Rotation-robust beamforming based on sound field interpolation with regularly circular microphone array

Yukoh Wakabayashi, Kouei Yamaoka, and Nobutaka Ono

Department of Computer science,
Graduate school of systems design,
Tokyo metropolitan university

Tokyo, Japan

IEEE ICASSP 2021

Organization

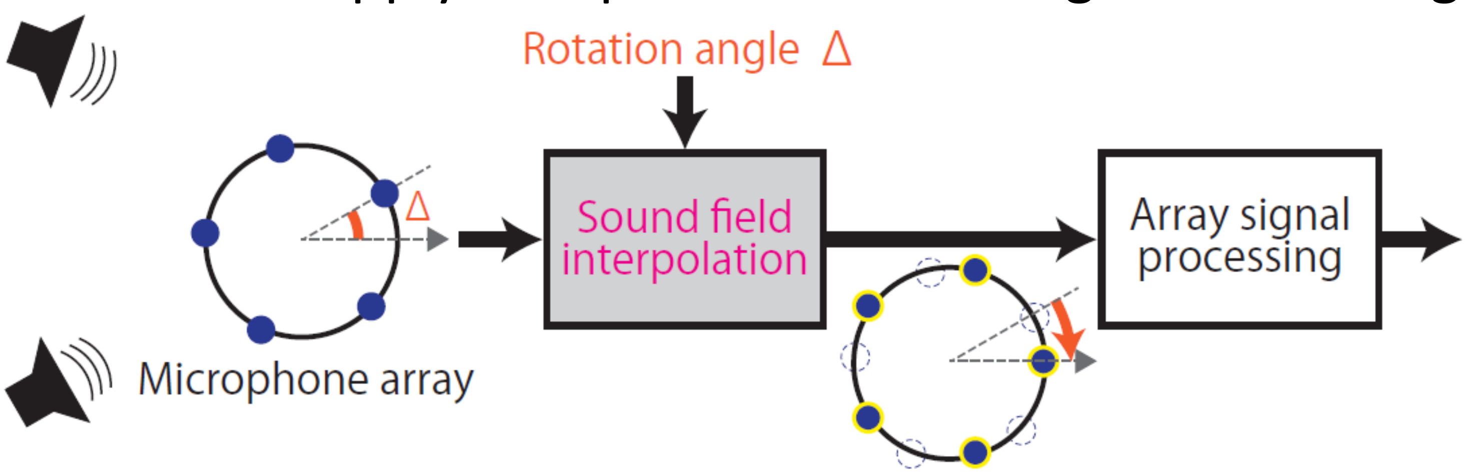
1. Introduction
2. Prior work
3. Sound field interpolation
   - Idea
   - Formulation
   - Remark
4. Application to beamforming
5. Evaluation
   - Interpolation performance
   - Signal enhancement performance by rot.-robust beamforming
6. Conclusion

Introduction

- Background
  - Array signal processing assumes a time-invariant ATS*.
  - ATS's variation makes re-estimation of spatial information (e.g., spatial filter and DOA†) necessary.
    ⇒makes online processing difficult.

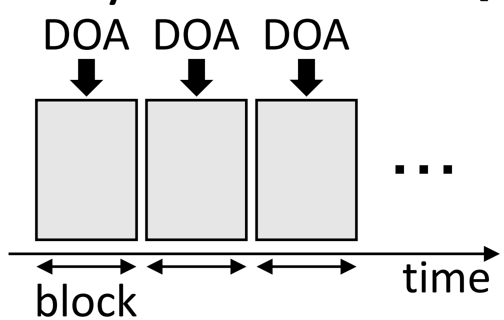
*: acoustic transfer system
†: direction of arrival
**: circular microphone array

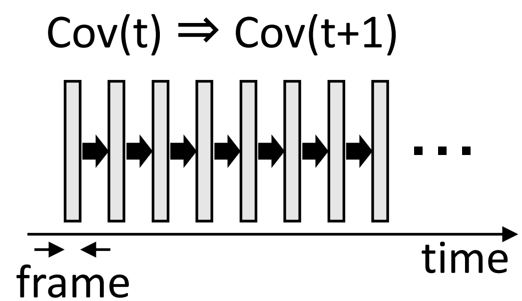
- Motivation
  - We want to follow the ATS's variation caused by CMA** rotation.
  - We want to apply interpolation to existing beamformings.



Prior work for ATS's variation

- Case of source movement
  - Blockwise spatial filter estimation using DOA information [Nikunen+, 2018], [Naqvi+, 2011]
  - Sequential covariance estimation every time-frequency bins with a Bayesian tracker [Taseska+, 2018]





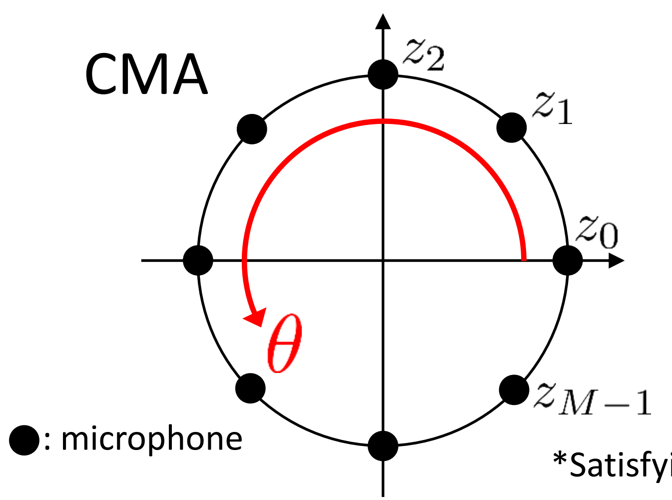
- Case of sensor movement
  - Motion compensation method [Tourbabin+, 2015]
    - Formulating circular harmonic domain rotation matrix
    - Applying to DOA estimation using a CMA

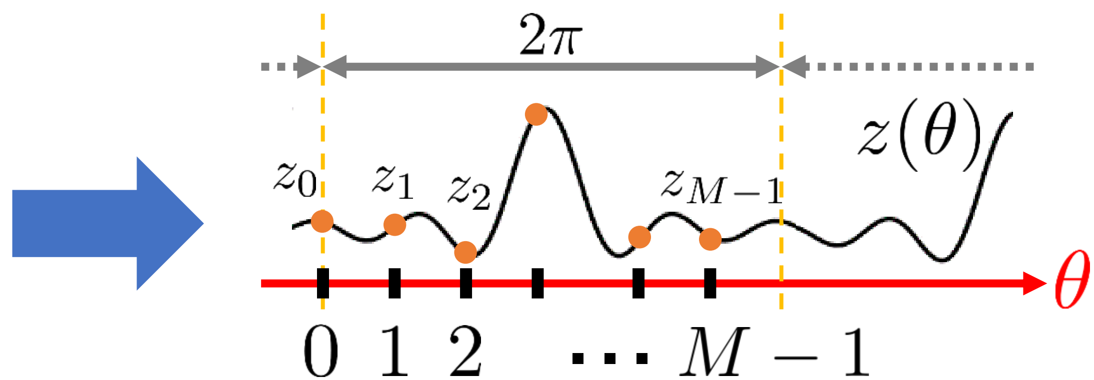
Idea of sound field interpolation

- Continuous sound field on circle's circumference, $z(\theta)$
  - $z(\theta)$ is a periodic function with 2π
  - Discretizing $z(\theta)$ with $2\pi/M$ intervals
    = Observing the sound field using an M-ch CMA

$$z_m = z\left(2\pi\frac{m}{M}\right), \quad m = 0, \ldots, M-1.$$

- $z(\theta)$ can be reconstructed by the discrete signal z_m if the sampling theorem is satisfied*, resulting in **sound field interpolation** possible.





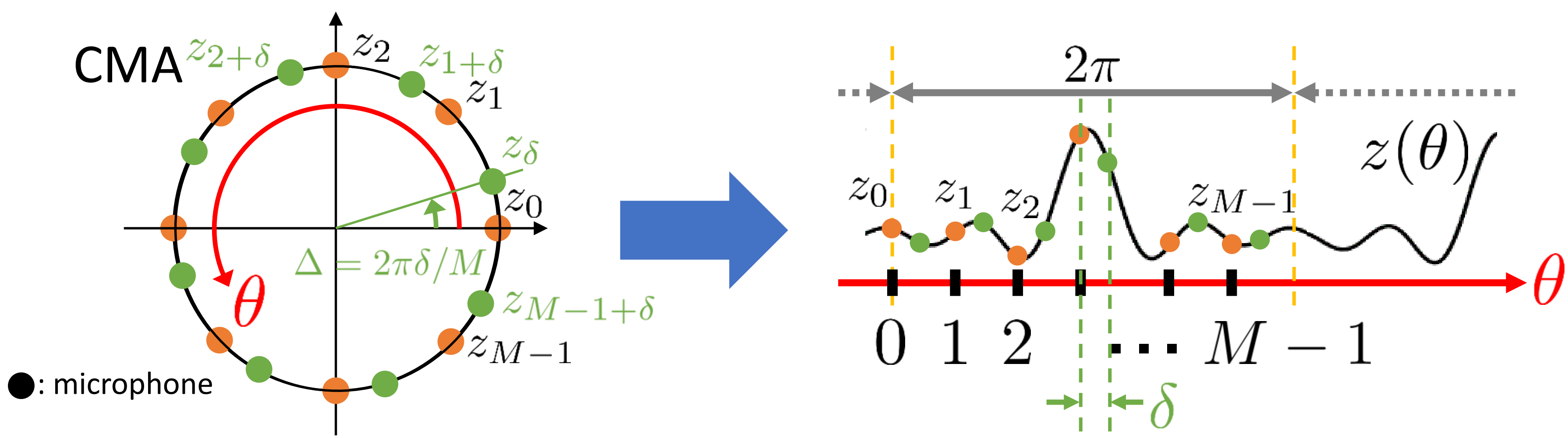
*Satisfying the sampling theorem = Band limitation of circular harmonics spectrum

Sound field interpolation using noninteger sample shift

- Relationship b/w CMA rotation and sample shift
  - Sound field observed by a CMA in the reference position, z_m
  - Sound field observed by the CMA rotated Δ (=$2\pi\delta/M$) [rad]
    = δ-sample shifted z_m in the spatial axis

$$z_{m+\delta} = z\left(2\pi\frac{m}{M} + \Delta\right)$$

- The above equation enables estimating $z_{m+\delta}$ from z_m, m=0,…, M-1.



Formulation of linear interpolation

- Linear representation of δ-sample shifted sound field using the sample shift theorem in the DFT

$$z_{m+\delta} = \mathscr{F}_{\mathrm{D}}^{-1}\left[\mathscr{F}_{\mathrm{D}}\left[z_m\right]e^{j\Delta k}\right]$$

$$= \frac{1}{M}\sum_{k=-M/2+1}^{M/2}\left(Z_k e^{j\Delta k}\right)e^{-j2\pi\frac{mk}{M}} \stackrel{\mathrm{def}}{=} \sum_{n=0}^{M-1} z_n u_{m,n,\delta}$$

$\mathscr{F}_{\mathrm{D}}$: DFT operation

- Sound field interpolation using sinc function (It is derived from the equation above*.)

*: The detailed derivation is appended on the last page.

$$u_{m,n,\delta} = \begin{cases} \frac{1+(-1)^{n-m}e^{j\delta\pi}}{M} + \frac{\mathrm{sinc}\left(L/2\right)}{\mathrm{sinc}\left(L/M\right)} \cdot \cos\left(\frac{M+2}{2M}L\pi\right), & M \text{ is even}, \\ \frac{1}{M} + \frac{M-1}{M}\frac{\mathrm{sinc}\left(L(M-1)/2M\right)}{\mathrm{sinc}\left(L/M\right)} \cdot \cos\left(\frac{M+1}{2M}L\pi\right), & M \text{ is odd}. \end{cases}$$

$$L = n - m - \delta$$

Remark of formulation

- How to handle Nyquist frequency (NyqF) component

$$z_{m+\delta} = \sum_{n=0}^{M-1} z_n u_{m,n,\delta} \quad (*)$$

$$u_{m,n,\delta} = \begin{cases} \frac{1+(-1)^{n-m}}{M} + \frac{\mathrm{sinc}\left(L/2\right)}{\mathrm{sinc}\left(L/M\right)} \cdot \cos\left(\frac{M+2}{2M}L\pi\right), & M \text{ is even}, \\ \frac{1}{M} + \frac{M-1}{M}\frac{\mathrm{sinc}\left(L(M-1)/2M\right)}{\mathrm{sinc}\left(L/M\right)} \cdot \cos\left(\frac{M+1}{2M}L\pi\right), & M \text{ is odd}. \end{cases}$$

  - When M is even, a noninteger sample shift results in a complex-valued term even if the sign of the NyqF is positive or negative, and which causes some contradiction;
    e.g., sample shift of a real-valued even point signal by $(*)$ translates to a complex-valued signal.
  - In this study, we neglect the NyqF component by setting δ=0.

Matrix representation

- Formulation of the relationship b/w the observation by a Δ-rotated CMA, **x**, and by the CMA w/o rotation, $\mathbf{x}_0$

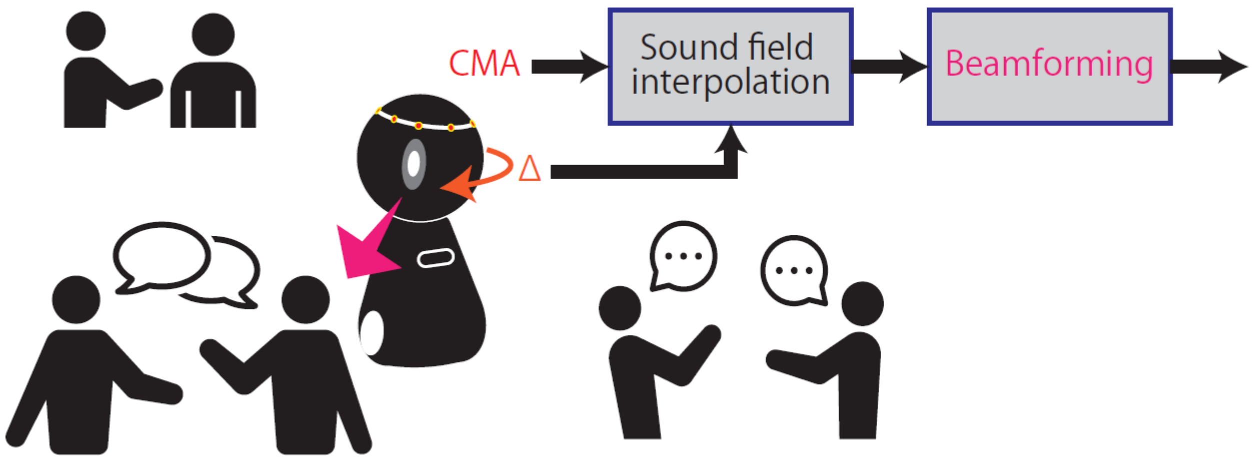
$$
\begin{aligned}
\boldsymbol{x} &= \begin{bmatrix} x_1 & \cdots & x_M \end{bmatrix}^{\mathrm{T}} \\
&= \begin{bmatrix} z\Big(2\pi(0+\delta)/M\Big) & \cdots & z\Big(2\pi(M-1+\delta)/M\Big) \end{bmatrix}^{\mathrm{T}} \\
&= \begin{bmatrix} u_{0,0,\delta} & u_{0,1,\delta} & \cdots & u_{0,M-1,\delta} \\ u_{1,0,\delta} & u_{1,1,\delta} & \cdots & u_{1,M-1,\delta} \\ \vdots & \vdots & \ddots & \vdots \\ u_{M-1,0,\delta} & u_{M-1,1,\delta} & \cdots & u_{M-1,M-1,\delta} \end{bmatrix} \begin{bmatrix} z_0 \\ z_1 \\ \vdots \\ z_{M-1} \end{bmatrix} \\
&= \boldsymbol{U}_{\Delta} \boldsymbol{x}_0 .
\end{aligned}
$$

- Note: $\mathbf{U}_{\Delta}$ is a cyclic matrix & does not depend on frequencies

[cf.] Our method has a tight relationship to the circular harmonics domain rotation matrix used by the motion compensation method [Tourbabin+, 2015].

Applying interpolation to beamforming

- Situation: A user or humanoid robot
  - wears a CMA on the head.
  - rotates the head to listen to ambient conversations attentively.



- Problem:
  - The rotation angle Δ is given.
  - We estimate the observation **before CMA rotation**.
  - We do beamforming using interpolated M-ch signals.

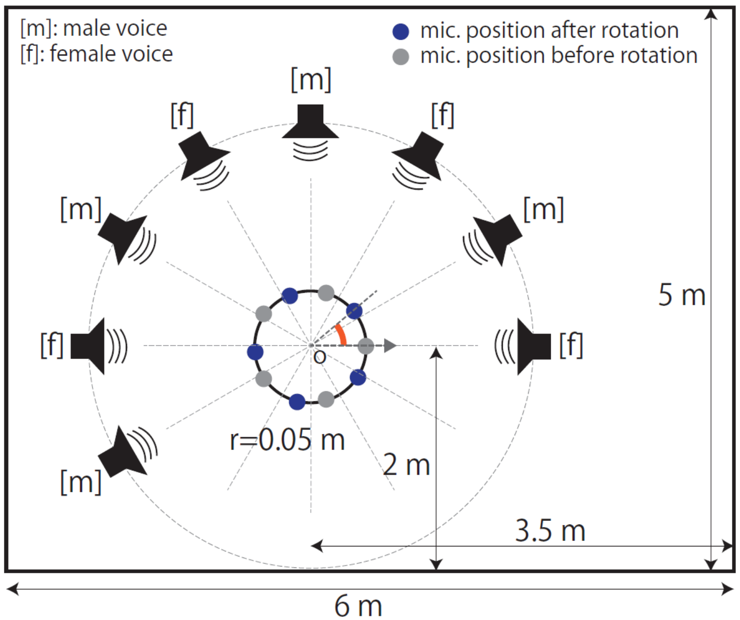
Evaluation condition

- Mixture : 2 sources from SiSEC [Araki+, 2012], fs=16 kHz, 12 envs.
- RIR : Simulator [Habets, 2008], RT60≒100 ms
- STFT : 64 ms Hamming window with 1/8 shifts

- **Exp. 1 : interpolation performance**
  - Array rot. Δ : 10, 20, and 30 deg
  - Measure : SER (signal-to-error ratio)

$$\mathrm{SER}_{m,k} = 10\log_{10}\left(\frac{\sum_t |x_{m,t,k}|^2}{\sum_t |\hat{x}_{m,t,k} - x_{m,t,k}|^2}\right)$$

$x_{m,t,k} \in \mathbb{C}$



- **Exp. 2 : signal enhancement**
  - Array rot. Δ : 10, 20, and 36 deg
  - Beamformer : MPDR* + RTF** [Doclo+, 2015]
    - 2 methods of applying beamformer.
  - Measure : SDR, SIR [Vincent+, 2006]

*minimum power distortionless response
**relative transfer function

How to apply to beamforming

- **Method 1: Constantly use of pre-estimated spatial filter**
  - Pre-estimation of spatial filter when the CMA does not rotate

$$\boldsymbol{w} = \mathrm{MPDR}(\boldsymbol{V}, \boldsymbol{a}), \quad \boldsymbol{V} = \mathbb{E}\left[\boldsymbol{x}_0 \boldsymbol{x}_0^{\mathrm{H}}\right]$$

  - Interpolation & beamforming

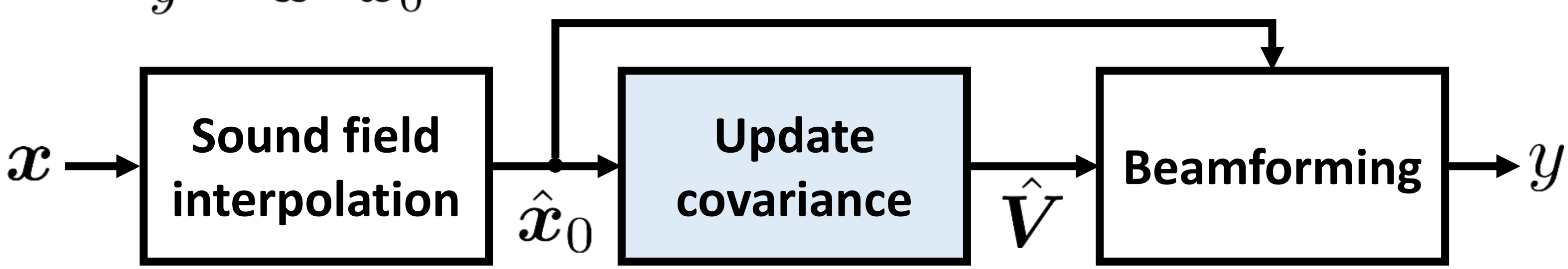
$$y = \boldsymbol{w}^{\mathrm{H}} \hat{\boldsymbol{x}}_0 = \boldsymbol{w}^{\mathrm{H}} \left(\boldsymbol{U}_{-\Delta} \boldsymbol{x}\right)$$

Δ : rotation angle
$\boldsymbol{a}$: transfer function
$\boldsymbol{x}_0$: observation in the reference position

- **Method 2: Re-estimation of spatial filter**

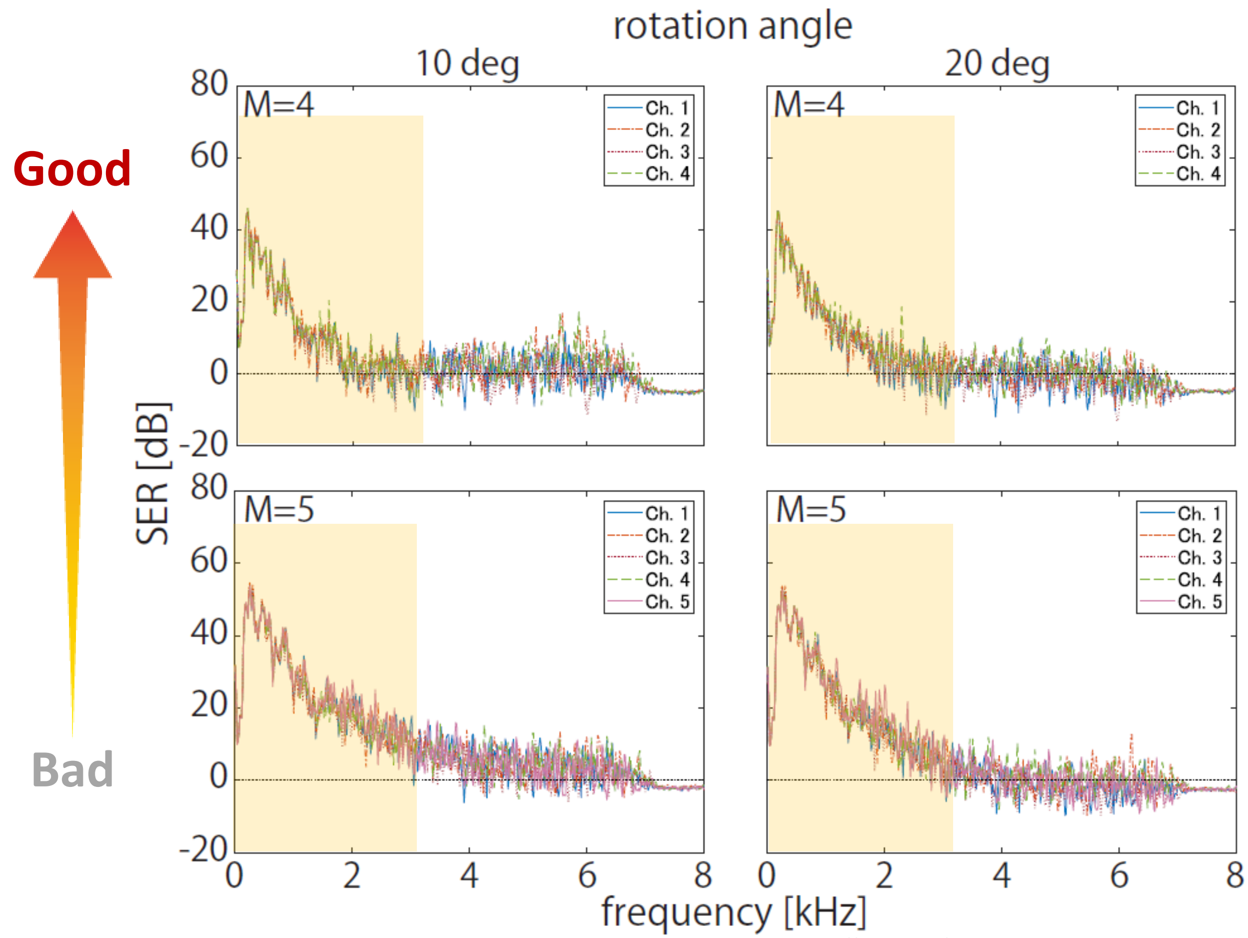
$$\tilde{\boldsymbol{w}} = \mathrm{MPDR}(\hat{\boldsymbol{V}}, \boldsymbol{a}), \quad \hat{\boldsymbol{V}} = \mathbb{E}\left[\hat{\boldsymbol{x}}_0 \hat{\boldsymbol{x}}_0^{\mathrm{H}}\right], \quad \hat{\boldsymbol{x}}_0 = \boldsymbol{U}_{-\Delta} \boldsymbol{x}$$

$$y = \tilde{\boldsymbol{w}}^{\mathrm{H}} \hat{\boldsymbol{x}}_0$$



Exp. 1: Interpolation performance 1/2

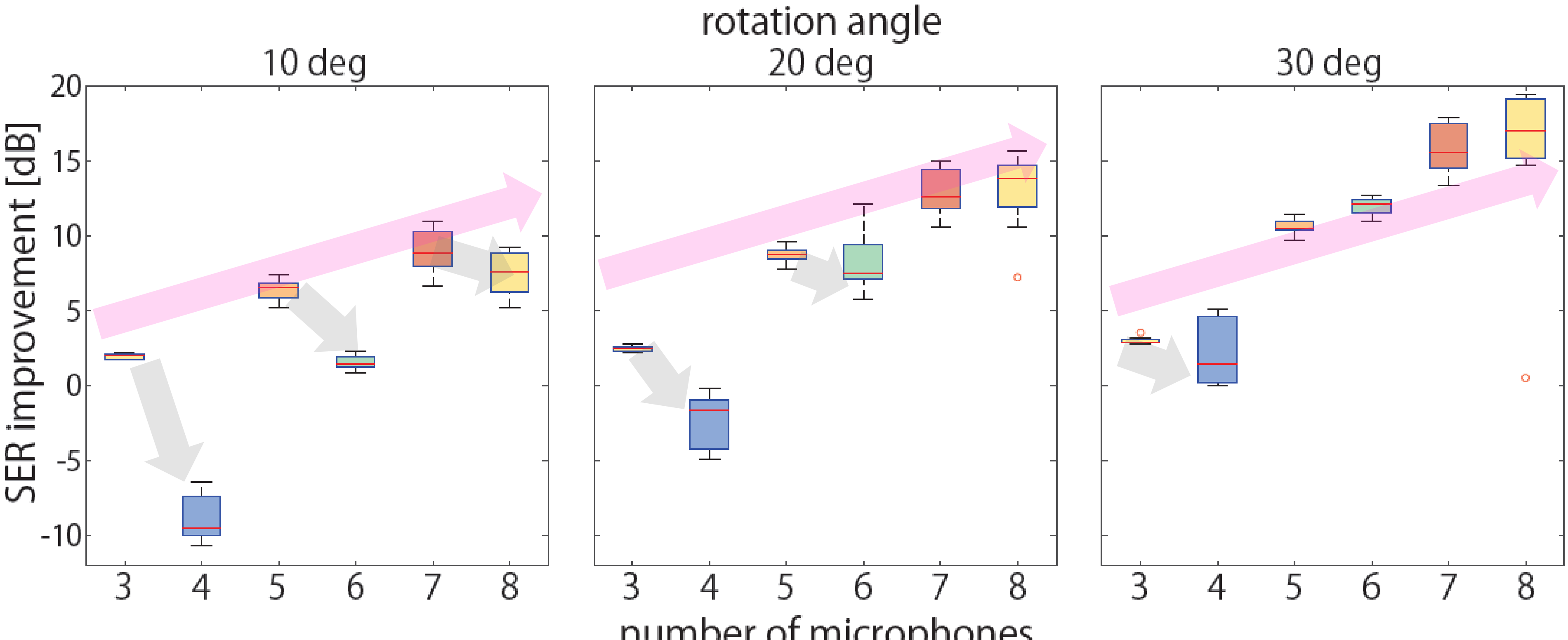
- Examples of SER when 2 sources are active
  - Lower band interpolation worked well, but not higher one.



Exp. 1: Interpolation performance 2/2

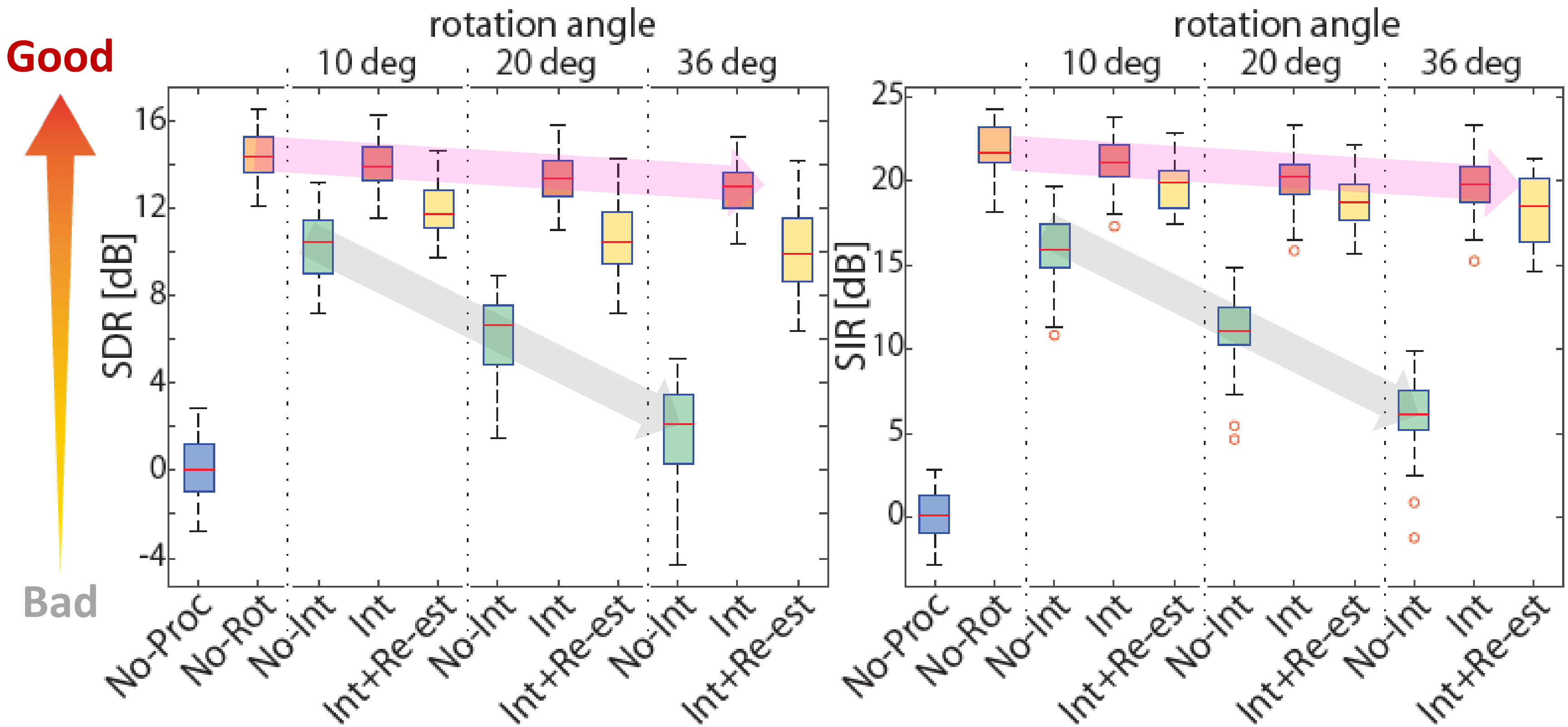
- Averaged SER improvement (up to 3 kHz)
  - More the number of microphones improves performance.↑
  - NyqF components decrease performance when M=4, 6↓

    e.g., when M=4, only 3 components contribute to interpolation.



Exp. 2: Signal enhancement

- SDR & SIR, M=5
  - In **No-Int** case, bigger rotation decreases SDR & SIR.
  - **Int** & **Int+Re-est** come close to the performance of **No-Rot**.



No-Proc : unprocessed, **No-Rot : no rotation**, **No-Int : no interpolation**, **Int : interpolation**, **Int＋Re-est : interpolation + filter re-estimation**

Conclusion

- Summary
  - Sound field interpolation method for rotation-robust beamforming using CMA.
  - Lower band interpolation accuracy was high when M is odd.
  - Higher band interpolation accuracy was low, but applying to beamforming worked well.
- Future work
  - Applying to different array processing, e.g., source separation
  - Clarifying relation to circular harmonics domain beamforming

Derivation (1/2)

When M is even

$$W = e^{-j2\pi/M}$$
$$0 < \delta < 1$$

$$z_{m+\delta} = z\left(2\pi\frac{m}{M} + \Delta\right) = \mathscr{F}_{\mathrm{D}}^{-1}\left[\mathscr{F}_{\mathrm{D}}\left[z_m\right]e^{j\Delta k}\right]$$

$$= \frac{1}{M}\sum_{k=-M/2+1}^{M/2} Z_k e^{j\Delta k} W^{-mk}$$

DC, Positive freq., NyqF component, Negative freq.

$$= \frac{1}{M}\left\{Z_0 + \sum_{k=1}^{M/2-1} Z_k e^{j\Delta k} W^{-mk} + Z_{\frac{M}{2}} e^{j\Delta\frac{M}{2}} W^{-m\frac{M}{2}} + \sum_{k=-M/2+1}^{-1} Z_k e^{j\Delta k} W^{-mk}\right\}$$

$$= \frac{1}{M}\left\{\left(\sum_{n=0}^{M-1} z_n\right) + \sum_{k=1}^{M/2-1}\left(\sum_{n=0}^{M-1} z_n W^{nk} e^{j\Delta k}\right)W^{-mk} + \left(\sum_{n=0}^{M-1} z_n W^{n\frac{M}{2}} e^{j\Delta\frac{M}{2}}\right)W^{-m\frac{M}{2}}\right.$$
$$\left. + \sum_{k=1}^{M/2-1}\overline{\left(\sum_{n=0}^{M-1} z_n W^{nk} e^{j\Delta k}\right)}W^{mk}\right\}$$

This term is a complex value because δ is a noninteger.

$$= \frac{1}{M}\sum_{n=0}^{M-1} z_n \underbrace{\left\{1 + \left(\sum_{k=1}^{M/2-1} W^{(n-m-\delta)k}\right) + W^{(n-m-\delta)\frac{M}{2}} + \left(\sum_{k=1}^{M/2-1} W^{-(n-m-\delta)k}\right)\right\}}_{u_{m,n,\delta}}$$

Derivation (2/2)

Set $L = n - m - \delta$

$$
\begin{aligned}
u_{m,n,\delta} &= \frac{1}{M}\left\{1 + \underline{W^{L\frac{M}{2}}} + \sum_{k=1}^{M/2-1} W^{Lk} + \sum_{k=1}^{M/2-1} W^{-Lk}\right\} \\
&= \frac{1}{M}\left\{1 + W^{L\frac{M}{2}} + W^{L}\frac{1-W^{L\frac{M}{2}}}{1-W^{L}} + W^{-L}\frac{1-W^{-L\frac{M}{2}}}{1-W^{-L}}\right\} \\
&= \frac{1}{M}\left\{1 + W^{L\frac{M}{2}} + W^{\frac{L}{4}(M+2)}\frac{W^{-L\frac{M}{4}}-W^{L\frac{M}{4}}}{W^{-\frac{L}{2}}-W^{\frac{L}{2}}} + W^{-\frac{L}{4}(M+2)}\frac{W^{L\frac{M}{4}}-W^{-L\frac{M}{4}}}{W^{\frac{L}{2}}-W^{-\frac{L}{2}}}\right\} \\
&= \frac{1}{M}\left(1 + W^{L\frac{M}{2}} + W^{\frac{L}{4}(M+2)}\frac{\sin(\pi L/2)}{\sin(\pi L/M)} + W^{-\frac{L}{4}(M+2)}\frac{\sin(\pi L/2)}{\sin(\pi L/M)}\right) \\
&= \frac{1}{M}\left(1 + W^{L\frac{M}{2}}\right) + \frac{1}{2}\cdot\frac{\operatorname{sinc}(L/2)}{\operatorname{sinc}(L/M)}\left(W^{\frac{L}{4}(M+2)} + W^{-\frac{L}{4}(M+2)}\right) \\
&= \frac{1}{M}\left(1 + e^{-jL\pi}\right) + \frac{\operatorname{sinc}\Big(L/2\Big)}{\operatorname{sinc}\Big(L/M\Big)}\cdot\cos\left(\frac{M+2}{2M}L\pi\right).
\end{aligned}
$$