Low-Latency Real-Time Blind Source Separation with Binaural Directional Hearing Aids

Masahiro Sunohara[1], Chiho Haruta[1], Nobutaka Ono[2]

[1] RION CO., LTD. [2] National Institute of Informatics

CHAT-2017, 19th August, 2017

Outline of presentation

I. Introduction

II. Low-latency real-time BSS

III. Directional microphone in hearing aids

IV. Evaluation

V. Conclusion

Ⅰ. Introduction

Background

- Hearing-impaired listeners find it difficult to understand speech in noisy environments.

Crowded restaurant
- Speech
- Background music
- Clatter of dishes

⋮

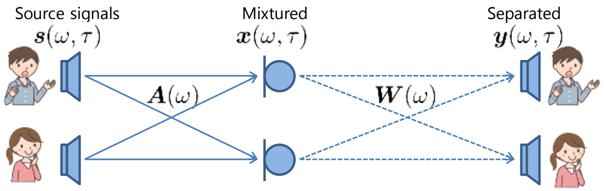
- In these situations, it is difficult to focus a desired sound.
- Unfortunately, current hearing aids are often ineffective in these situations.

The purpose of this study:

Improving speech communication for hearing-impaired persons in noisy environments using hearing aids.

To focus a target sound

- We consider a multi-microphone system in this study.
- Beamforming is one of the familiar technique for solving this problem, however, a perfect voice activity detection (VAD) or prior information of a target sound source are required.



- Blind source separation (BSS) is an effective technique to extract a desired source without VAD or prior information of a target source.

Blind source separation (BSS) technique

- For convolutive mixtures, independent vector analysis (IVA) [Kim2006, Hiroe2006] in the frequency domain have been developed as a standard technique of the BSS.
- There is a state-of-the-art approach for the IVA: Auxiliary-function-based IVA (AuxIVA) [Ono2011]

  Fast convergence speed, Low calculation cost, No permutation ambiguity

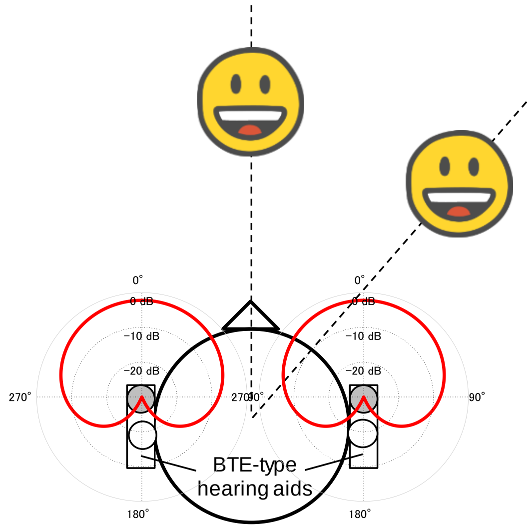
- However, frequency-domain BSSs (including AuxIVA) have a long algorithmic delay of at least one frame length.

- We have proposed a low-latency algorithm for real-time BSS based on the online AuxIVA [Sunohara 2017].
  (Algorithmic delay < 10 ms, flame length 4096@16 kHz)

Directional microphones + Low-latency AuxIVA ?

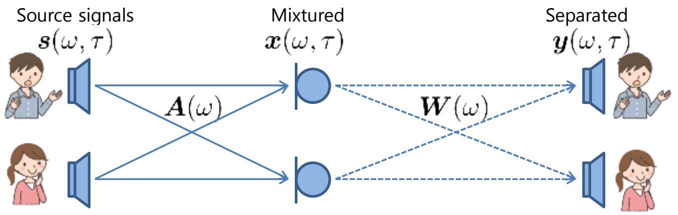
- Bilateral directional microphones have been widely used in actual hearing aids to improve the SNR of front speech signals.
- We investigate the separation performance of binaural BSS based on the low-latency online AuxIVA with directional microphones.

Directional microphones



+

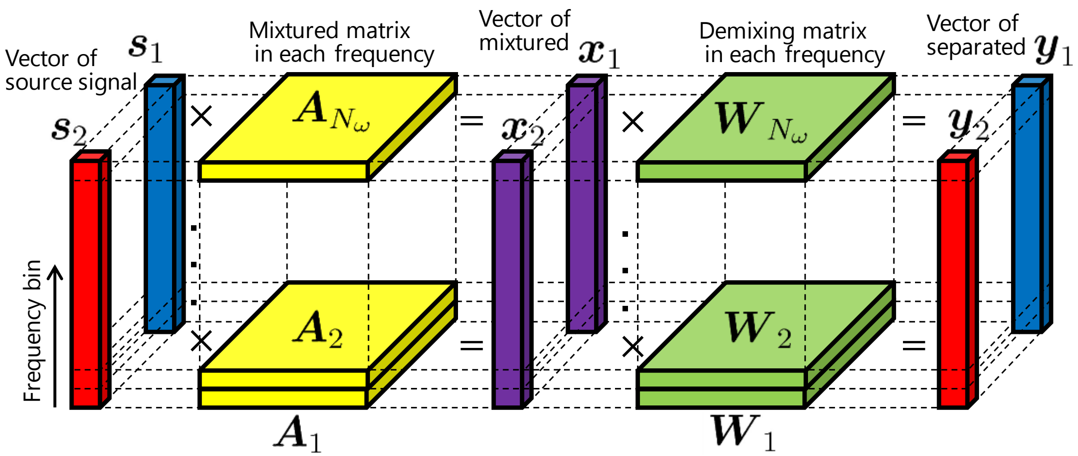
Binaural BSS based on the low-latency online AuxIVA



?

Ⅱ. Low-latency real-time BSS

Overview of online AuxIVA [Taniguchi 2014]



$$\boldsymbol{x}(\omega,\tau) = \boldsymbol{A}(\omega)\boldsymbol{s}(\omega,\tau)$$

$$\boldsymbol{y}(\omega,\tau) = \boldsymbol{W}(\omega)\boldsymbol{x}(\omega,\tau)$$

Demixing Matrix $\boldsymbol{W}$ is estimated to separate $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ independently with considering higher-order correlation between frequency bins.

Cost function

$$J(\boldsymbol{W}) = \frac{1}{N_\tau}\sum_{\tau=1}^{N_\tau}\sum_{k=1}^{K} G(\boldsymbol{y}_k(\tau)) - \sum_{\omega=1}^{N_\omega} \log|\det \boldsymbol{W}(\omega)|$$

(Supposing a spherical laplace distribution)

Weighted covariance matrix update

$$r_k(\tau) = \sqrt{\sum_{\omega=1}^{N_\omega} \left|\boldsymbol{w}_k^h(\omega;\tau)\boldsymbol{x}(\omega,\tau)\right|^2},$$

$$V_k(\omega;\tau) = \alpha V_k(\omega;\tau-1) + (1-\alpha)\frac{G'(r_k(\tau))}{r_k(\tau)}\boldsymbol{x}(\omega,\tau)\boldsymbol{x}^h(\omega,\tau),$$

Demixing matrix update

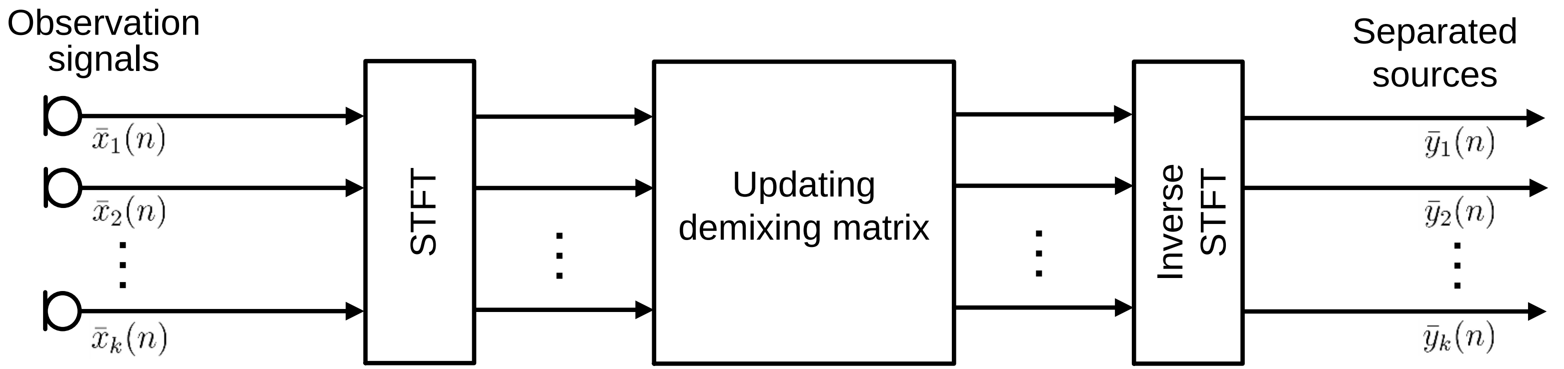
$$W(\omega;\tau) = W(\omega;\tau-1).$$

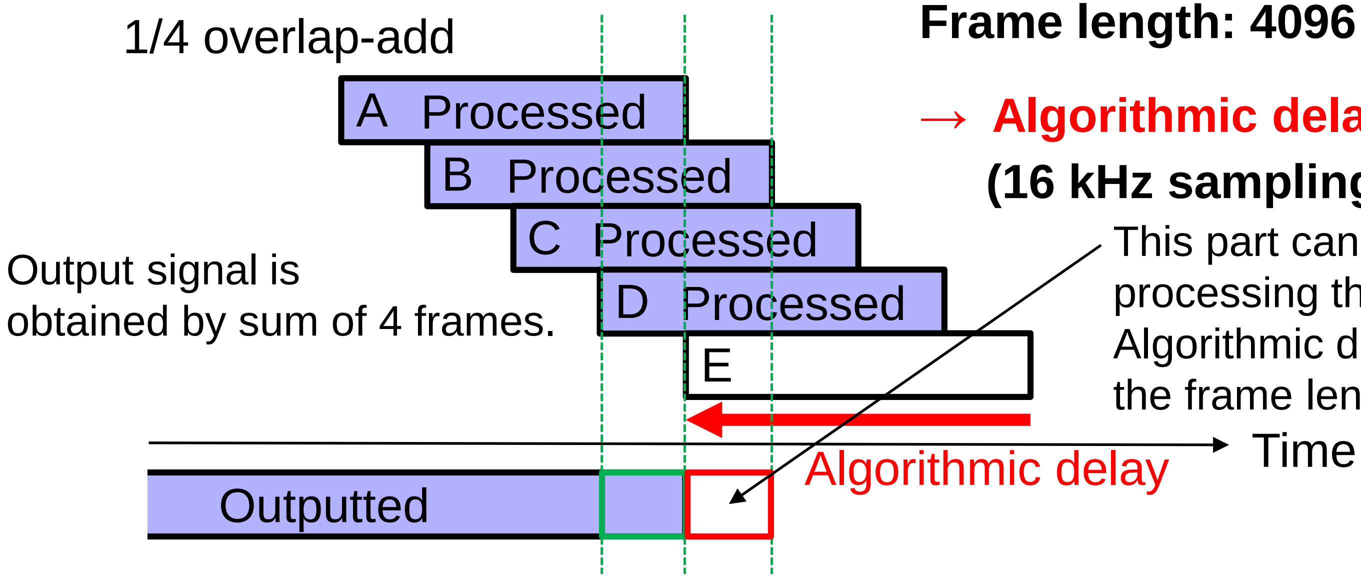
$$\boldsymbol{w}_k(\omega;\tau) \leftarrow (W(\omega;\tau)V_k(\omega;\tau))^{-1}\boldsymbol{e}_k,$$

$$\boldsymbol{w}_k(\omega;\tau) \leftarrow \boldsymbol{w}_k(\omega;\tau)/\sqrt{\boldsymbol{w}_k^h(\omega;\tau)V_k(\omega;\tau)\boldsymbol{w}_k(\omega;\tau)},$$

Algorithmic delay of the frequency-domain BSS

Block diagram of the standard frequency-domain BSS (including AuxIVA)





Frame length: 4096 samples

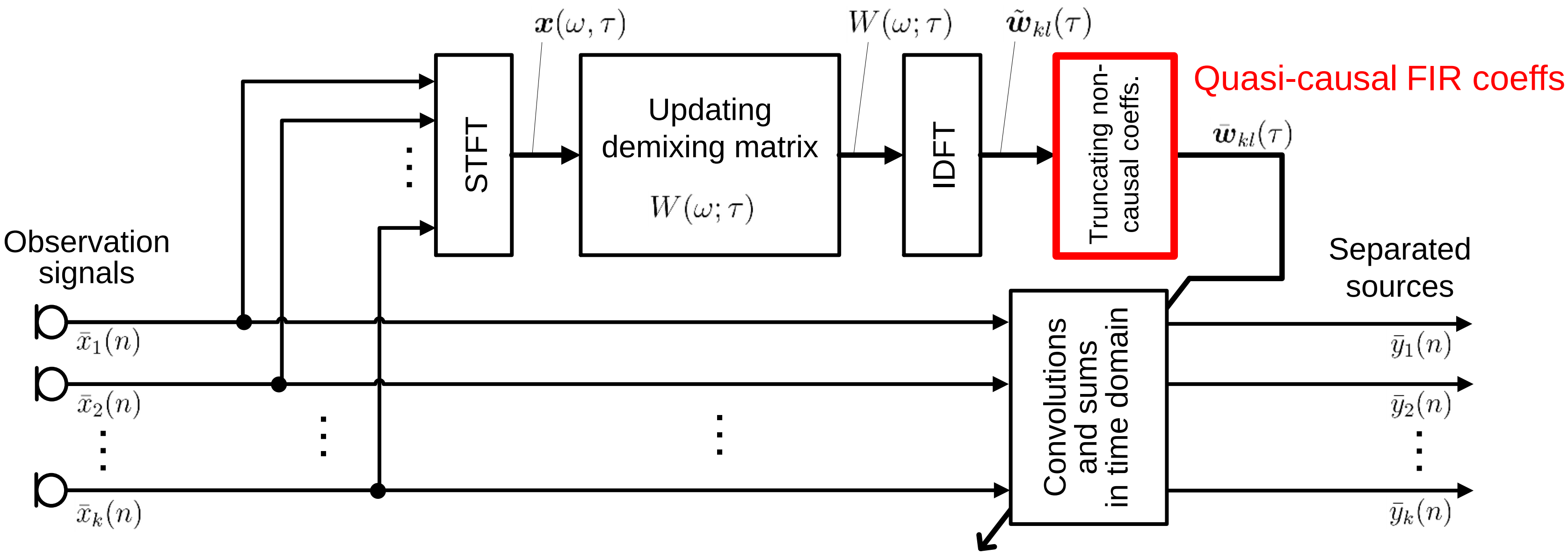
→ Algorithmic delay: 256 ms

(16 kHz sampling frequency)

This part can be calculated after processing the frame E. Algorithmic delay is equal to the frame length.

Image of the algorithmic delay for frequency-domain BSS

Low-latency online AuxIVA [Sunohara 2017]



Block diagram of a low-latency version of the online AuxIVA

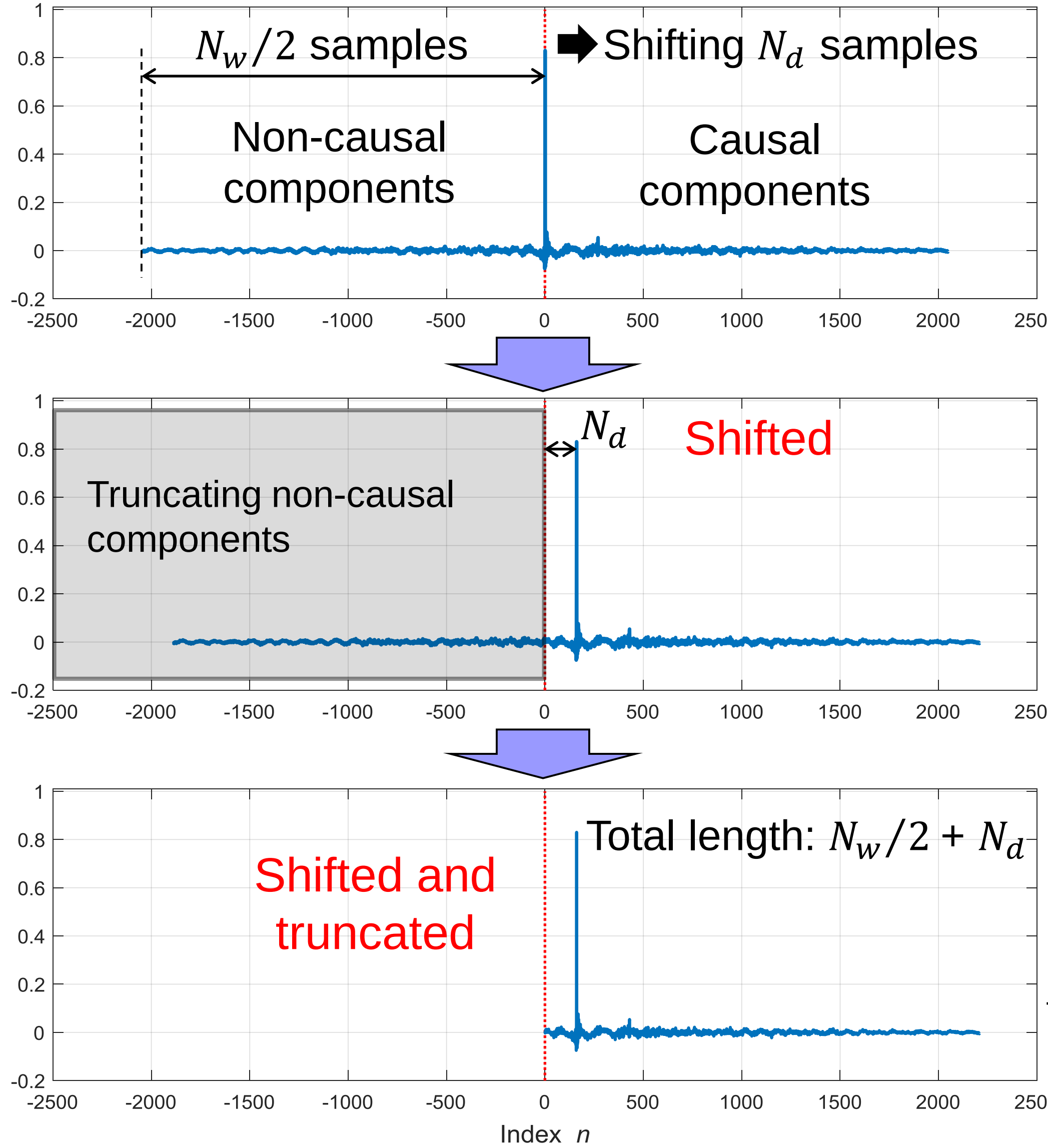
- For separating the sources using quasi-causal FIR filters in the time domain.

Realization as quasi-causal FIR filter

F_s: 16 kHz, Frame length N_w: 4096, Number of shifting N_d: 160

Original demixing FIR filter coefficients $\widetilde{w}_{kl}(n;\tau)$

Algorithmic Delay $N_w/2$ samples → 128 ms

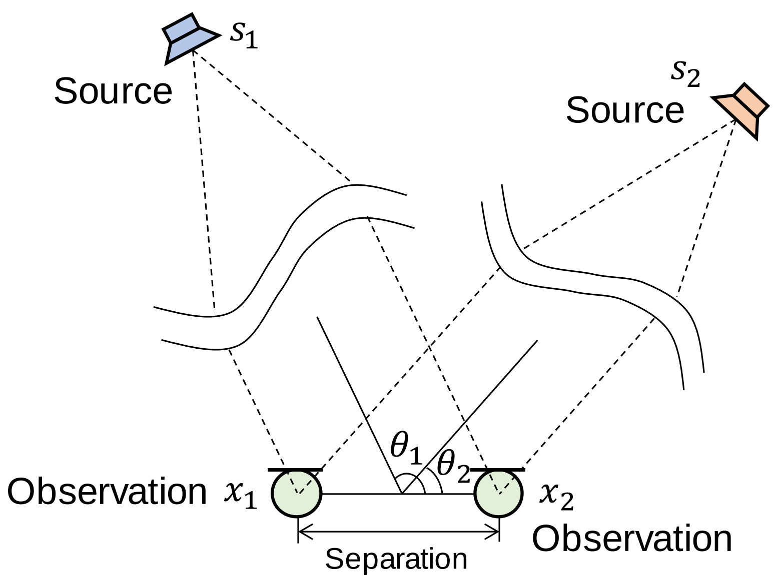


Shifted and truncated demixing FIR filter coefficients $\overline{w}_{kl}(n;\tau)$

Algorithmic Delay N_d samples → **10 ms**

Causality of demixing impulse response

- If all the non-causal components of the demixing FIR filter are originally zero, the algorithmic delay of the system can theoretically be zero without degradation.
- For simple model consisting of two sources and two mics, a theoretical sufficient condition for the ideal separation filters to be causal is obtained as .. [Sunohara 2017]



Simple model

$$[\log a(\theta_2) - \log a(\theta_1)] \cdot [\tau(\theta_2) - \tau(\theta_1)] < 0 \quad (2)$$

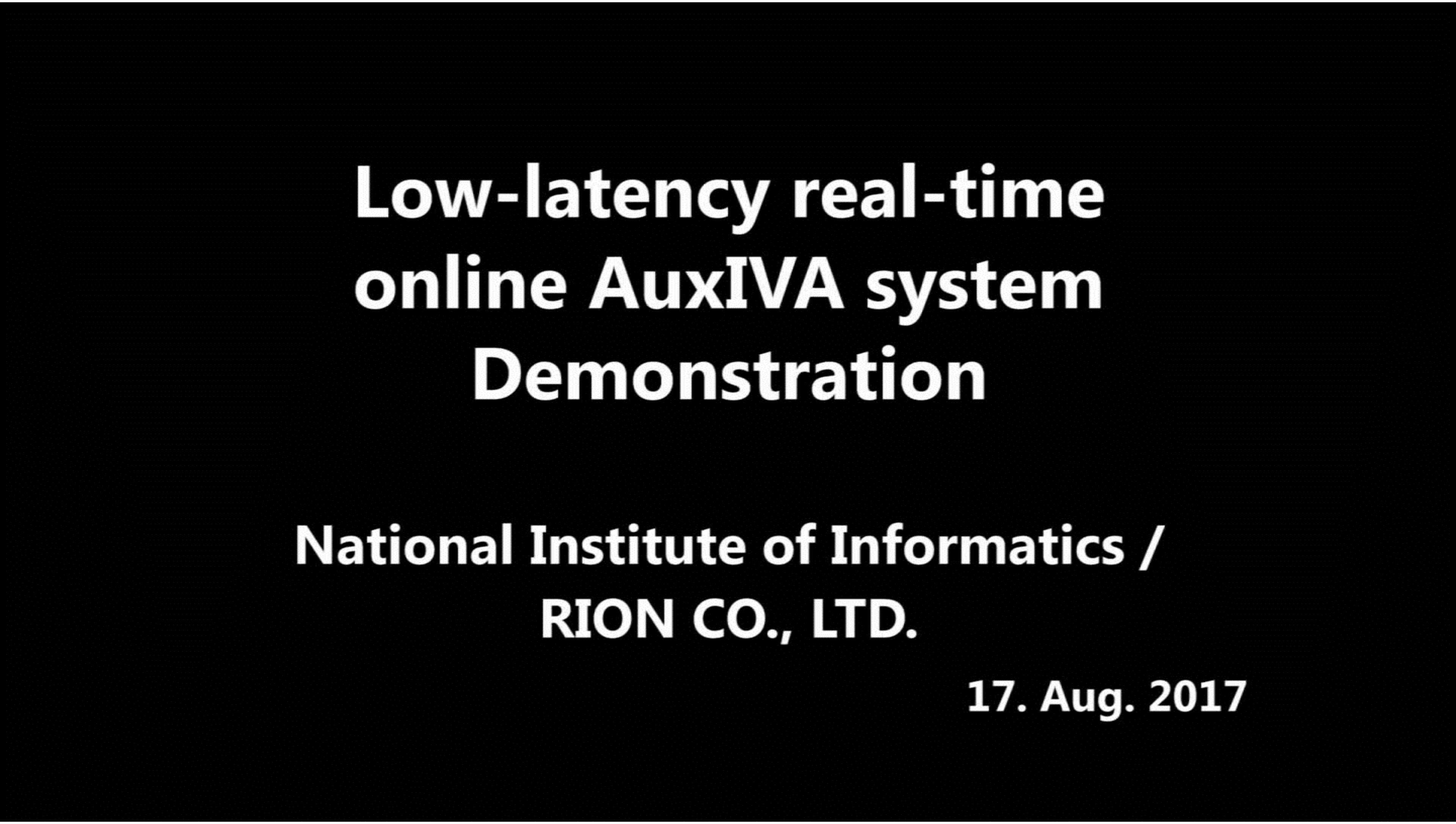
"An earlier channel is louder."

→ All non-causal components become 0.

$$\begin{cases} a_k = a(\theta_k)\text{: amplitude ratio ..} \\ \tau_k = \tau(\theta_k)\text{: time difference ..} \end{cases}$$

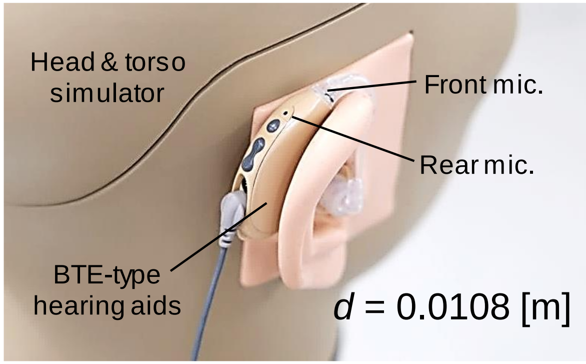
of the second channel relative to the first channel for a source with direction θ_k.

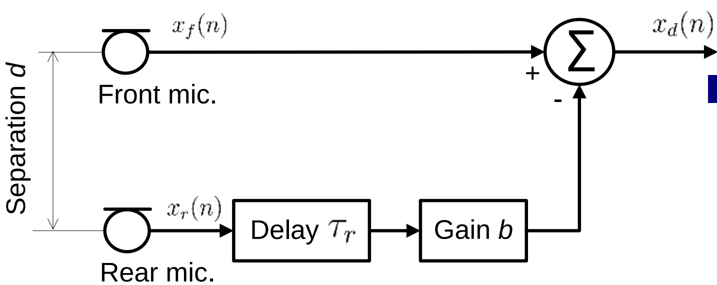
Demonstration

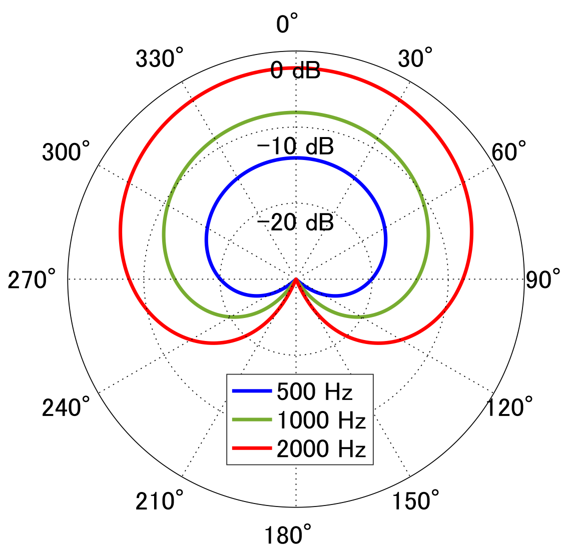


Ⅲ. Directional microphone in hearing aids

Directional microphone in hearing aids







- Directivity in a hearing aid is produced by a pair of omnidirectional microphone.
- When $\tau_r = d/c$, the directional pattern becomes cardioid.
- Sensitivity of the response at the lower frequencies is attenuated by 6 dB / octave.
- These spatial directional responses may affect the causality of the demixing impulse response

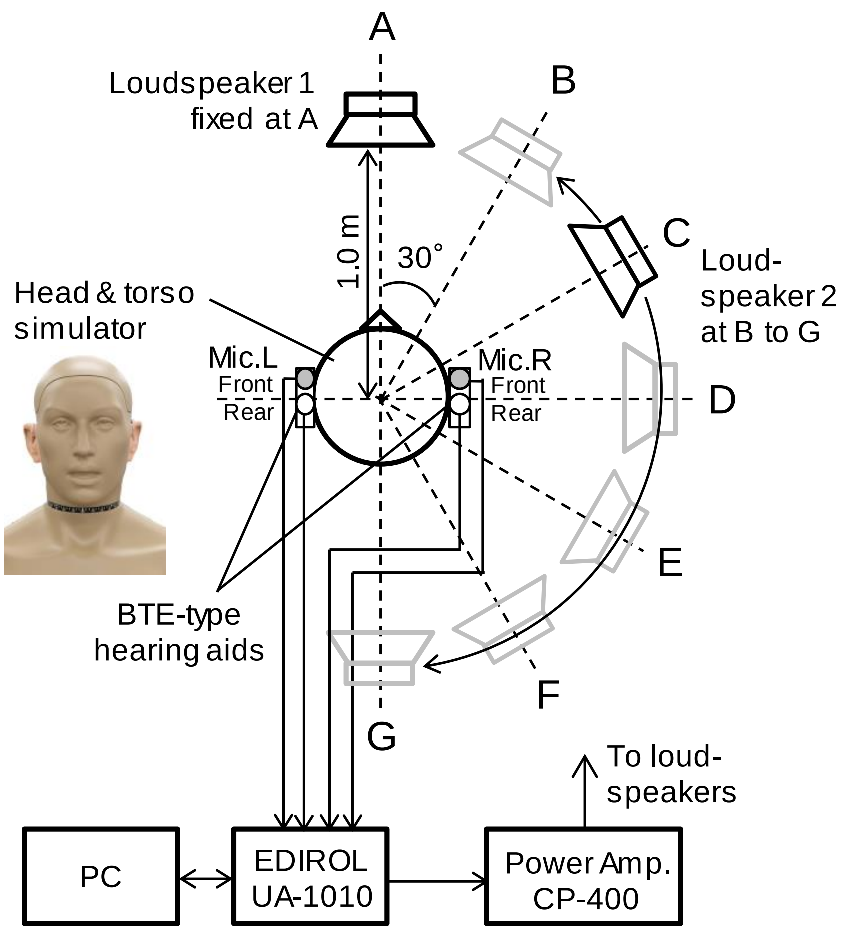
$$[\log a(\theta_2) - \log a(\theta_1)] \cdot [\tau(\theta_2) - \tau(\theta_1)] < 0 \quad (2)$$

- Separation performance of the low-latency BSS ? → investigated

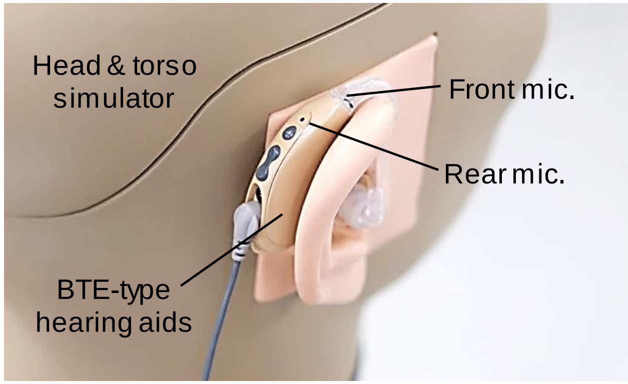
Ⅳ. Evaluation

Evaluation - experimental setup

Binaural BTE-type hearing aids with omni / directional mics



Meeting room (5.0×10.0×2.7 m)



Reverberation time: 650 ms at 500 Hz

Evaluation - conditions

- Sources: RWCP Japanese News Speech Corpus (Signal length: 30 s × 10 set for each direction)
- Microphone spacing: 18cm
- Microphone: Omnidirectional / Directional
- Sampling frequency: 16 kHz
- Frame length: 4096 samples
- Frame shift: 1024 samples (75 % Overlap)
- Window function: Hanning
- Evaluation index: Signal-to-interference ratio (SIR)

	Low-latency AuxIVA 128 ms	**Low-latency AuxIVA 10 ms**
FIR coeff. shift: N_d	2048 samples	160 samples
Algorithmic delay	128 ms	10 ms

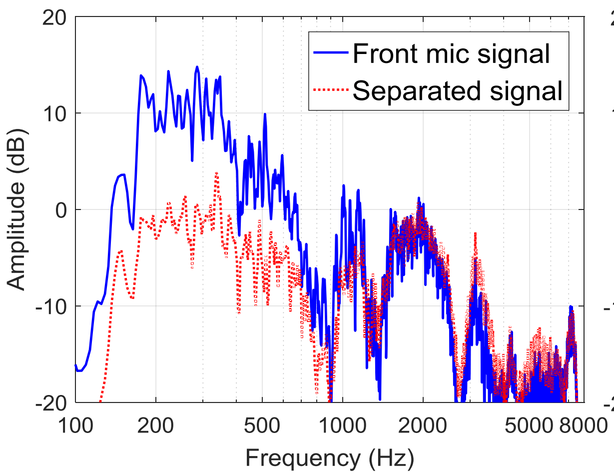
Compensation for directional response

- For fair comparison, it is necessary to compensate the difference in the sensitivity response associated with the directivity.
- We derive a compensation filter by minimizing the following cost function as post-processing :

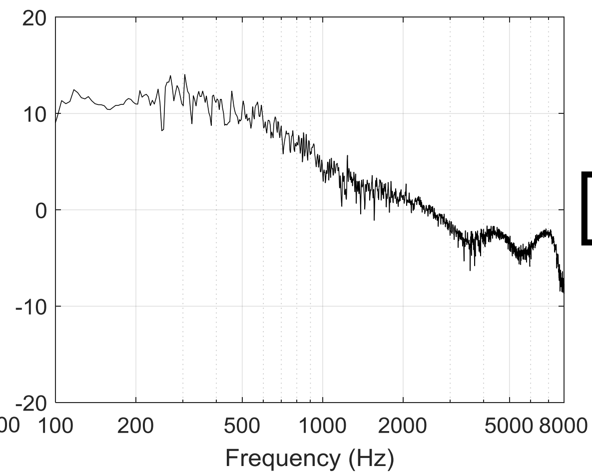
$$J(\boldsymbol{C}(\omega)) = \sum_{\tau} \left| r(\omega,\tau) - \boldsymbol{C}^H(\omega)\boldsymbol{Y}(\omega,\tau) \right|^2$$

$\begin{bmatrix} y_1(\omega,\tau) \\ y_2(\omega,\tau) \end{bmatrix}$: Separated signals

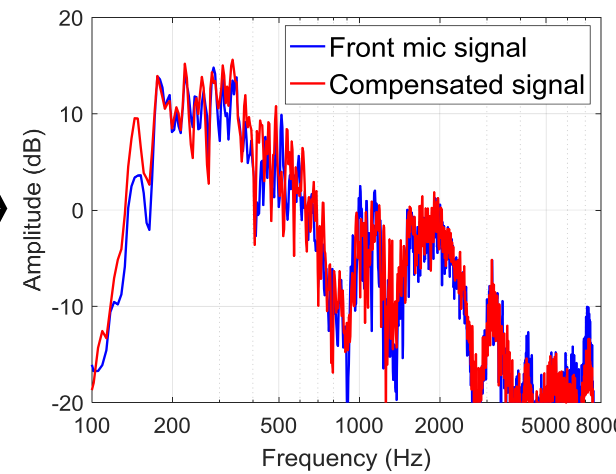
$r(\omega,\tau)$: STFT of the front mic, $\begin{bmatrix} c_1(\omega) \\ c_2(\omega) \end{bmatrix}$: Compensation filter



Averaged amplitude responses of the front mic(omni) and separated(directional) signal

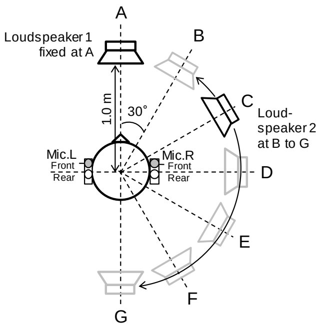


Amplitude response of Compensated filter $c_k(\omega)$

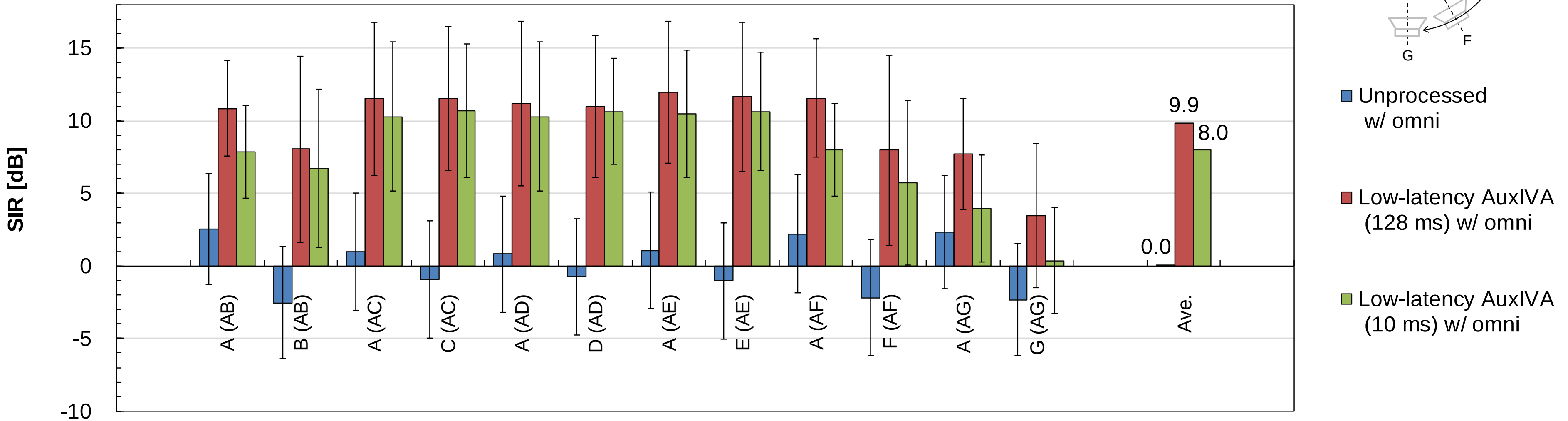


Averaged amplitude responses of the front mic(omni) and compensated separated signal

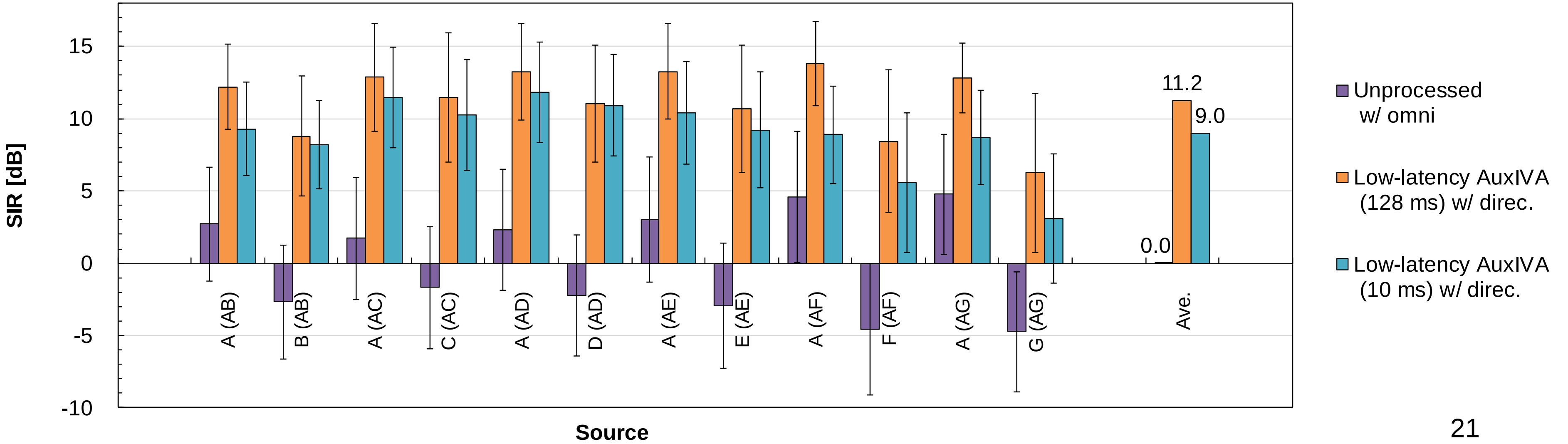
Results



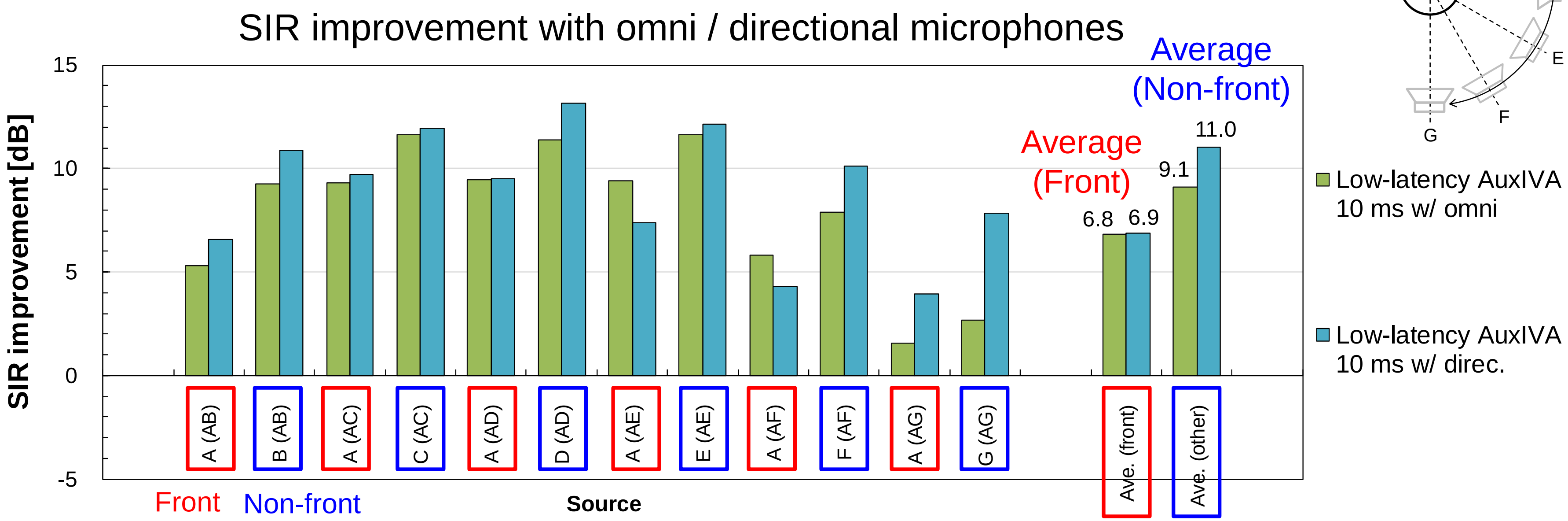
Separation performance with omnidirectional microphones

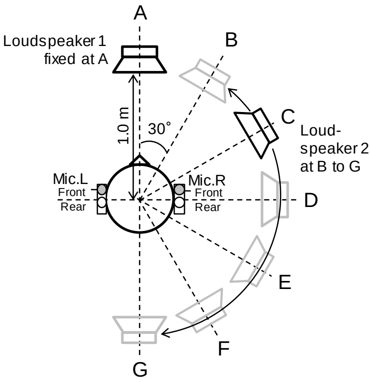


Separation performance with directional microphones



Discussion





- Separation performance with directional microphone for the front source is almost same as that with omnidirectional microphone.
- Separation performance with directional microphone for non-front source is better than that with omnidirectional microphone.

Conclusion

- We evaluated the separation performance of low-latency online AuxIVA with directional microphones for binaural hearing aids.
- The averaged SIR of the low-latency (10 ms) AuxIVA with directional microphone was 9.0 dB, which was 1.0 dB better than that with omnidirectional microphones.

- Future work:
  - Listening tests to verify the proposed system.
  - Prototyping the real-time system.

Thank you for your attention !!