Multi-channel Source Separation 1

안인규 (Inkyu An)

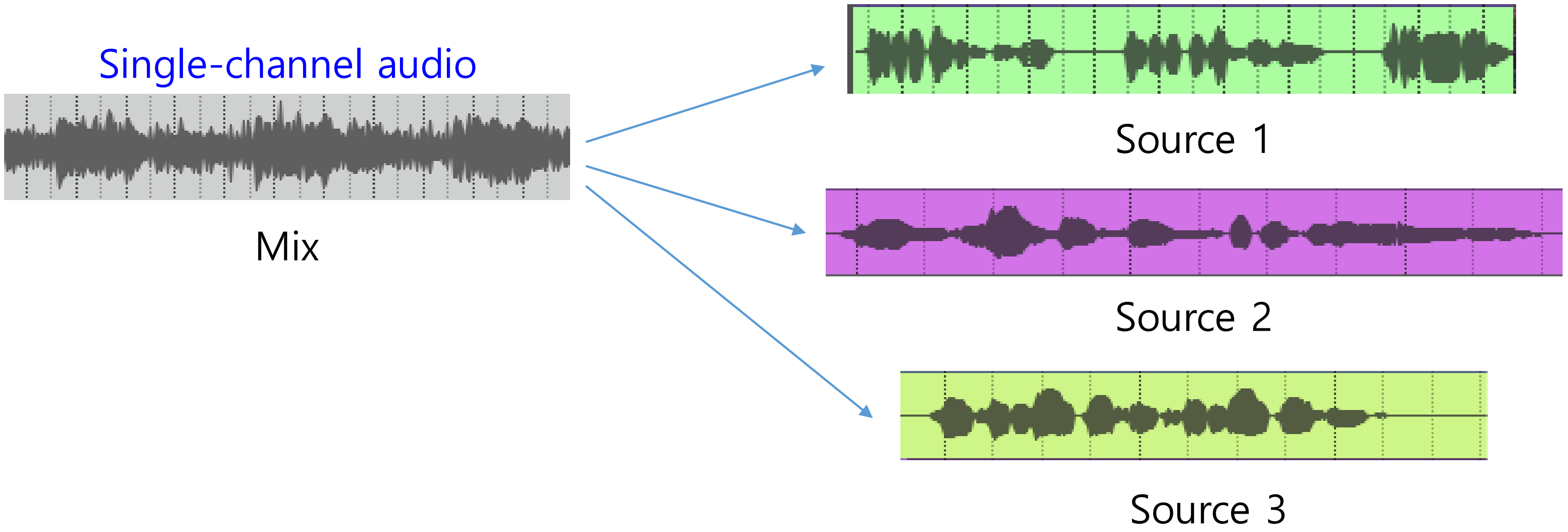
Speech And Audio Recognition
(오디오 음성인식)

https://mairlab-km.github.io/

This lecture material refers to
https://github.com/yandexdataschool/speech_course?tab=readme-ov-file and
https://github.com/markovka17/dla

Blind Source Separation

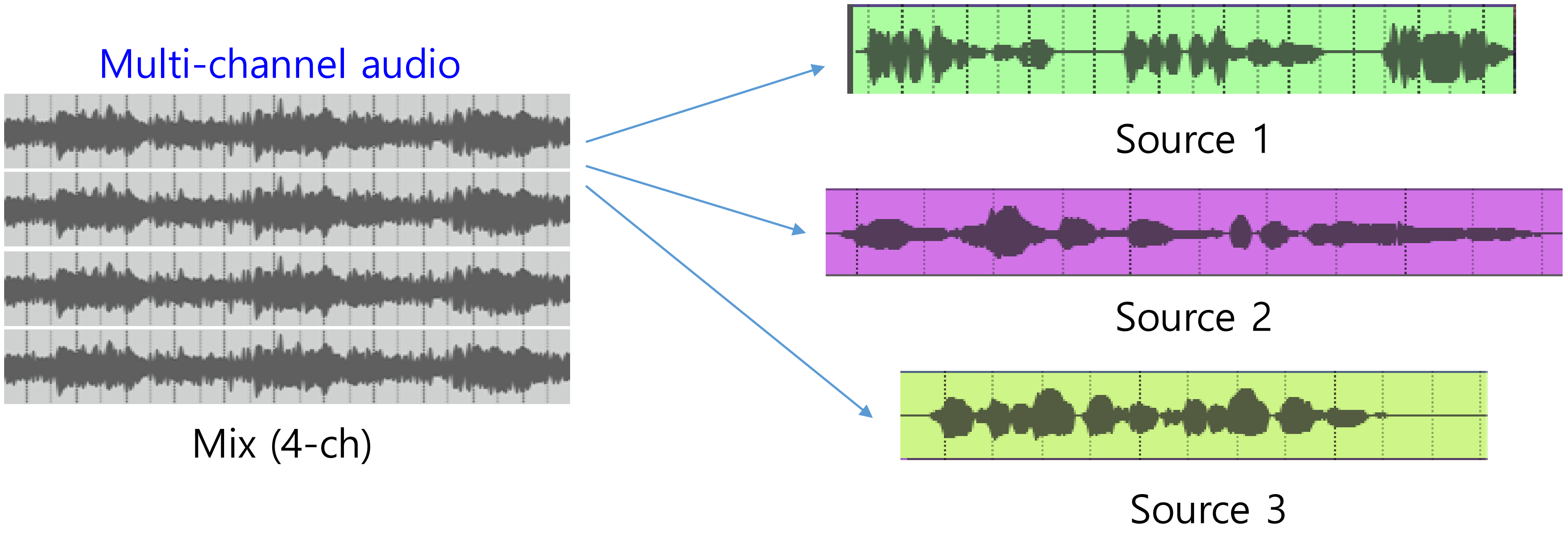
- **Goal**: extract K sources from the noisy mixture **<u>w/o (or with very little) information about the mixing process</u>**



- We know denoising and specific guided separation, how can we apply DL here?

Multi-channel Source Separation

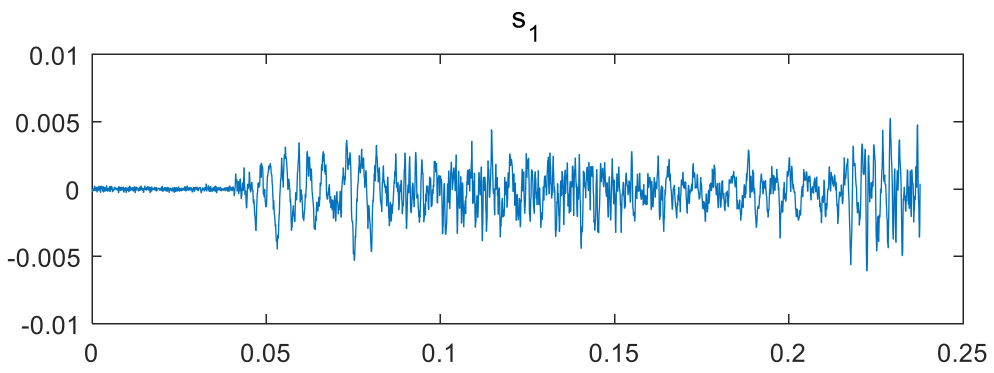
- **Goal**: extract K sources from the noisy mixture of <u>multi-channel audio</u>

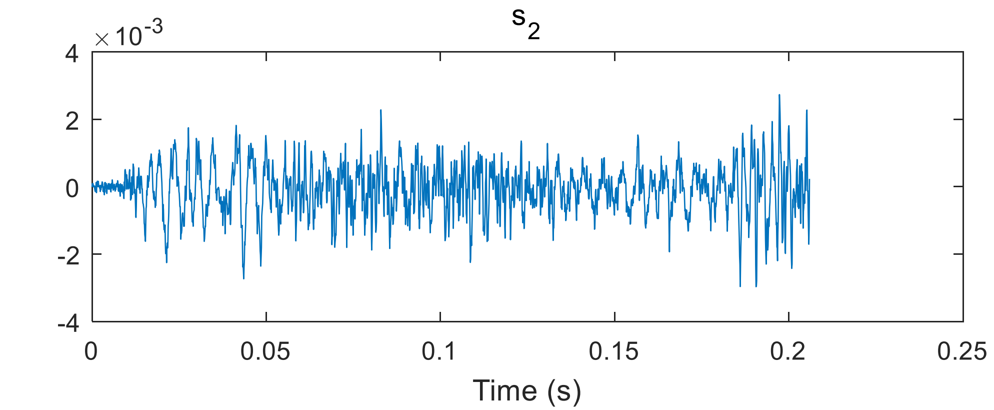


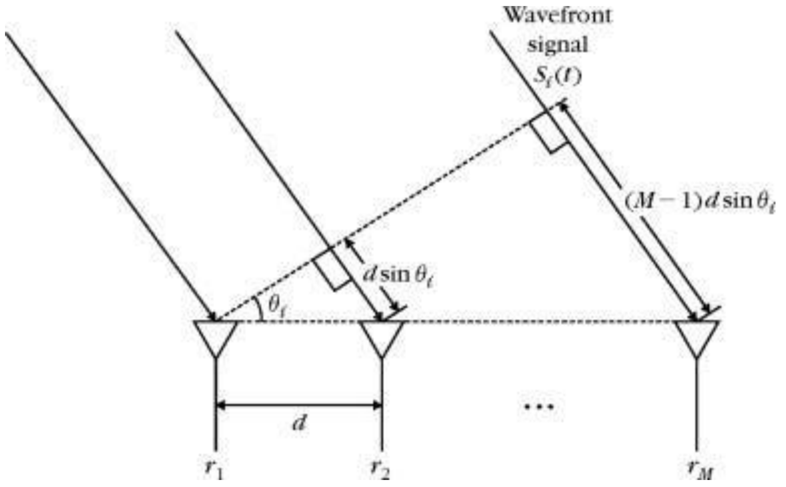
- How can we measure the multi-channel audio signal?

Multi-channel Source Separation

- Microphone Array
  - Consider the Uniform Linear Array
  - What's different about the audio collected at each microphone?

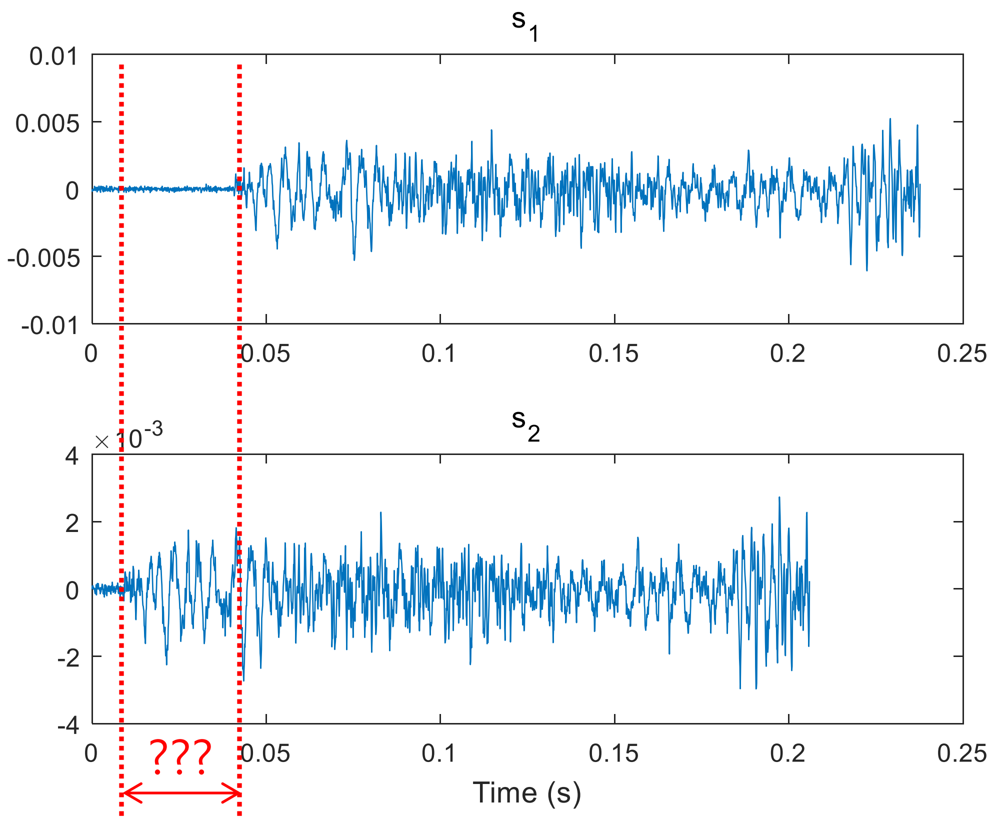


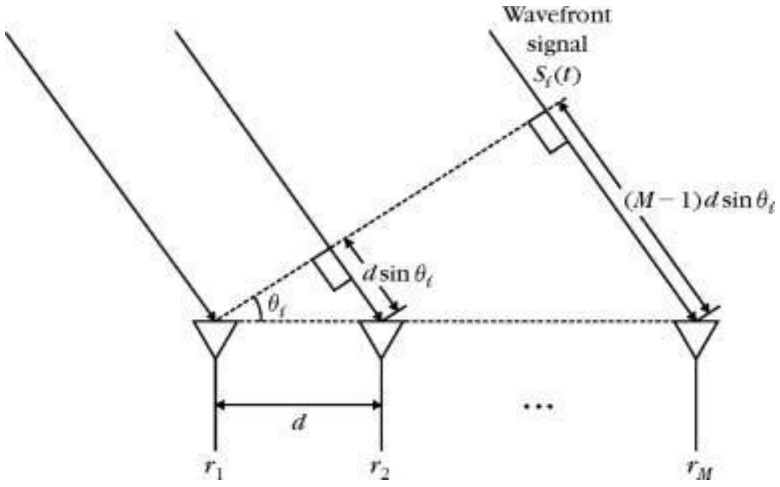




Multi-channel Source Separation

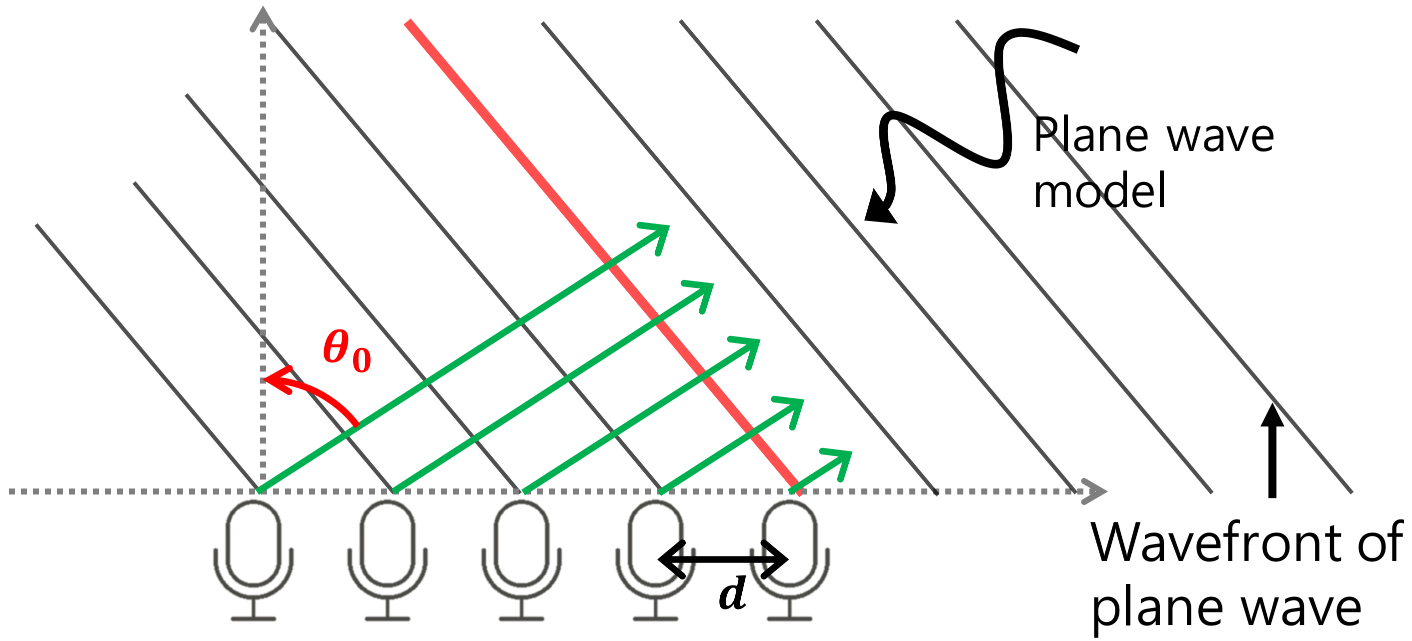
- Microphone Array
  - Consider the Uniform Linear Array
  - What's different about the audio collected at each microphone?





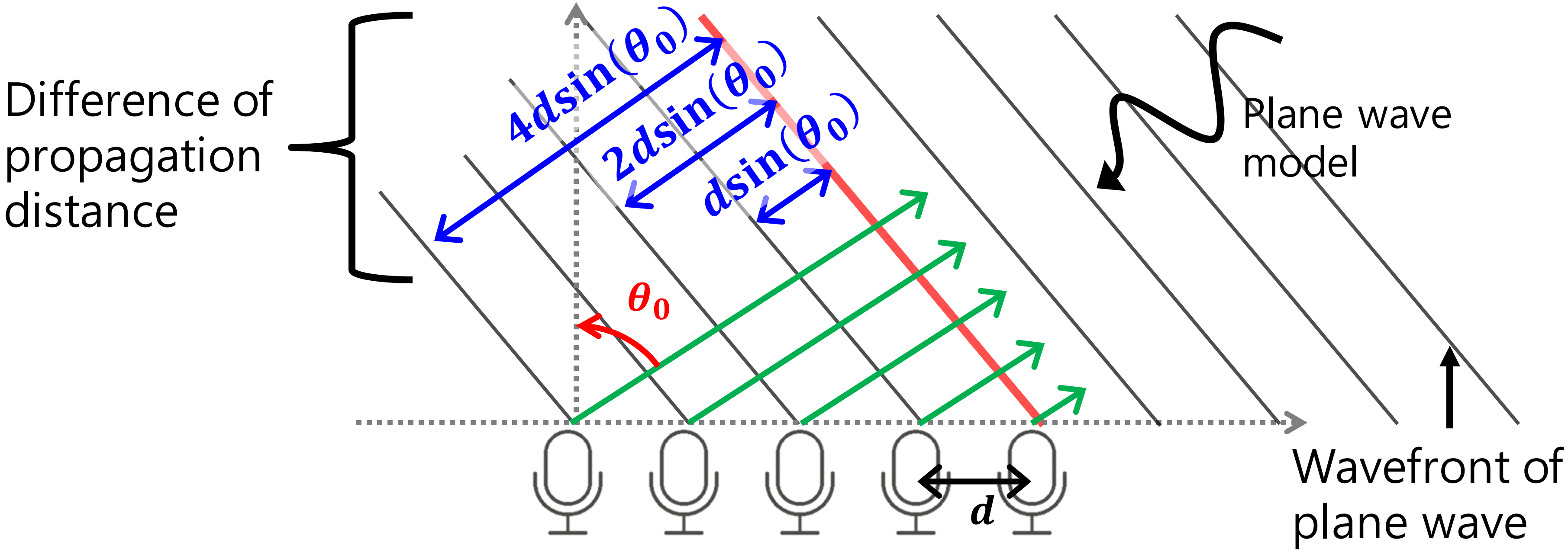
Multi-channel Source Separation

- Microphone Array
  - Consider the Uniform Linear Array
  - What’s different about the audio collected at each microphone?

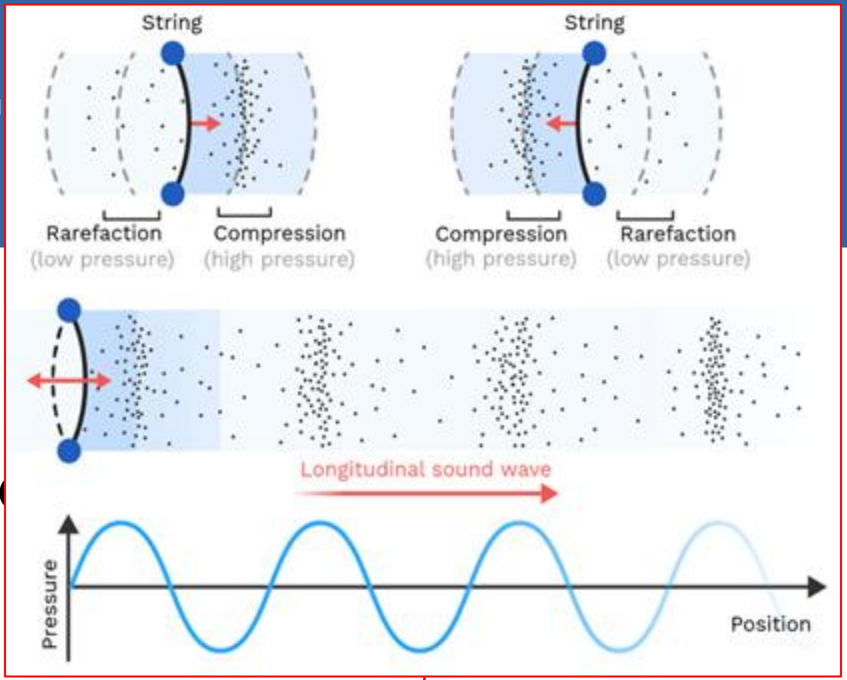


Multi-channel Source Separation

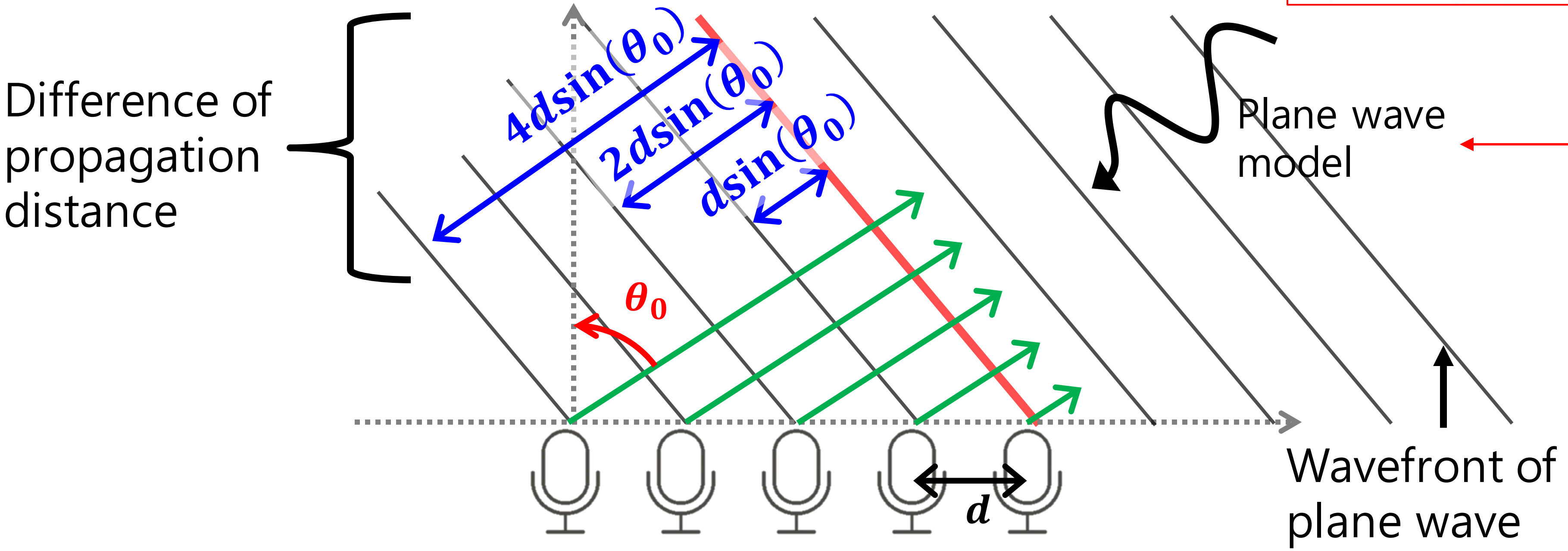
- Microphone Array
  - Consider the Uniform Linear Array
  - What's different about the audio collected at each microphone?



Multi-channel Source Separ



- Microphone Array
  - Consider the Uniform Linear Array
  - What's different about the audio collected at ea



Multi-channel Source Separation

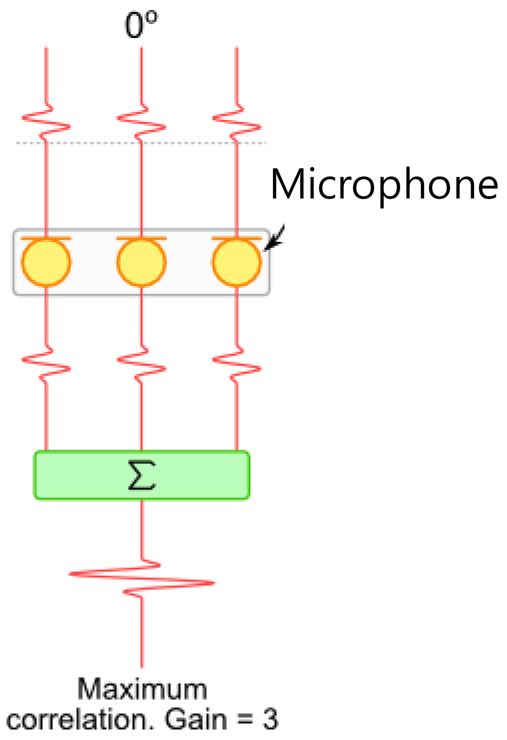
- Microphone Array
  - Is it possible to extract only the audio signal coming from a specific direction? → **Beamforming**
  - What is the beamforming algorithm?
    - Focusing a specific direction, $\boldsymbol{\theta}$, by using beamforming weights (spatial filter)

$$y(\boldsymbol{\theta}, \boldsymbol{t}) = \boldsymbol{w}(\boldsymbol{\theta}, -\boldsymbol{t})^{\boldsymbol{H}} \underset{\text{Convolution}}{*} \boldsymbol{x}(\boldsymbol{t}) \quad \overset{}{\underset{\text{FFT \& iFFT}}{\Longleftrightarrow}} \quad y(\boldsymbol{\theta}, \boldsymbol{\omega}) = \boldsymbol{w}(\boldsymbol{\theta}, \boldsymbol{\omega})^{\boldsymbol{H}}\, \boldsymbol{x}(\boldsymbol{w})$$

Beamforming output | Weight (Spatial filter) | Measured pressure signals

Multi-channel Source Separation

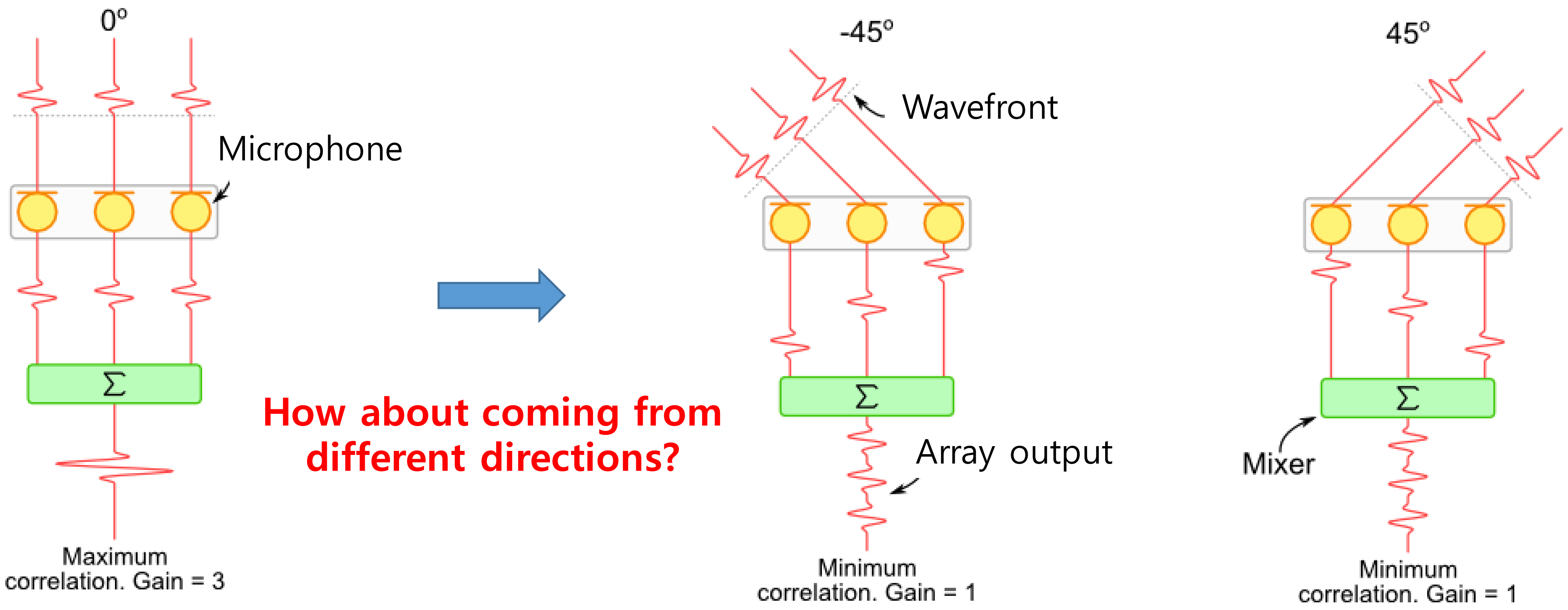
- Microphone Array
  - Is it possible to extract only the audio signal coming from a specific direction? → **Beamforming**
  - A simple method? → **Add all the multi-channel signals together**



https://medium.com/@itberrios6/introduction-to-beamforming-part-1-92767408f483

Multi-channel Source Separation

- Microphone Array
  - Is it possible to extract only the audio signal coming from a specific direction? → **<u>Beamforming</u>**
  - A simple method? → **Add all the multi-channel signals together**



How about coming from different directions?

https://medium.com/@itberrios6/introduction-to-beamforming-part-1-92767408f483

Multi-channel Source Separation

- Microphone Array
  - Is it possible to extract only the audio signal coming from a specific direction? → **Beamforming**
  - What is the beamforming algorithm?
    - Focusing a specific direction, $\boldsymbol{\theta}$, by using beamforming weights (spatial filter)

$$y(\boldsymbol{\theta}, \boldsymbol{t}) = \boldsymbol{w}(\boldsymbol{\theta}, -\boldsymbol{t})^{\boldsymbol{H}} * \boldsymbol{x}(\boldsymbol{t}) \quad \Longleftrightarrow \quad y(\boldsymbol{\theta}, \boldsymbol{\omega}) = \boldsymbol{w}(\boldsymbol{\theta}, \boldsymbol{\omega})^{\boldsymbol{H}}\, \boldsymbol{x}(\boldsymbol{w})$$

Convolution

FFT & iFFT

Beamforming output

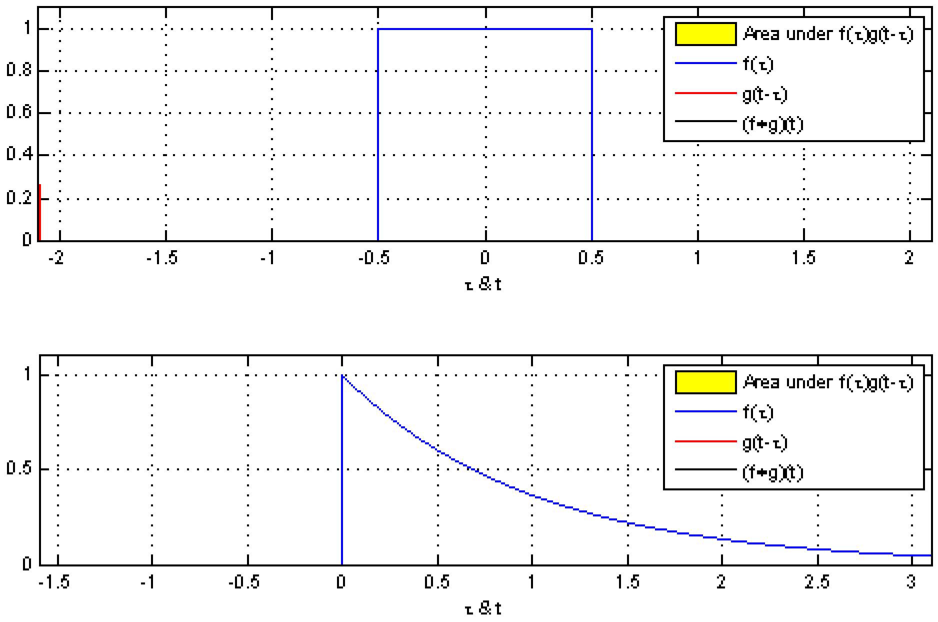
Weight (Spatial filter)

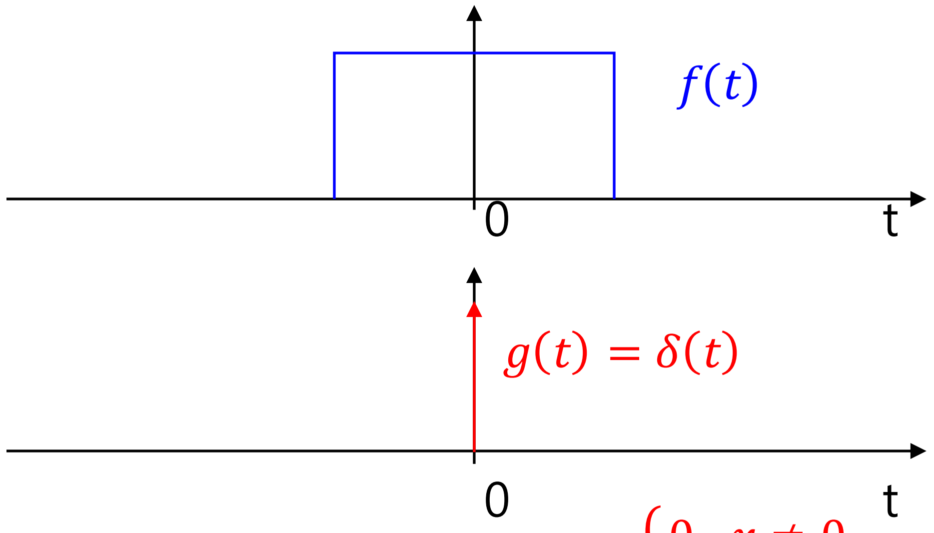
Measured pressure signals

Multi-channel Source Separation

- Microphone Array
  - Convolution: $(f * g)(t) := \int_{-\infty}^{\infty} f(\tau)g(t-\tau)\,d\tau.$

Example of Convolution





Dirac delta function $\delta(t) = \begin{cases} 0, & x \neq 0 \\ \infty, & x = 0 \end{cases}$

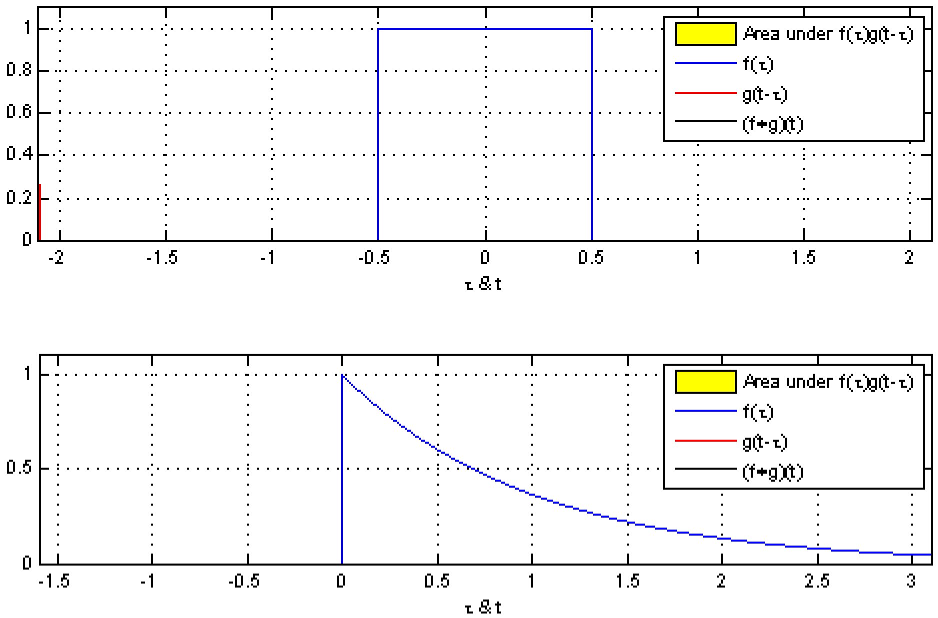
$$\int_{-\infty}^{\infty} \delta(t)dt = 1$$

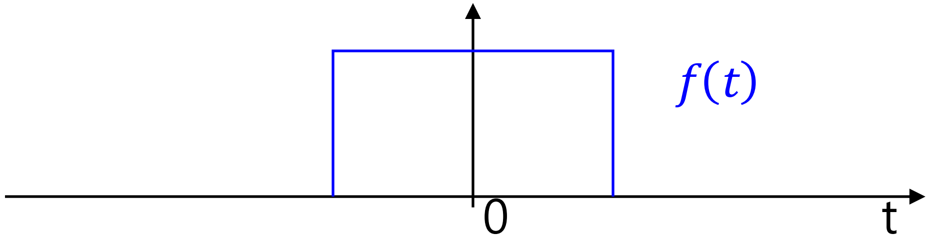
$(f * g)(t)$?

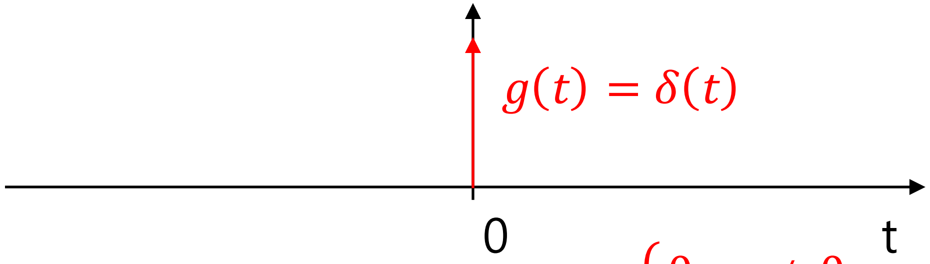
Multi-channel Source Separation

- Microphone Array
  - Convolution: $(f * g)(t) := \int_{-\infty}^{\infty} f(\tau) g(t-\tau)\, d\tau.$

Example of Convolution



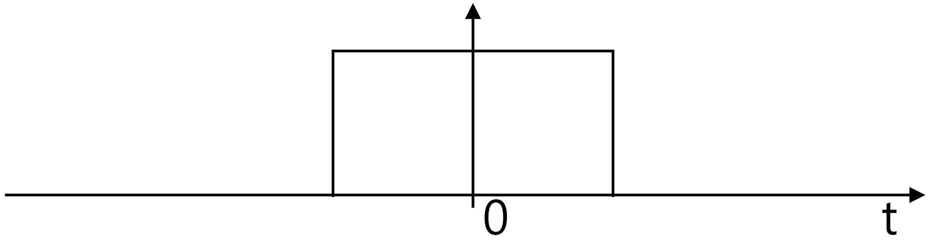




Dirac delta function $\delta(t) = \begin{cases} 0, & x \neq 0 \\ \infty, & x = 0 \end{cases}$

$$\int_{-\infty}^{\infty} \delta(t) dt = 1$$

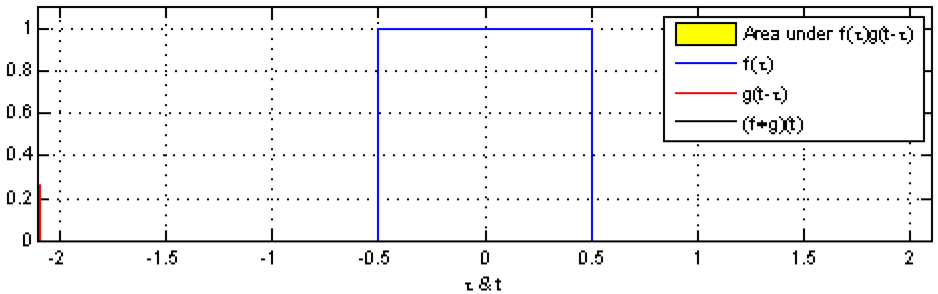
$(f * g)(t)$?

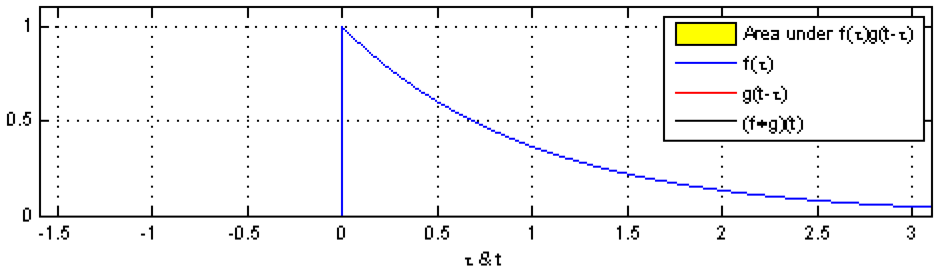


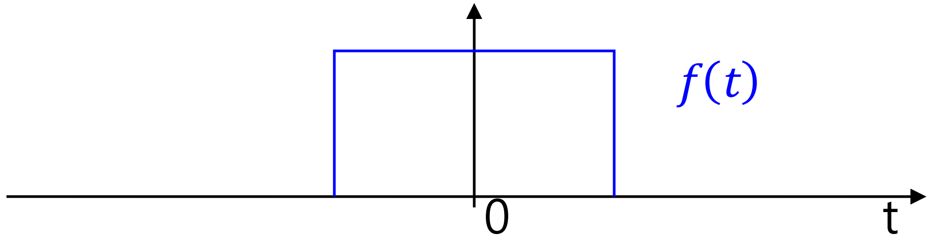
Multi-channel Source Separation

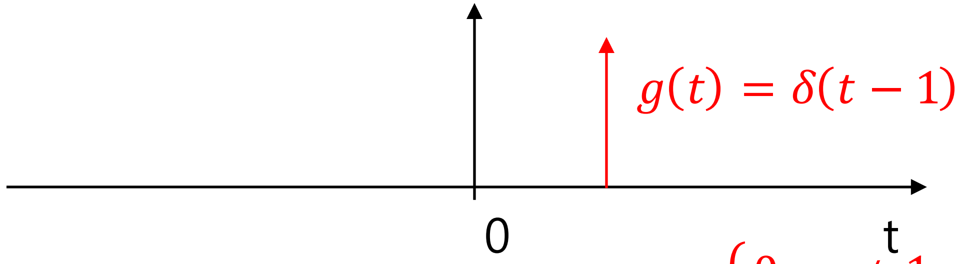
- Microphone Array
  - Convolution: $(f * g)(t) := \int_{-\infty}^{\infty} f(\tau) g(t-\tau)\, d\tau.$

Example of Convolution









Dirac delta function $\delta(t-1) = \begin{cases} 0, & x \neq 1 \\ \infty, & x = 1 \end{cases}$

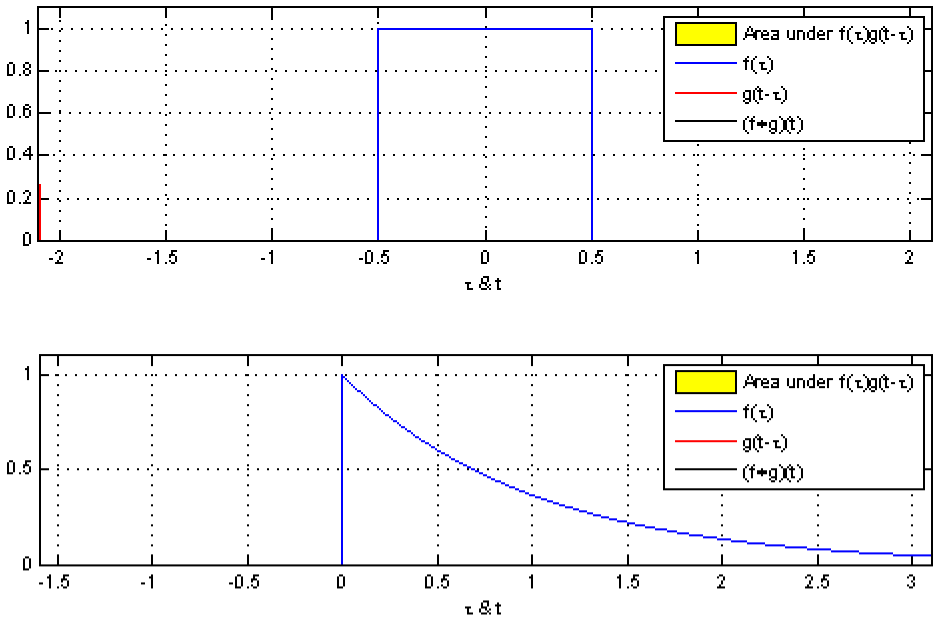
$$\int_{-\infty}^{\infty} \delta(t-1)\, dt = 1$$

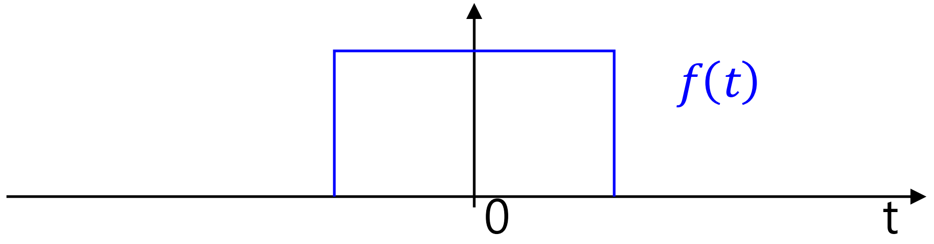
$(f * g)(t)$?

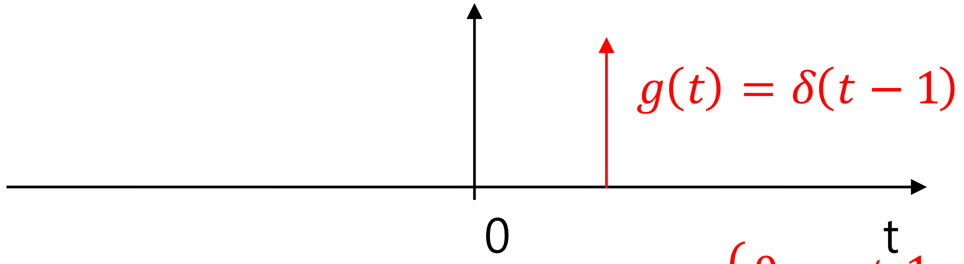
Multi-channel Source Separation

- Microphone Array
  - Convolution: $(f * g)(t) := \int_{-\infty}^{\infty} f(\tau) g(t - \tau)\, d\tau.$

Example of Convolution



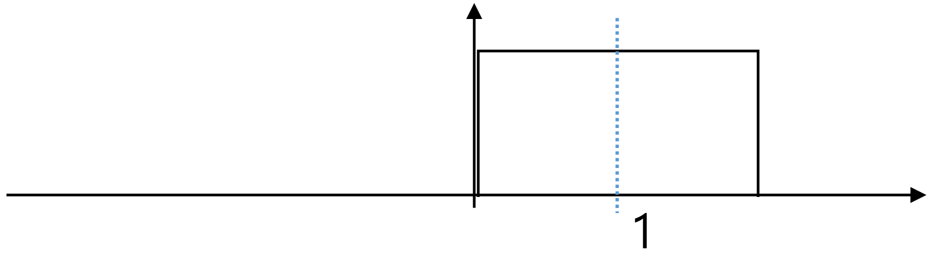




Dirac delta function $\delta(t-1) = \begin{cases} 0, & x \neq 1 \\ \infty, & x = 1 \end{cases}$

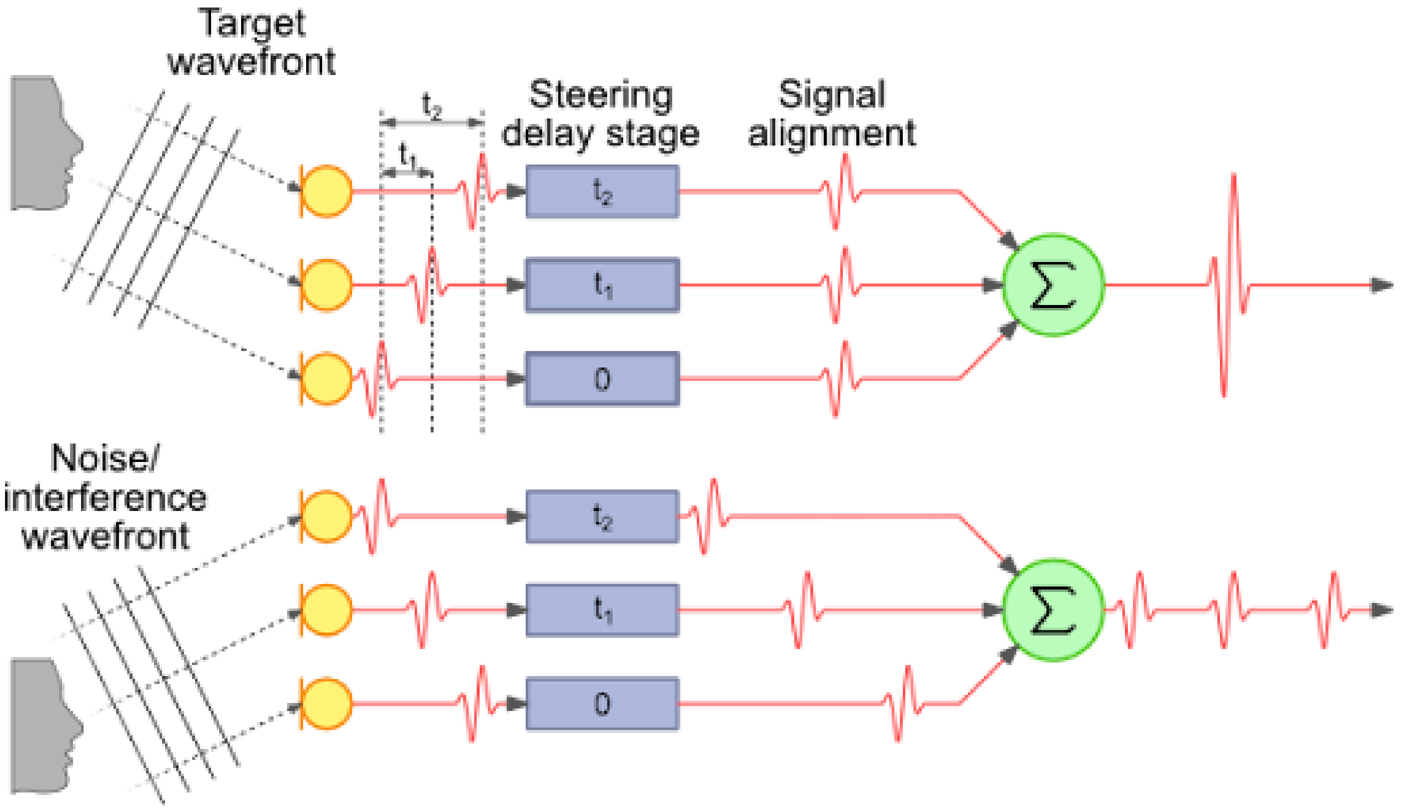
$$\int_{-\infty}^{\infty} \delta(t-1)\, dt = 1$$

$(f * g)(t)?$



Multi-channel Source Separation

- Delay-and-Sum Beamformer
  - Consider the propagation delays!



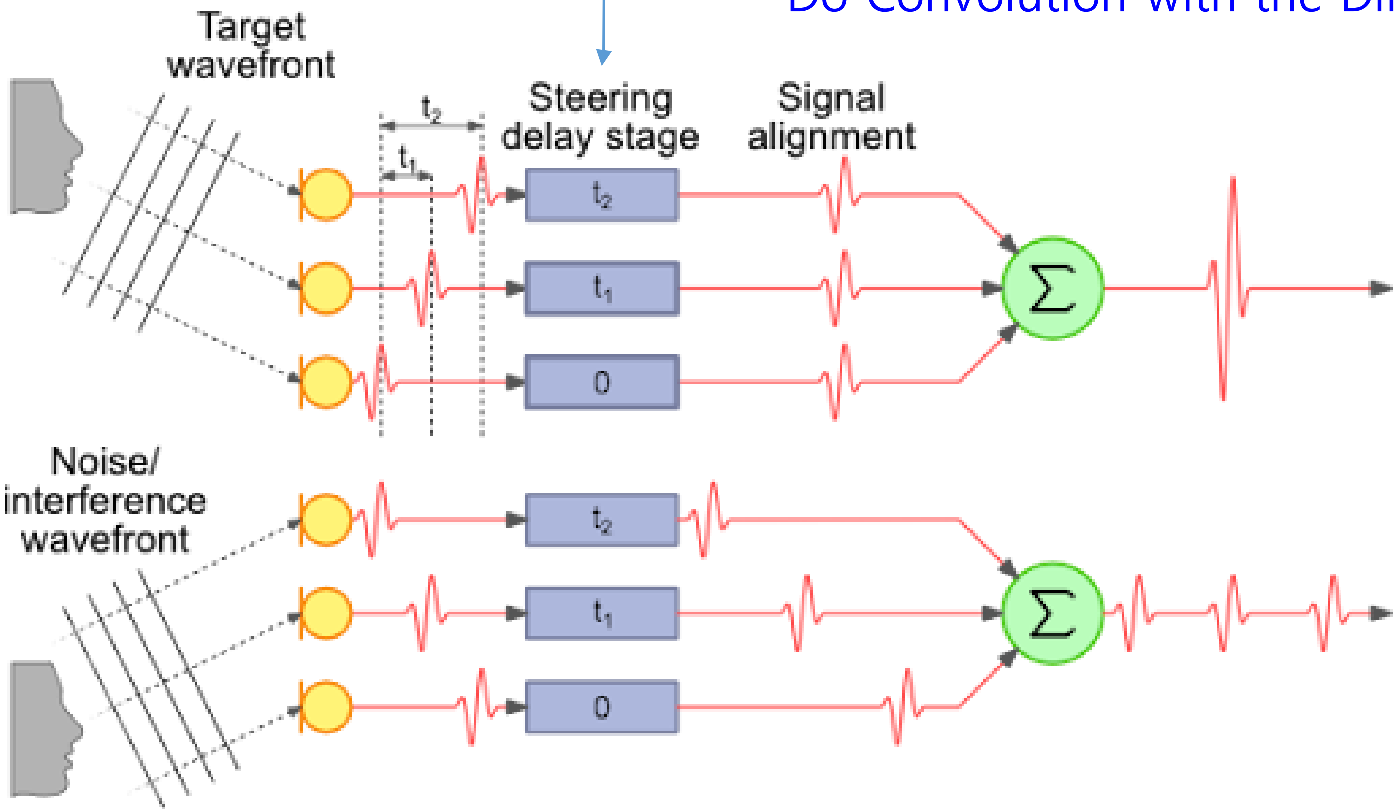
https://medium.com/@itberrios6/introduction-to-beamforming-part-1-92767408f483

Multi-channel Source Separation

- Delay-and-Sum Beamformer
  - Consider the propagation delays!

$g(t) = \delta(t + t_1)$

Do Convolution with the Dirac delta function!



https://medium.com/@itberrios6/introduction-to-beamforming-part-1-92767408f483

Multi-channel Source Separation

- Microphone Array
  - Is it possible to extract only the audio signal coming from a specific direction? → **Beamforming**
  - What is the beamforming algorithm?
    - Focusing a specific direction, $\boldsymbol{\theta}$, by using beamforming weights (spatial filter)

$$y(\boldsymbol{\theta}, \boldsymbol{t}) = \boldsymbol{w}(\boldsymbol{\theta}, -\boldsymbol{t})^{\boldsymbol{H}} \underset{\text{Convolution}}{*} \boldsymbol{x}(\boldsymbol{t}) \quad \underset{\text{FFT \& iFFT}}{\Longleftrightarrow} \quad \underbrace{y(\boldsymbol{\theta}, \boldsymbol{\omega})}_{\text{Beamforming output}} = \underbrace{\boldsymbol{w}(\boldsymbol{\theta}, \boldsymbol{\omega})^{\boldsymbol{H}}}_{\text{Weight (Spatial filter)}} \underbrace{\boldsymbol{x}(\boldsymbol{\omega})}_{\text{Measured pressure signals}}$$

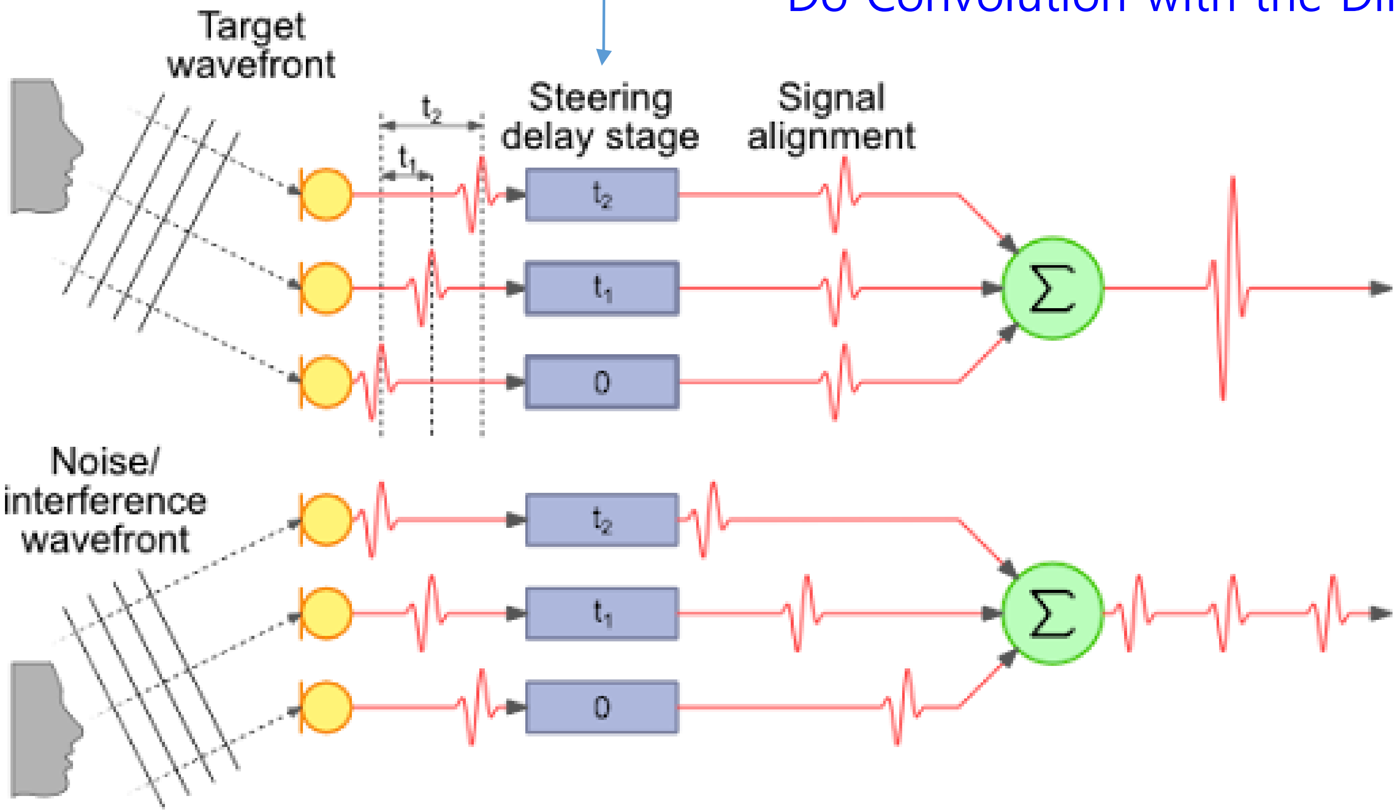
$$\boldsymbol{\delta}(\boldsymbol{t} + \boldsymbol{t}_1) \quad \underset{\text{FFT \& iFFT}}{\Longleftrightarrow} \quad \boldsymbol{e}^{+j\omega t_1} \quad \Rightarrow \quad \boldsymbol{y}(\boldsymbol{\theta}, \boldsymbol{\omega}) = \begin{bmatrix} \boldsymbol{e}^{+j\omega t_1} \\ \dots \\ \boldsymbol{e}^{+j\omega t_C} \end{bmatrix}^{\boldsymbol{H}} \cdot \boldsymbol{x}(\boldsymbol{\omega})$$

Multi-channel Source Separation

- Delay-and-Sum Beamformer
  - Consider the propagation delays!

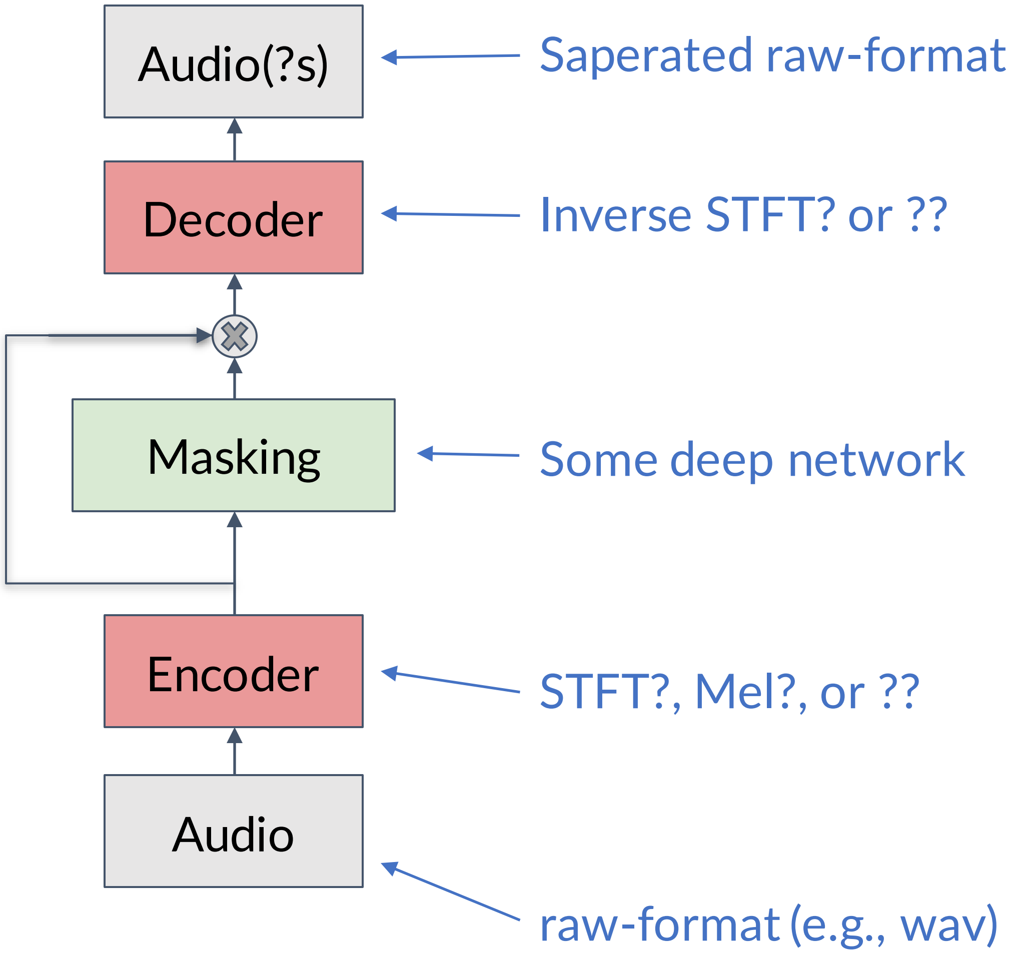
$g(t) = \delta(t + t_1)$

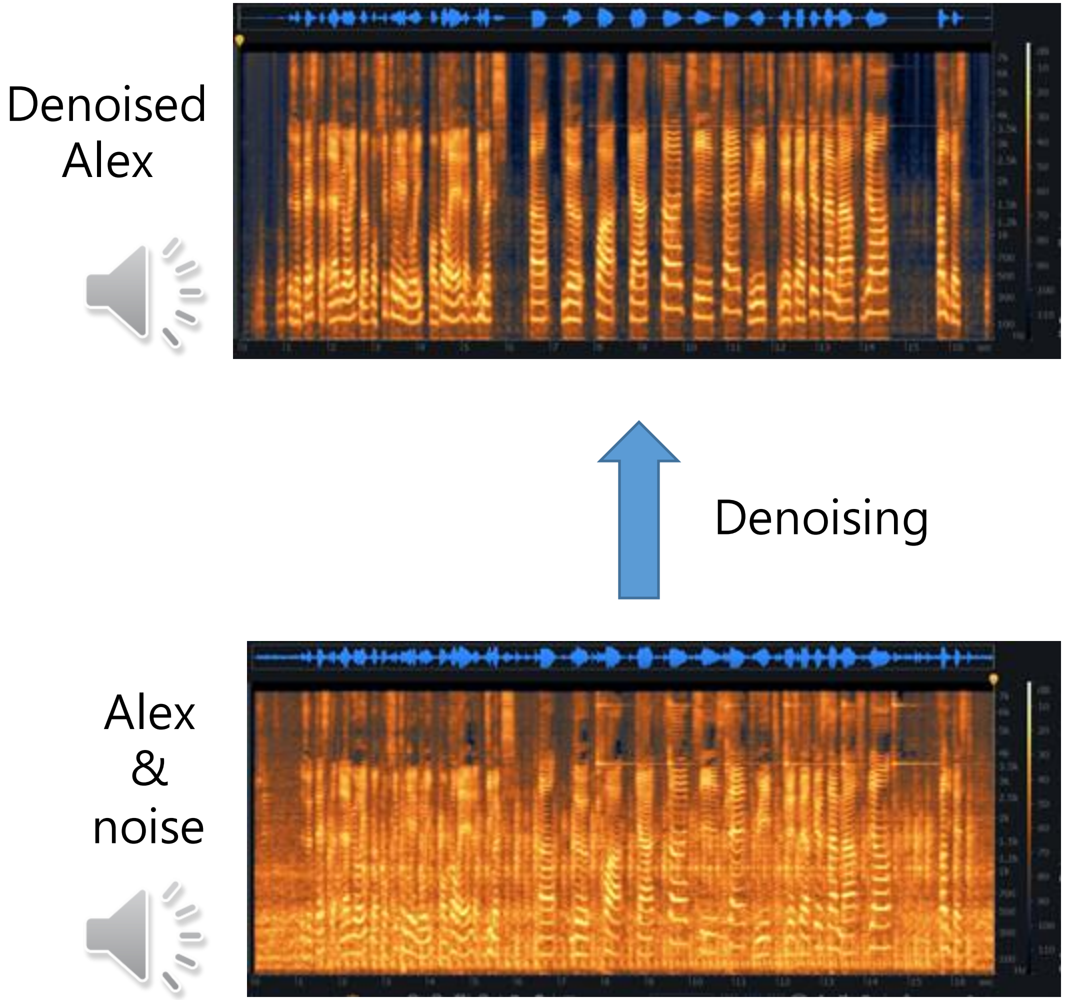
Do Convolution with the Dirac delta function!



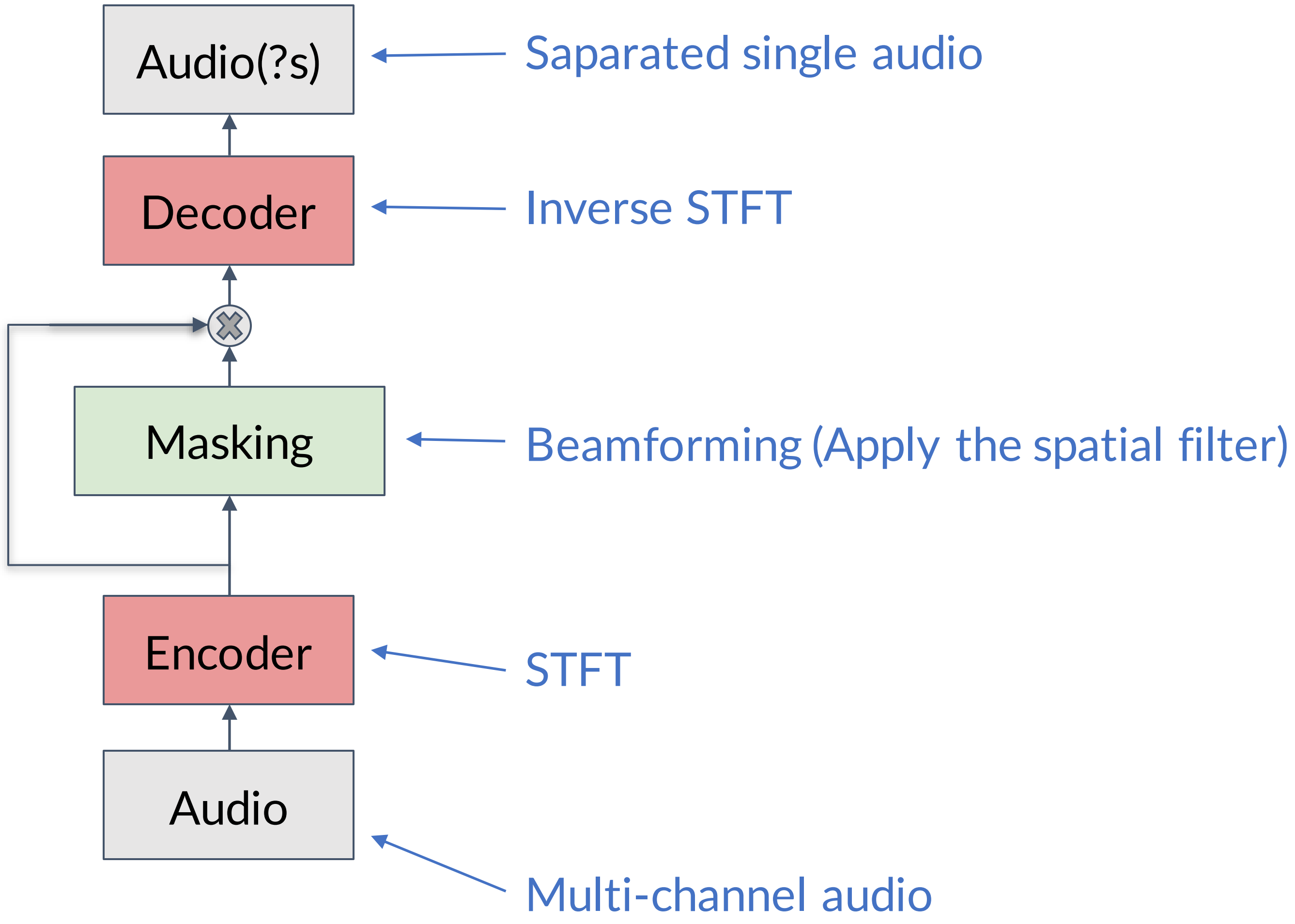
https://medium.com/@itberrios6/introduction-to-beamforming-part-1-92767408f483

Encoder-Separation-Decoder (ESD)



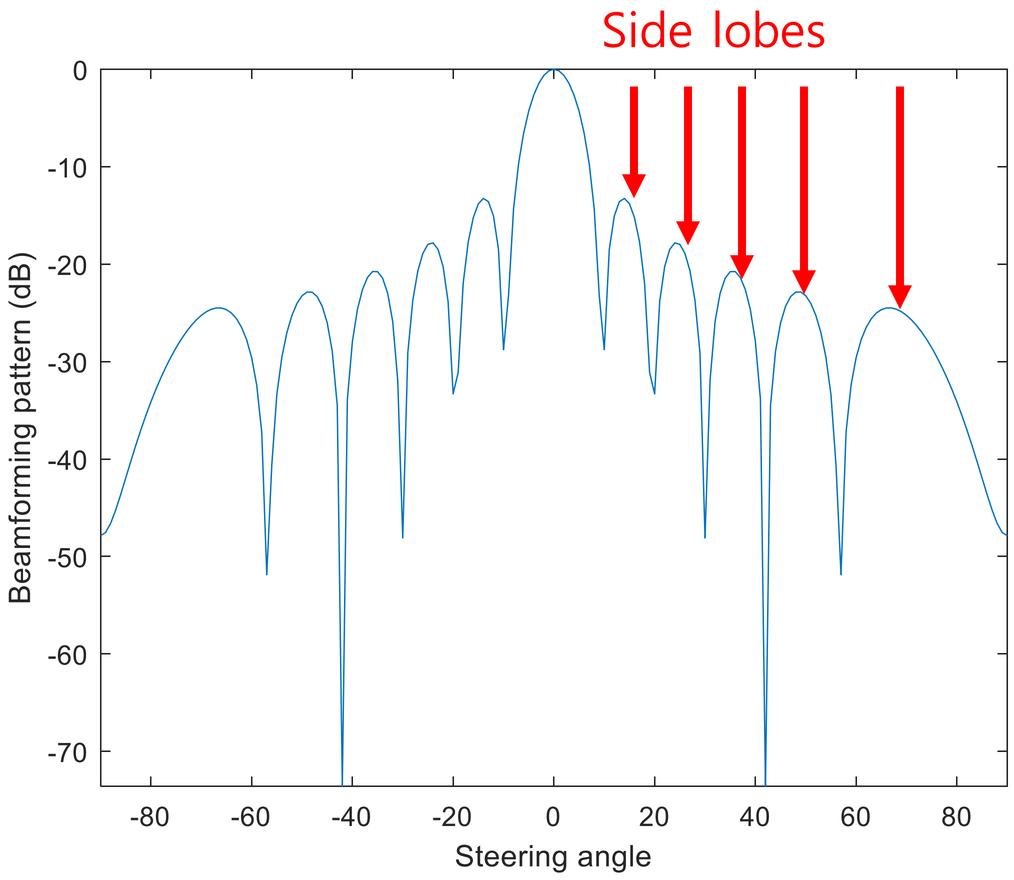


Multi-Channel Encoder-Separation-Decoder (ESD)



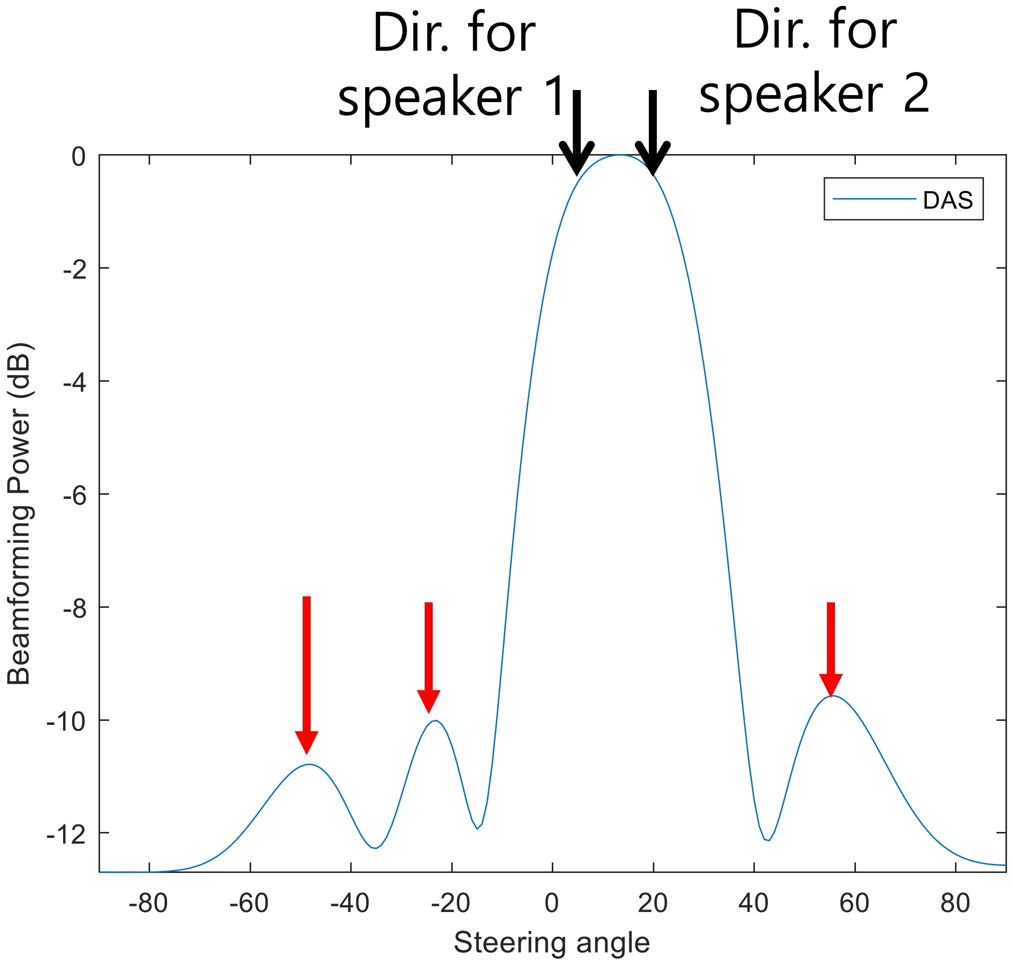
Multi-channel Source Separation

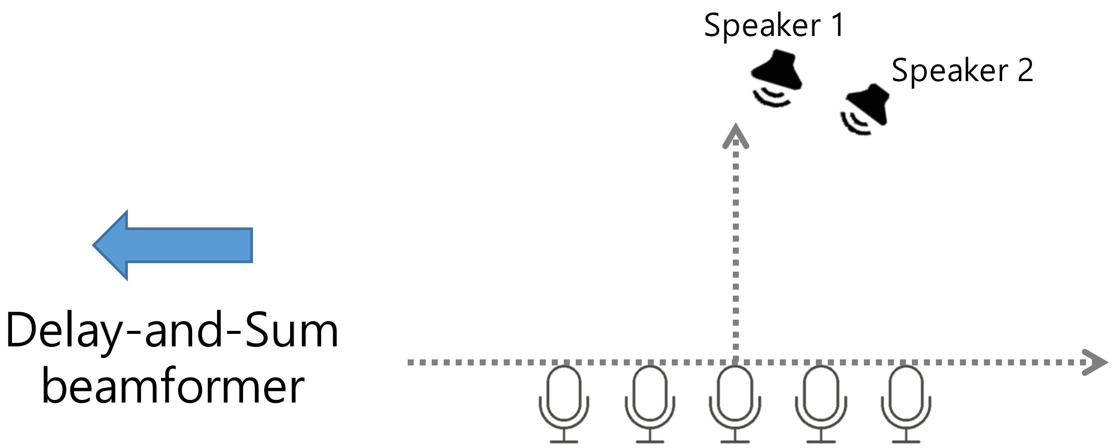
- The problem of the Delay-and-Sum Beamformer
  - The beamforming pattern of DAS beamformer is not sharp enough!



Multi-channel Source Separation

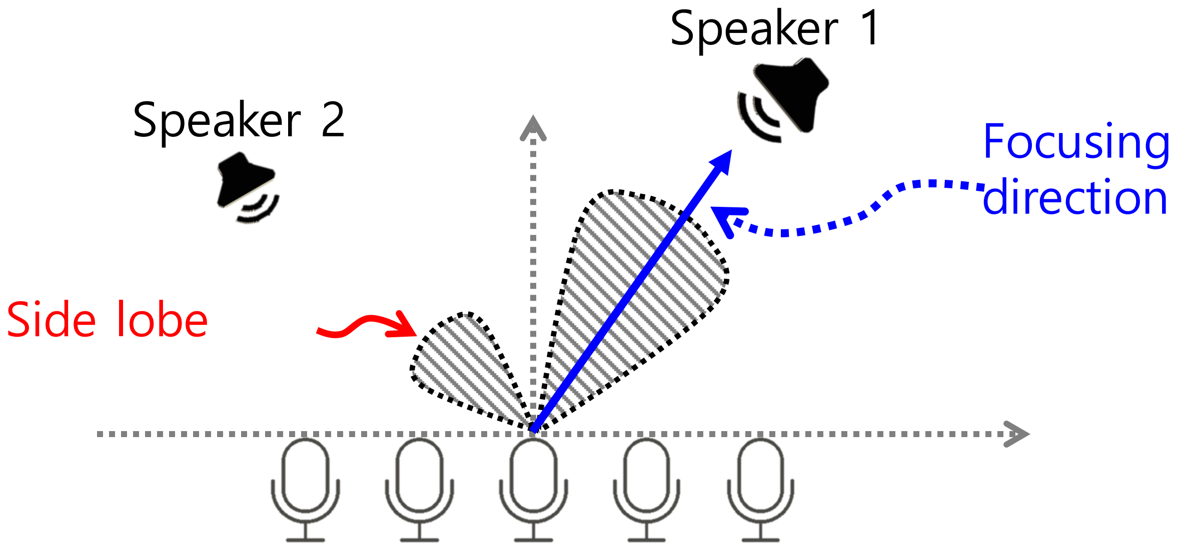
- The problem of the Delay-and-Sum Beamformer
  - The beamforming pattern of DAS beamformer is not sharp enough!





Multi-channel Source Separation

- The problem of the Delay-and-Sum Beamformer
  - The beamforming pattern of DAS beamformer is not sharp enough!

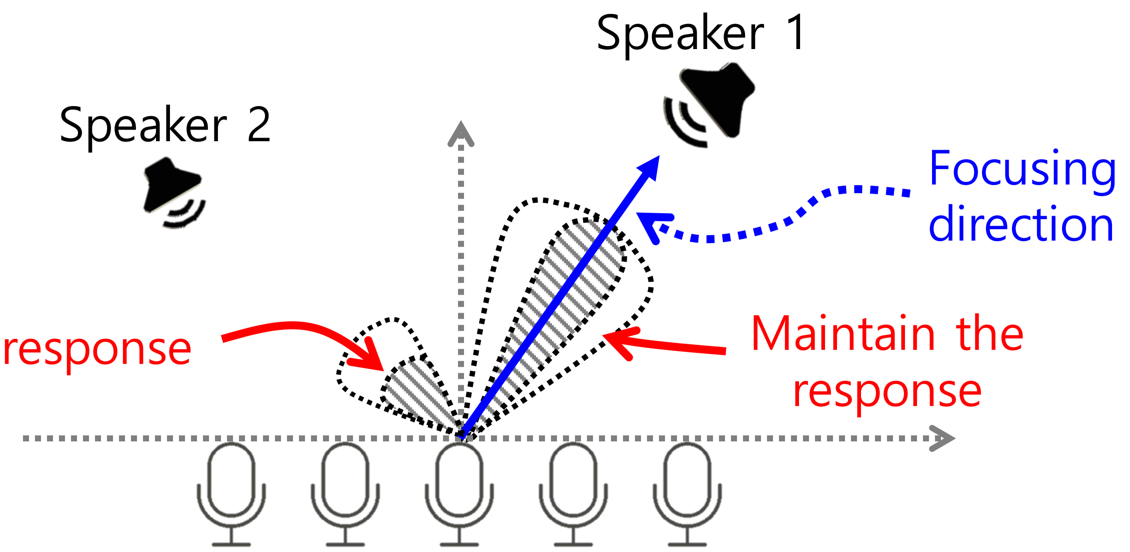


- How to reduce the response for the speaker 2? when focusing on the speaker 1…

MVDR beamformer!

Multi-channel Source Separation

- MVDR beamformer
  - Design a spatial filter w that only responds to a specific direction



  - Design a spatial filter w that minimizes the total beamforming power but maintain the response in the aiming angle

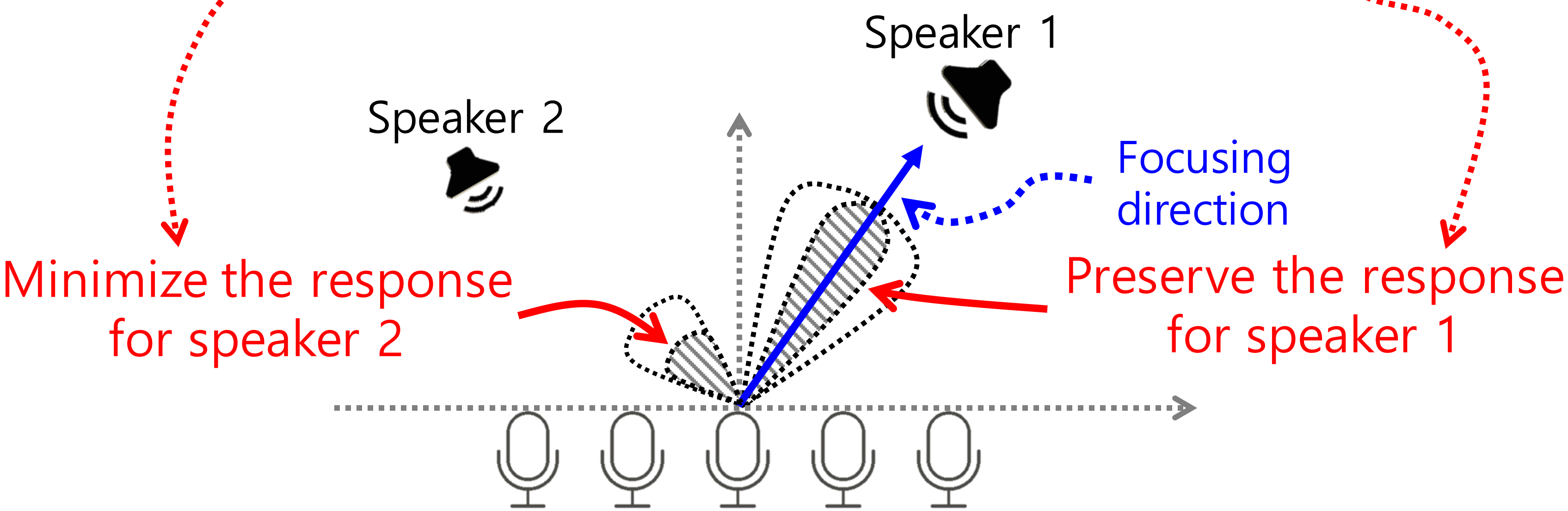
$$\textbf{Minimize } \boldsymbol{\beta} = \boldsymbol{w}^H E[\boldsymbol{x}(\boldsymbol{\omega})\boldsymbol{x}(\boldsymbol{\omega})^H]\boldsymbol{w} \qquad \textbf{Subject to } \boldsymbol{w}^H\mathbf{h} = \boldsymbol{b}_0$$

Plane wave model

Multi-channel Source Separation

- MVDR beamformer
  - Design a spatial filter w that minimizes the total beamforming power but maintain the response in the aiming angle

$$\textbf{Minimize } \boldsymbol{\beta}(\boldsymbol{\theta}) = \boldsymbol{w}(\boldsymbol{\theta})^{H}\boldsymbol{R}\boldsymbol{w}(\boldsymbol{\theta}) \qquad \textbf{Subject to } \boldsymbol{w}(\boldsymbol{\theta})^{H}\mathbf{h}(\boldsymbol{\theta}) = \boldsymbol{b}_{\mathbf{0}}$$



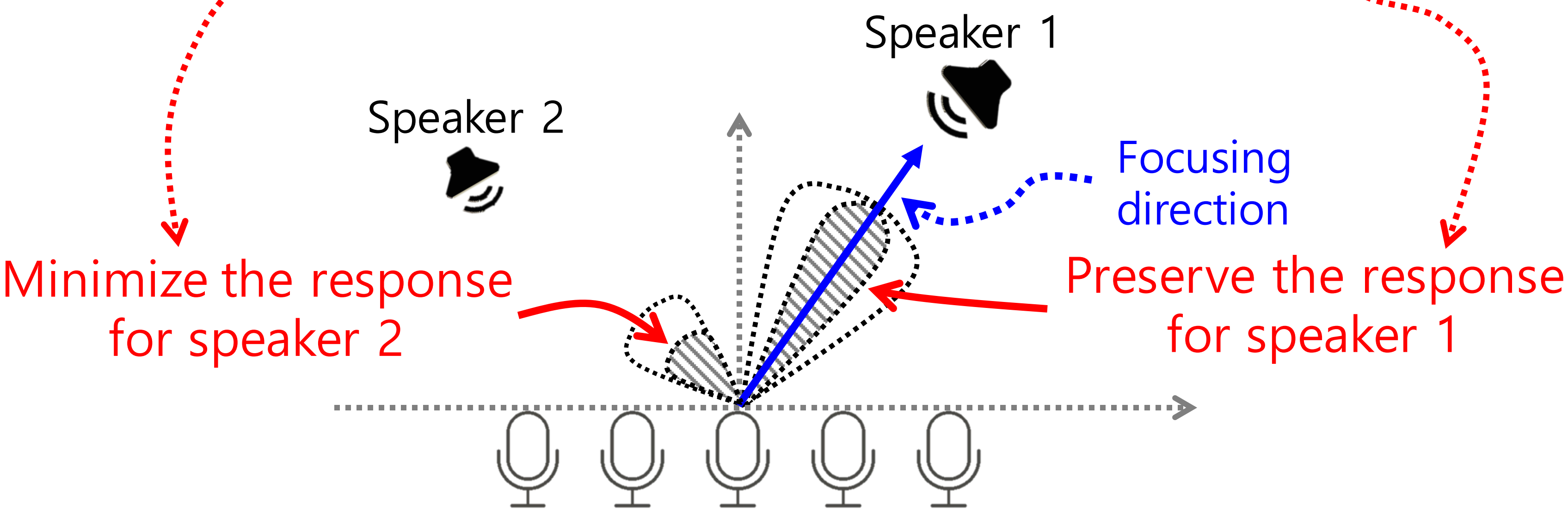
Multi-channel Source Separation

- MVDR beamformer
  - Design a spatial filter w that minimizes the total beamforming power but maintain the response in the aiming angle

$$\textbf{Minimize } \boldsymbol{\beta}(\boldsymbol{\theta}) = \boldsymbol{w}(\boldsymbol{\theta})^{H}\boldsymbol{R}\boldsymbol{w}(\boldsymbol{\theta}) \qquad \textbf{Subject to } \boldsymbol{w}(\boldsymbol{\theta})^{H}\mathbf{h}(\boldsymbol{\theta}) = \boldsymbol{b}_{0}$$

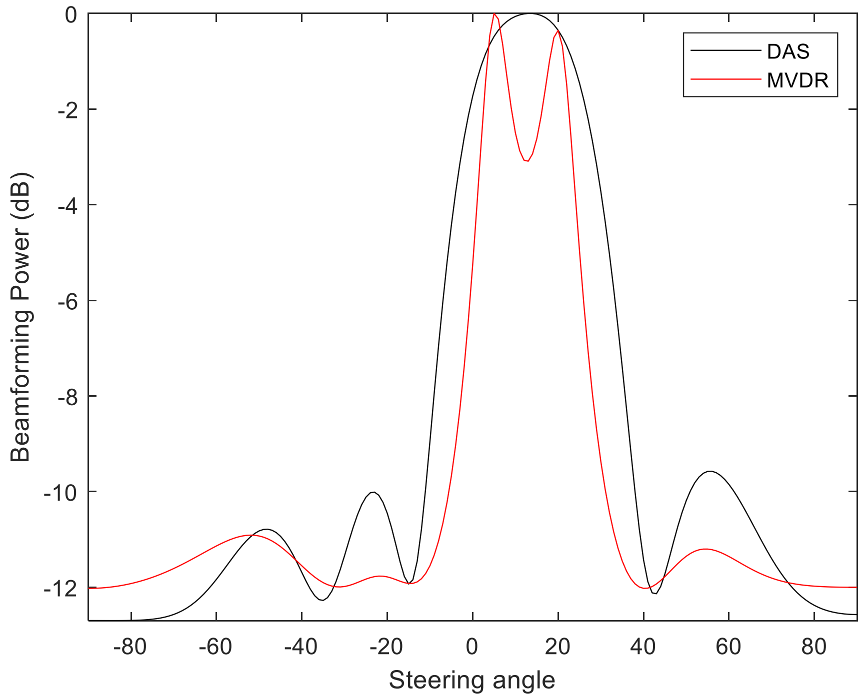
solve

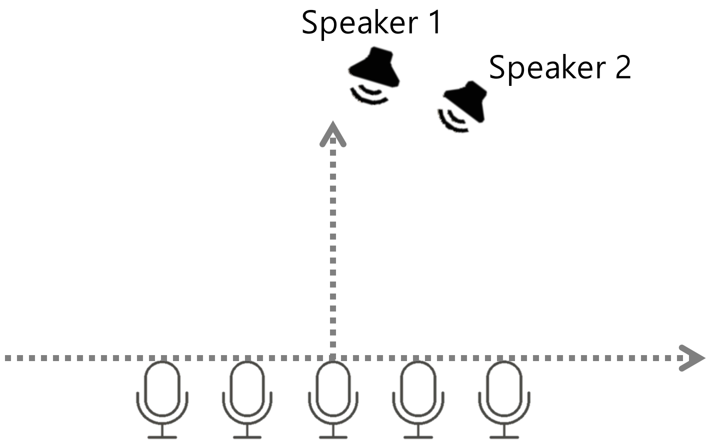
$$\Rightarrow \boldsymbol{w}(\boldsymbol{\theta}) = \frac{\boldsymbol{R}^{-1}\mathbf{h}(\boldsymbol{\theta})}{\mathbf{h}(\boldsymbol{\theta})^{H}\boldsymbol{R}^{-1}\mathbf{h}(\boldsymbol{\theta})}\boldsymbol{b}_{0}^{*}$$



Multi-channel Source Separation

- MVDR beamformer





Multi-channel Source Separation

- The limitation of the beamformer!
  - We have to know the direction-of-arrival (DoA) of the sound

$$y(\boldsymbol{\theta}, \boldsymbol{t}) = \boldsymbol{w}(\boldsymbol{\theta}, -\boldsymbol{t})^{\boldsymbol{H}} * \boldsymbol{x}(\boldsymbol{t}) \quad \Longleftrightarrow \quad y(\boldsymbol{\theta}, \boldsymbol{\omega}) = \boldsymbol{w}(\boldsymbol{\theta}, \boldsymbol{\omega})^{\boldsymbol{H}}\, \boldsymbol{x}(\boldsymbol{\omega})$$

Convolution

FFT & iFFT

Beamforming output

Weight (Spatial filter)

Measured pressure signals

- However, It is have to find out the accurate DoA…

Multi-channel Source Separation

- We have to know the direction-of-arrival (DoA) of the sound

$$y(\boldsymbol{\theta},\boldsymbol{t}) = \boldsymbol{w}(\boldsymbol{\theta},-\boldsymbol{t})^{\boldsymbol{H}} * \boldsymbol{x}(\boldsymbol{t}) \quad \Longleftrightarrow \quad y(\boldsymbol{\theta},\boldsymbol{\omega}) = \boldsymbol{w}(\boldsymbol{\theta},\boldsymbol{\omega})^{\boldsymbol{H}}\, \boldsymbol{x}(\boldsymbol{\omega})$$

Convolution

FFT & iFFT

Beamforming output

Weight (Spatial filter)

Measured pressure signals

- However, It is have to find out the accurate DoA…