framework was used, enabling the convergence to the global optimum.
  - A broadband frequency range was considered.
  - Sensors were placed close closer / further apart , to achieve compromise between directivity and noise robustness.

TECHNION
Israel Institute
of Technology

The Andrew and Erna Viterbi Faculty of
ELECTRICAL & COMPUTER
ENGINEERING

Future Research

- Modelling nonlinear loudspeaker distortion and background noise. The LMS approach works well only if those components are relatively small.
- Developing more accurate inter-frame and inter-sensor relations. For example, instead of the MTF approximation (*) using the Cross-Multiplicative Transfer Function (CMTF) approximation (**).
- Exploring other array types. The search is limited to non-uniform linear arrays. This will enable more complex ROIs too.
- Geometry optimization by other approaches, such as learning-based methods or deep neural networks (DNNs). Our framework runtime grows considerably with more optimization variables.

(*)Y. Avargel and I. Cohen, “On multiplicative transfer function approximation in the short-time fourier transform domain,” *IEEE Signal Processing Letters*, vol. 14, pp. 337–340, 2007.

(**) Y. Avargel and I. Cohen, “Adaptive system identification in the short-time fourier transform domain using cross-multiplicative transfer function approximation,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, pp. 162– 173, 2008.

Questions?

Thank you!

Special thanks to my supervisors Prof. Israel Cohen and Dr. Baruch Berdugo