Low-Latency Real-Time Blind Source Separation for Hearing Aids Based on Time-Domain Implementation of Online Independent Vector Analysis with Truncation of Non-causal Components

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

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

RION CO., LTD.

NII 大学共同利用機関法人 情報・システム研究機構 国立情報学研究所 National Institute of Informatics

9 March, 2017

Outline of presentation

I. Introduction

II. Low-latency scheme for real-time BSS

- Time-domain implementation
- Truncation of non-causal components

III. Causality of demixing impulse response

IV. Evaluation

V. Demonstration (video)

VI. Conclusion

Background

- Hearing-impaired listeners find it difficult to understand speech in noisy environments such as crowded restaurant.

Crowded restaurant

- Speech
- Music
- Clatter of dishes

⋮

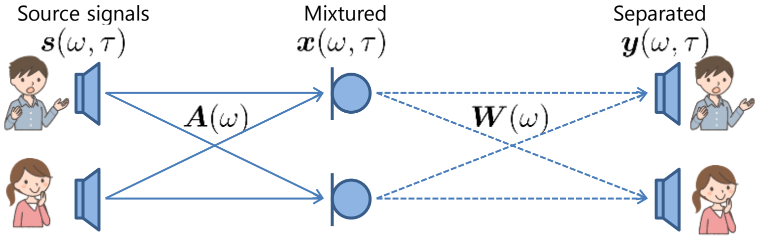
- In these situations, it is difficult to focus a desired sound.
- Unfortunately, current hearing aids are often ineffective in these situations.

The purpose of this study:

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

Blind source separation technique

- Blind source separation (BSS)
  - An effective technique to extract a desired source using only the information of the mixed signals observed by multi microphones.



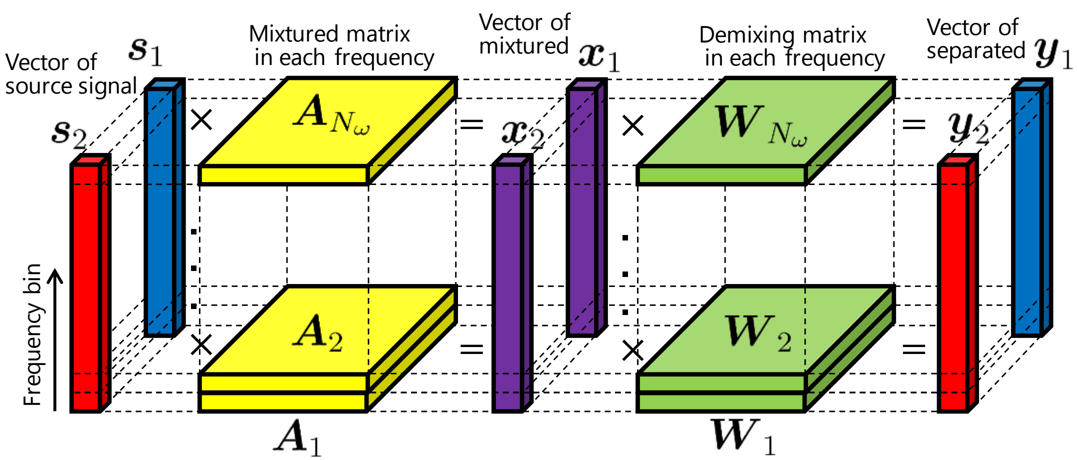
  - A Voice Activity Detection or prior information of a target source are not required. (Beamforming technique typically requires these information.)
  - For convolutive mixtures, independent vector analysis (IVA) [Kim2006, Hiroe2006] in the frequency domain have been developed as a standard technique.
  - There is a state-of-the-art approach for the IVA: Auxiliary-function-based IVA [Ono2011]

Auxiliary-function-based IVA (Ono 2011)

- Auxiliary-function-based Independent Vector Analysis (AuxIVA)
  - One of the frequency-domain approach for convolutive BSS.
  - Fast convergence speed.
  - Low calculation cost.
  - Permutation ambiguity problem is not required.
  - Online approach has been proposed. [Taniguchi2013]

We believe that the AuxIVA algorithm is suitable for the hearing aid system.

Algorithm of online AuxIVA



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

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

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)

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

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

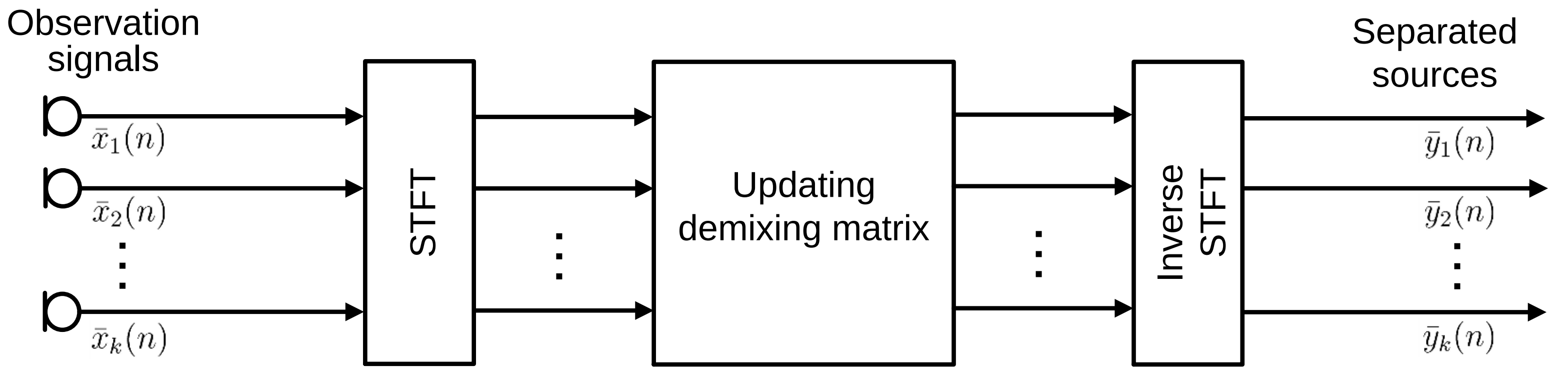
$$W(\omega;\tau) = W(\omega;\tau-1).$$

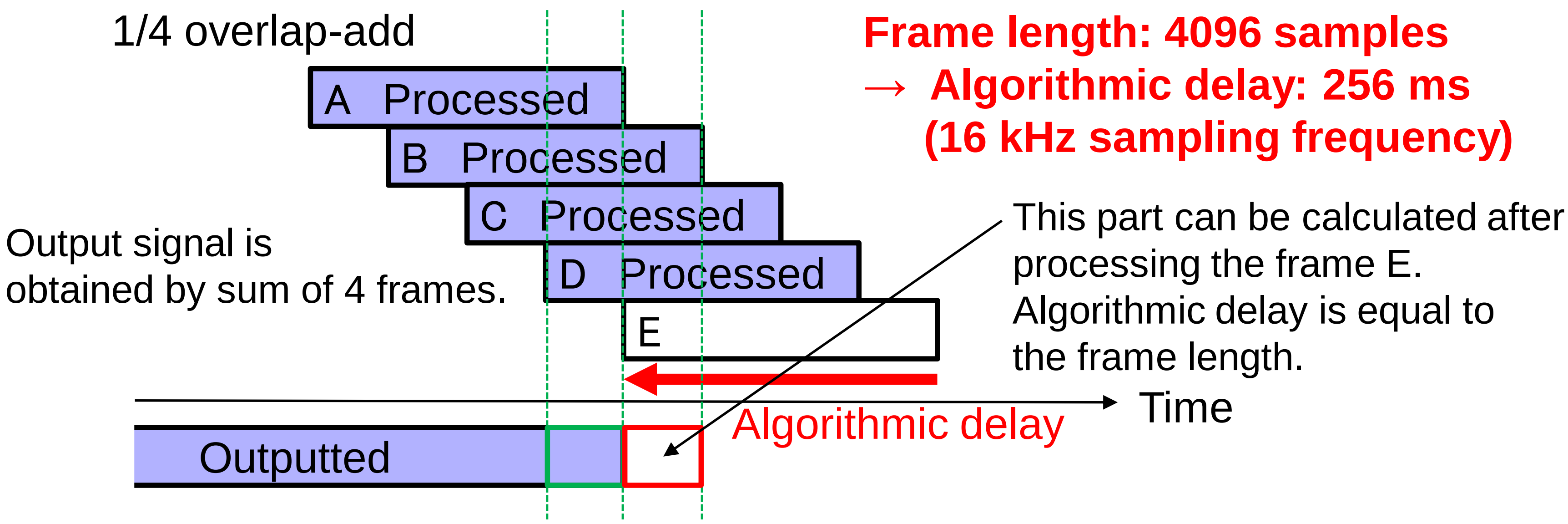
$$\boldsymbol{w}_k(\omega;\tau) \leftarrow (W(\omega;\tau)V_k(\omega;\tau))^{-1}\boldsymbol{e}_k,$$

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

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)

Image of the algorithmic delay for frequency-domain BSS

Problems caused by the algorithmic delay

For the frequency-domain BSS (including the AuxIVA), algorithmic delay at least one frame length is necessary for frame analysis.

- The algorithmic delay becomes 256 ms when the frame length is 4096 samples. (*Fs* = 16 kHz)
- Such a large delay causes various problems in a hearing aid system such as
  - Difficulty in speaking due to the delayed auditory feedback effect.
  - Sense of discomfort due to the loss of lip synchronization.

Reducing the latency (< 10ms) is highly important.

Proposed low-latency scheme

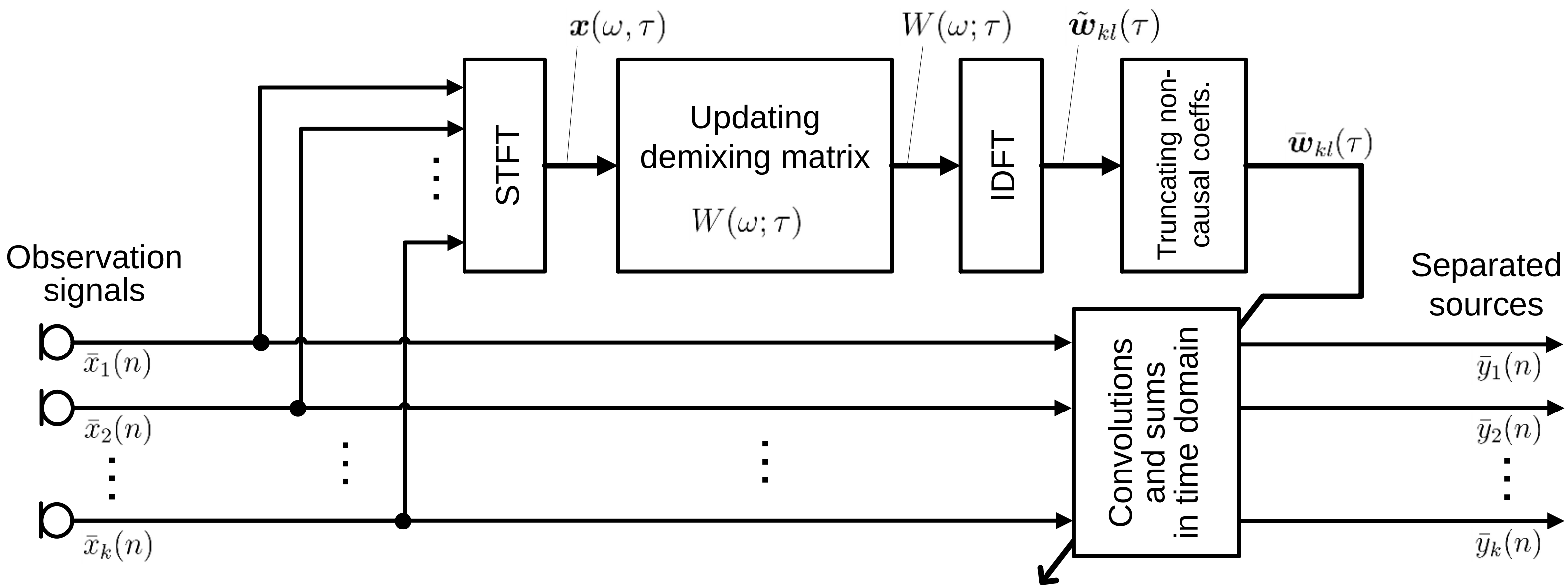
- To solve the problem of an inherent delay..

- We propose a real-time BSS algorithm with low-latency based on online AuxIVA for hearing aids.

i. Time-domain implementation

ii. Truncation of non-causal components

i. Time-domain implementation



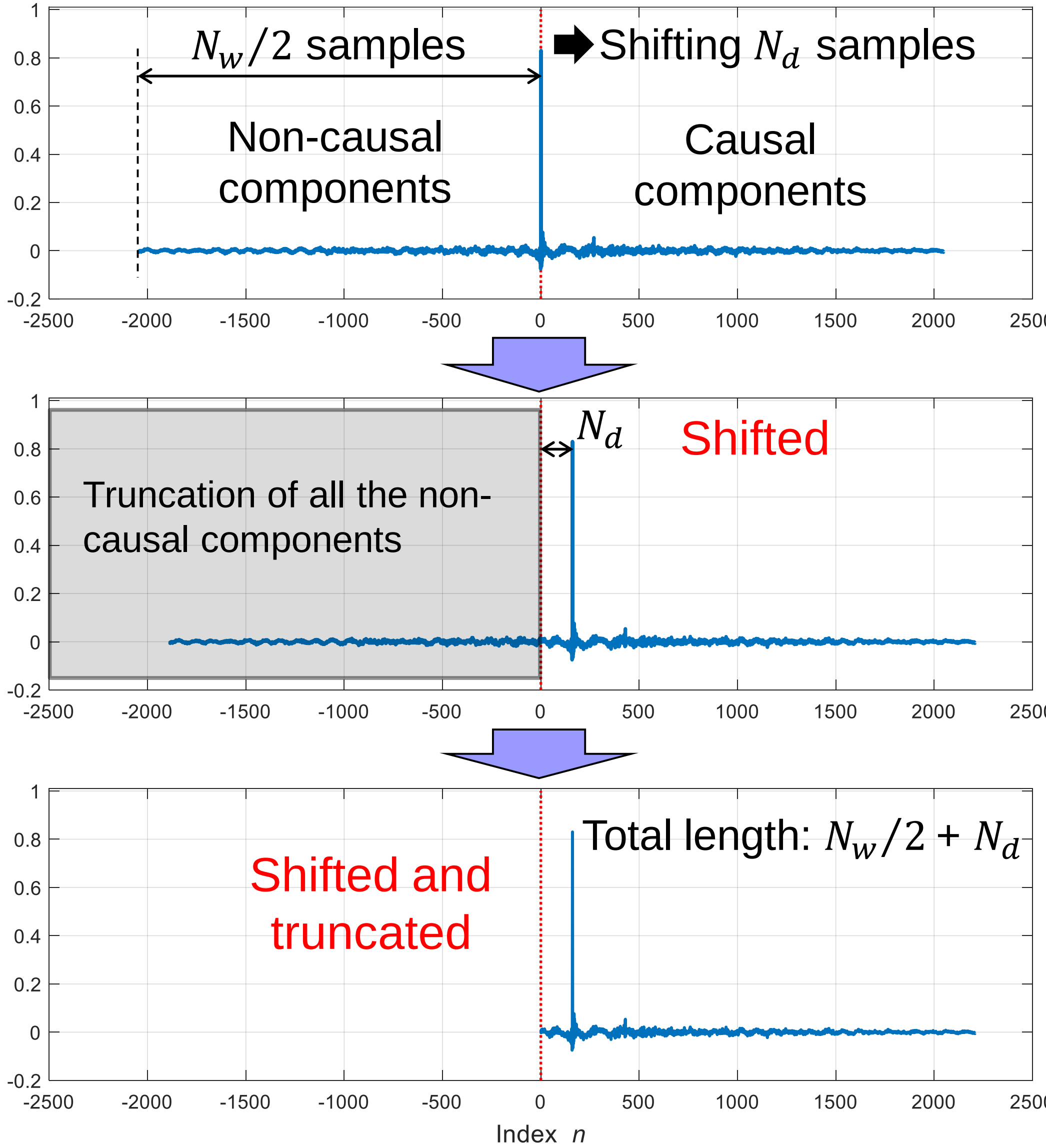
Block diagram of a time-domain implementation of the BSS

- Consisting of two paths
  1) For updating the demixing matrices in the frequency domain
  2) For separating the sources using FIR filters in the time domain
- This structure can shorten the algorithmic delay to a half of the frame length

ii. Truncation of non-causal components

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

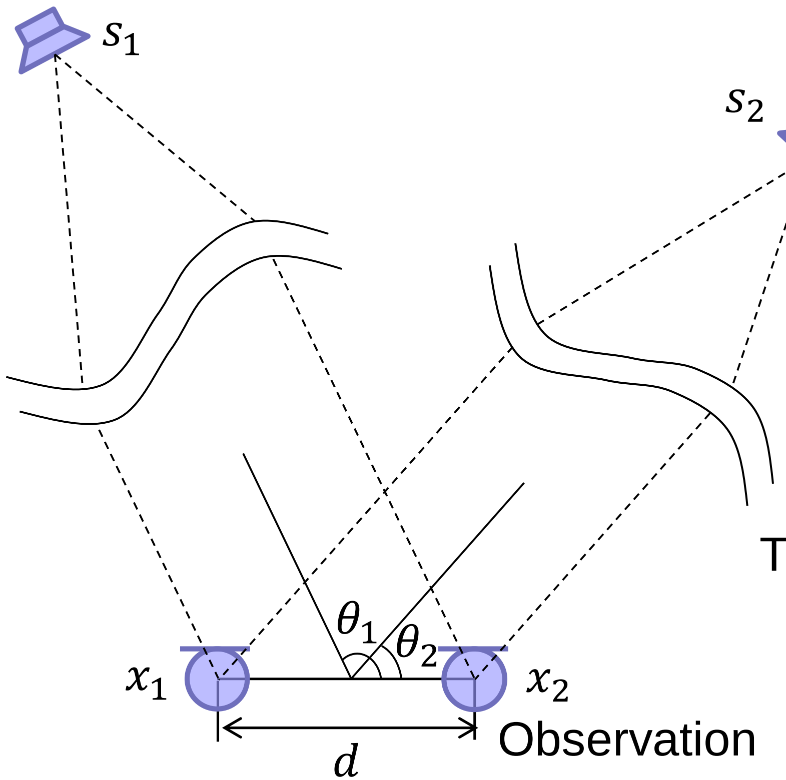
Shifted and truncated demixing FIR filter coefficients $\overline{w}_{kl}(n;\tau)$

Algorithmic Delay N_d samples → 10 ms

Causality of demixing impulse response

- Investigating the causality of impulse responses of ideal separation filters

Source



Simple model consisting of two sources and two observations

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

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

The microphone signals are given by:

$$\begin{pmatrix} x_1(\omega,\tau) \\ x_2(\omega,\tau) \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ a_1 e^{-j\omega\tau_1} & a_2 e^{-j\omega\tau_2} \end{pmatrix} \begin{pmatrix} s_1(\omega,\tau) \\ s_2(\omega,\tau) \end{pmatrix} \quad (13)$$

Then, the source signals can be expressed as

$$\begin{pmatrix} s_1(\omega,\tau) \\ s_2(\omega,\tau) \end{pmatrix} = \frac{1}{D} \begin{pmatrix} a_2 e^{-j\omega\tau_2} & -1 \\ -a_1 e^{-j\omega\tau_1} & 1 \end{pmatrix} \begin{pmatrix} x_1(\omega,\tau) \\ x_2(\omega,\tau) \end{pmatrix} \quad (14)$$

$$D = a_2 e^{-j\omega\tau_2} - a_1 e^{-j\omega\tau_1} = -a_1 e^{-j\omega\tau_1}(1 - R_a e^{-j\omega\Delta\tau})$$

$$R_a = a_2/a_1 \qquad \Delta\tau = \tau_2 - \tau_1$$

Causality of demixing impulse response

Demixing matrix W

$$\begin{pmatrix} s_1(\omega,\tau) \\ s_2(\omega,\tau) \end{pmatrix} = \begin{pmatrix} w_{11}(\omega) & w_{12}(\omega) \\ \boxed{w_{21}(\omega)} & w_{22}(\omega) \end{pmatrix} \begin{pmatrix} x_1(\omega,\tau) \\ x_2(\omega,\tau) \end{pmatrix}$$

$$\boxed{w_{21}(\omega)} = \sum_{m=0}^{\infty} \left(R_a e^{-j\omega\Delta\tau}\right)^m \quad (16)$$

$$R_a = a_2/a_1$$
$$\Delta\tau = \tau_2 - \tau_1$$

Inverse Fourier transform

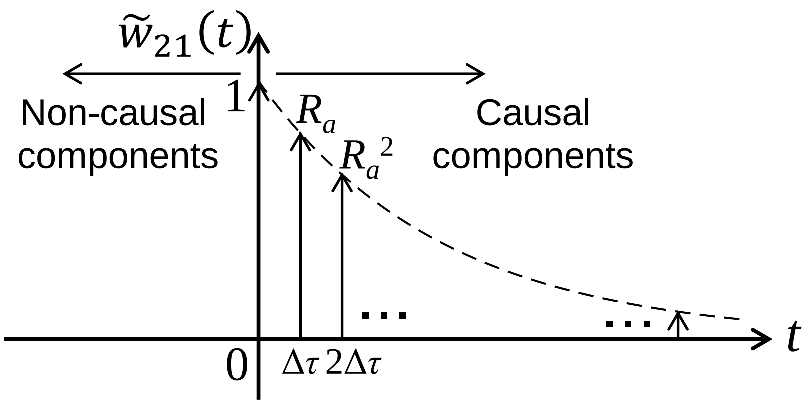
When $R_a < 1$ | When $R_a > 1$

Demixing impulse response

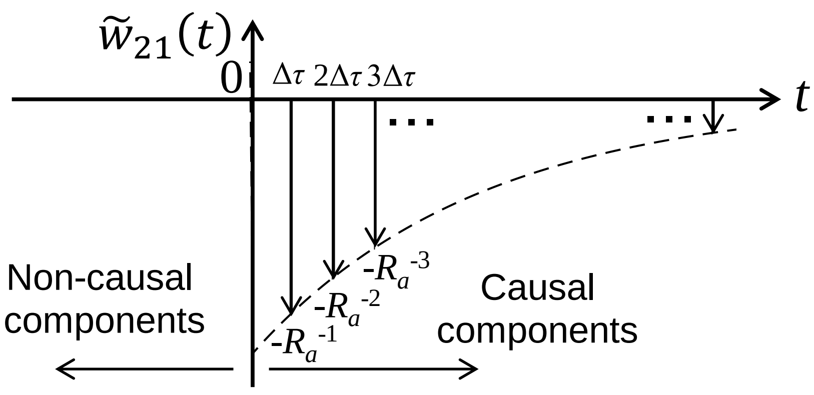
$$\tilde{w}_{21}(t) = \sum_{m=0}^{\infty} R_a^m \delta(t - m\Delta\tau) \quad (17)$$

$$\tilde{w}_{21}(t) = -\sum_{m=0}^{\infty} R_a^{-(m+1)} \delta(t + (m+1)\Delta\tau) \quad (18)$$

When $R_a < 1$ and $\Delta\tau > 0$



When $R_a > 1$ and $\Delta\tau < 0$



The demixing impulse response $\tilde{w}_{21}(t)$ exhibits in only causal-component.
$\tilde{w}_{11}(t)$, $\tilde{w}_{12}(t)$, and $\tilde{w}_{22}(t)$ have same properties.

Causality of demixing impulse response

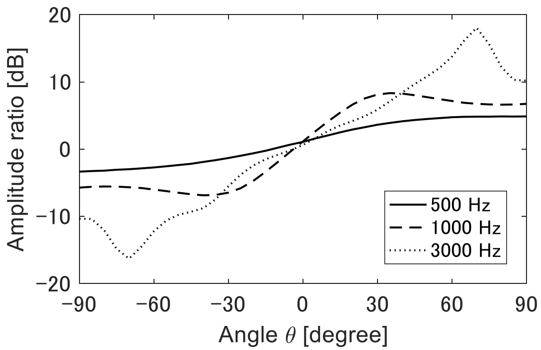
When $R_a < 1$ and $\Delta\tau > 0$

$a(\theta_1) > a(\theta_2)$ and $\tau(\theta_1) < \tau(\theta_2)$

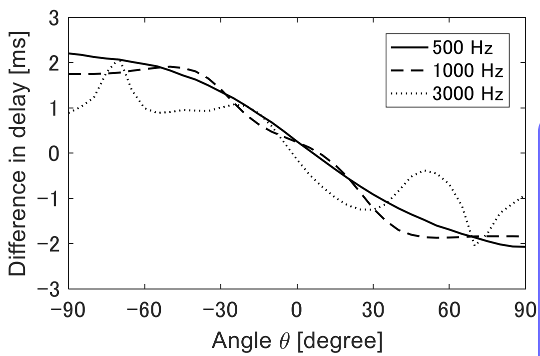
When $R_a > 1$ and $\Delta\tau < 0$

$a(\theta_1) < a(\theta_2)$ and $\tau(\theta_1) > \tau(\theta_2)$

A sufficient condition that the ideal separation filters are causal is
Amplitude ratio $a(\theta)$ → monotonically increasing function of θ
Time difference $\tau(\theta)$ → monotonically decreasing function of θ



(a) Amplitude ratio $a(\theta)$



(b) Time difference $\tau(\theta)$

The condition is roughly satisfactory.

We can expect that the performance degradation due to the truncation will not large in the case of hearing aids

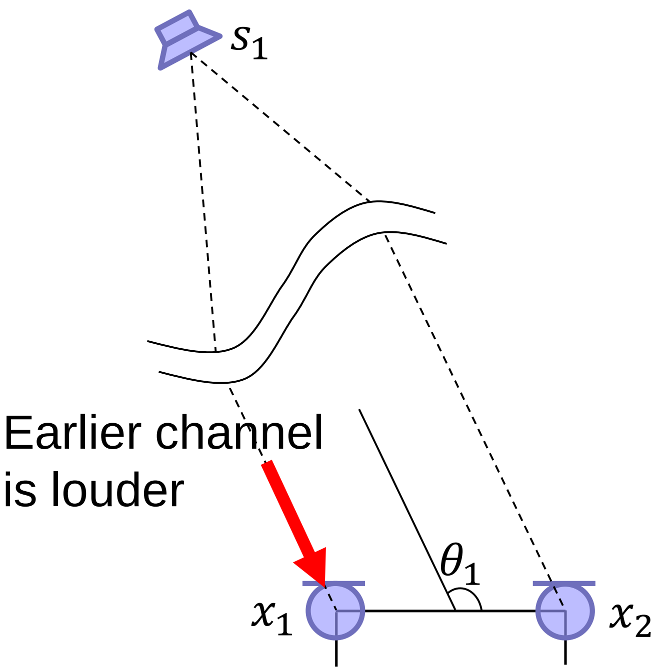
Amplitude ratio and time difference of the right channel relative to the left channel measured using a KEMAR dummy-head microphone

Conclusion of the causality

A sufficient condition that the ideal separation filters are causal is…

Amplitude ratio $a(\theta)$ → monotonically increasing function of θ

Time difference $\tau(\theta)$ → monotonically decreasing function of θ



"An earlier channel is louder."

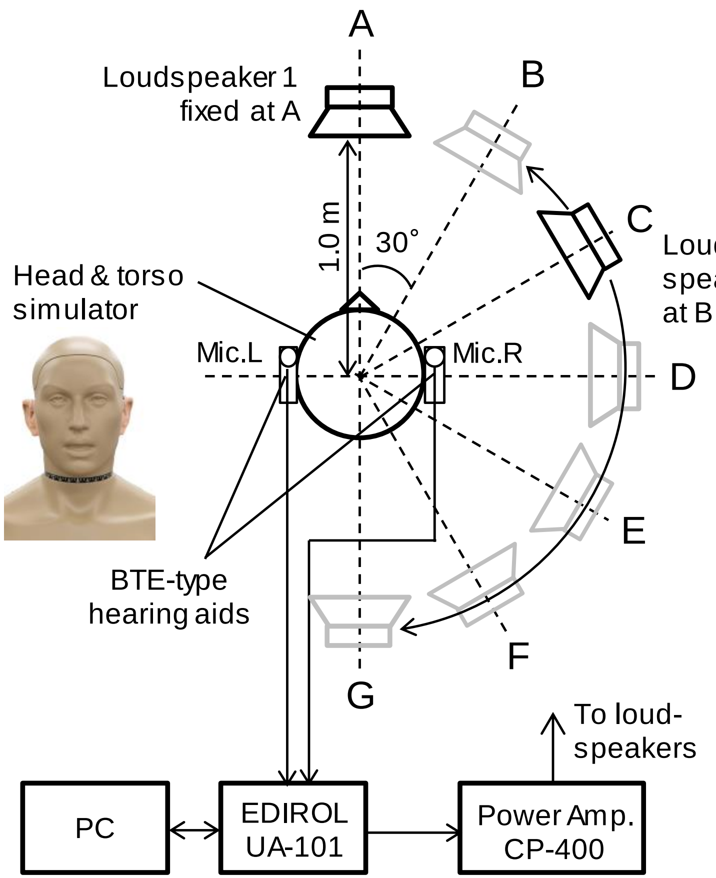
All non-causal components become 0.

We can truncate them.

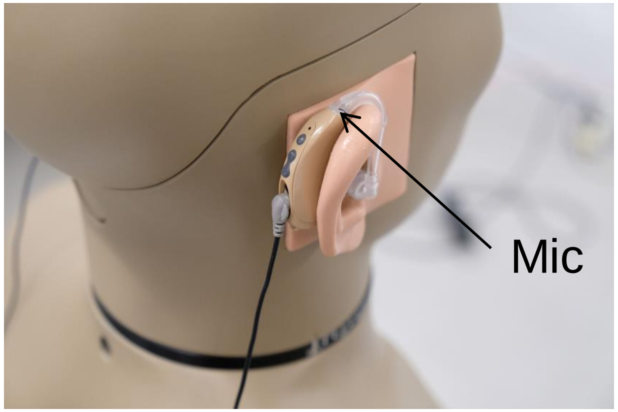
Algorithmic delay can be 0 theoretically.

Evaluation with PC simulation -conditions

KEMAR + BTE-type Hearing aids → 2ch mic. signal



Meeting room (5.0×10.0×2.7 m)



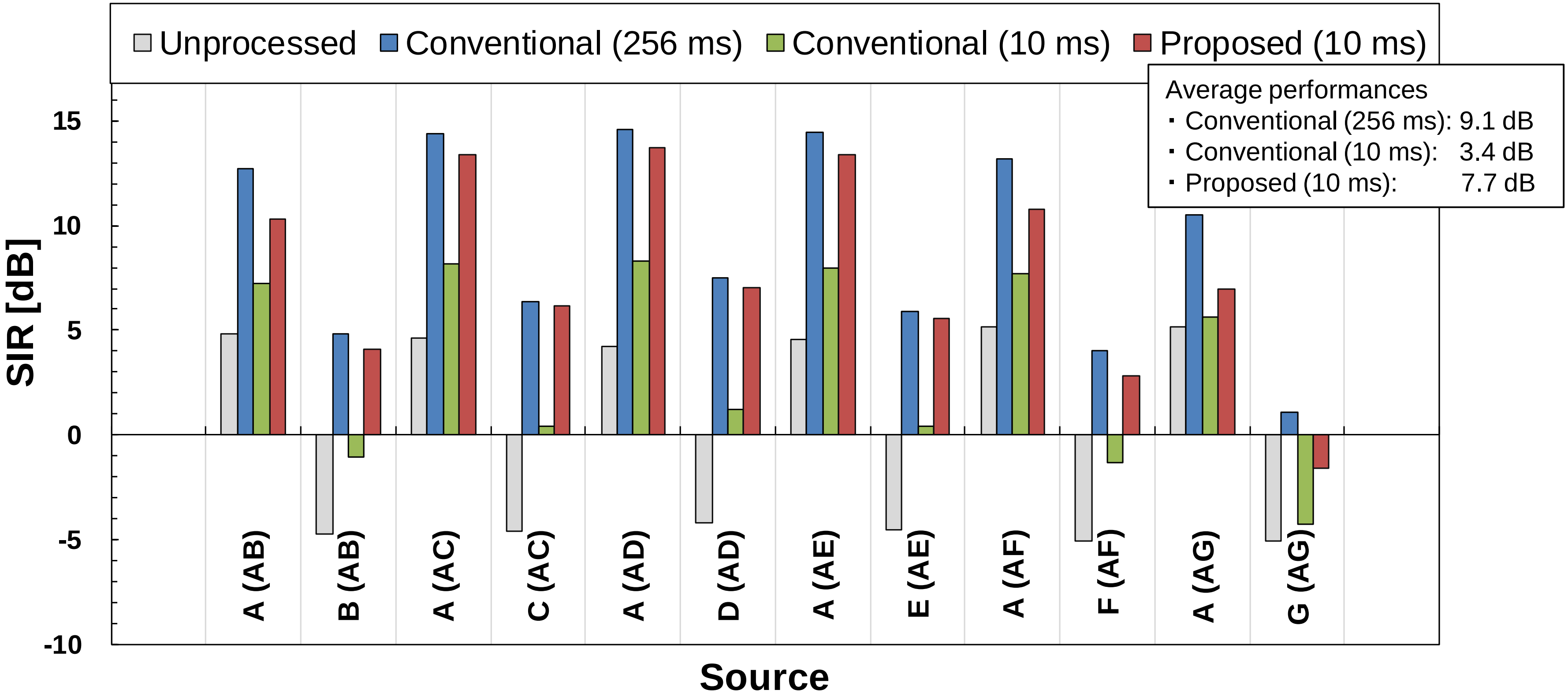
Evaluation with PC simulation -conditions

- Sources: RWCP Japanese News Speech Corpus (Signal length: 30 s × 10 set)
- Microphone spacing: 18cm
- Reverberation time: 650 ms at 500 Hz
- Sampling frequency: 16 kHz
- Evaluation index: Signal-to-interference ratio (SIR)

Algorithm	Conventional 256 ms	Conventional 10 ms	Proposed
Frame length	4096	160	4096
Frame shift	1024	40	1024
Window function	Hanning	Hanning	Hanning
Forgetting factor	0.98	0.98	0.98
Algorithmic delay	256 ms	10 ms	0 to 10 ms (N_d: 0 to 160)

Result 1

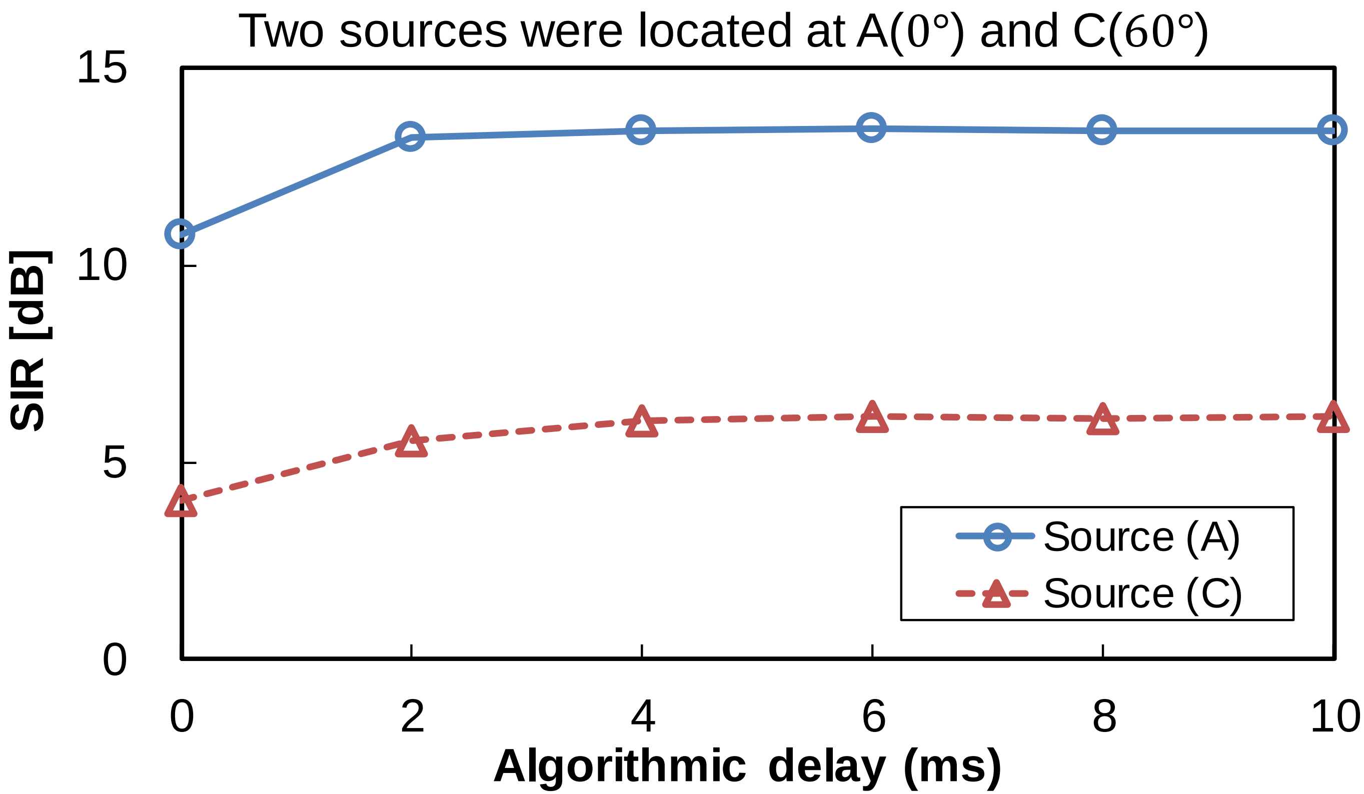
A(AB) denotes source A in a mixture of source A and B



Separation performance of the proposed algorithm compared with that of the conventional algorithm with algorithmic delays of 256 ms and 10 ms

The proposed algorithm shows better separation performance, which was on average only 1.4 dB less than that of the conventional algorithm with an algorithmic delay of 256 ms.

Result 2



Resultant SIRs of the proposed algorithm for algorithmic delays from 0 to 10 ms. (N_d was set from 0 to 160 samples)

Algorithmic delays of 2 ms and above resulted in better performance.

Demonstration video of a real-time system

- A real-time system of proposed low-latency online AuxIVA algorithm implemented on notebook

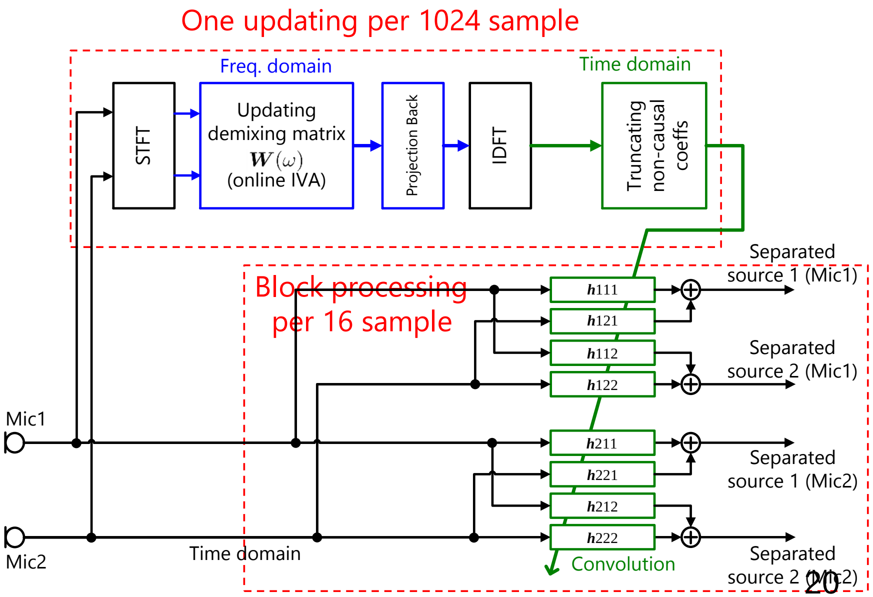
- Windows 8.1 Core i5 2GHz
- Visual C++, Eigen & FFTW library
- Multithread processing
- Sampling frequency: 16 kHz
- Frame length: 4096, Frame shift: 1024

Total latency

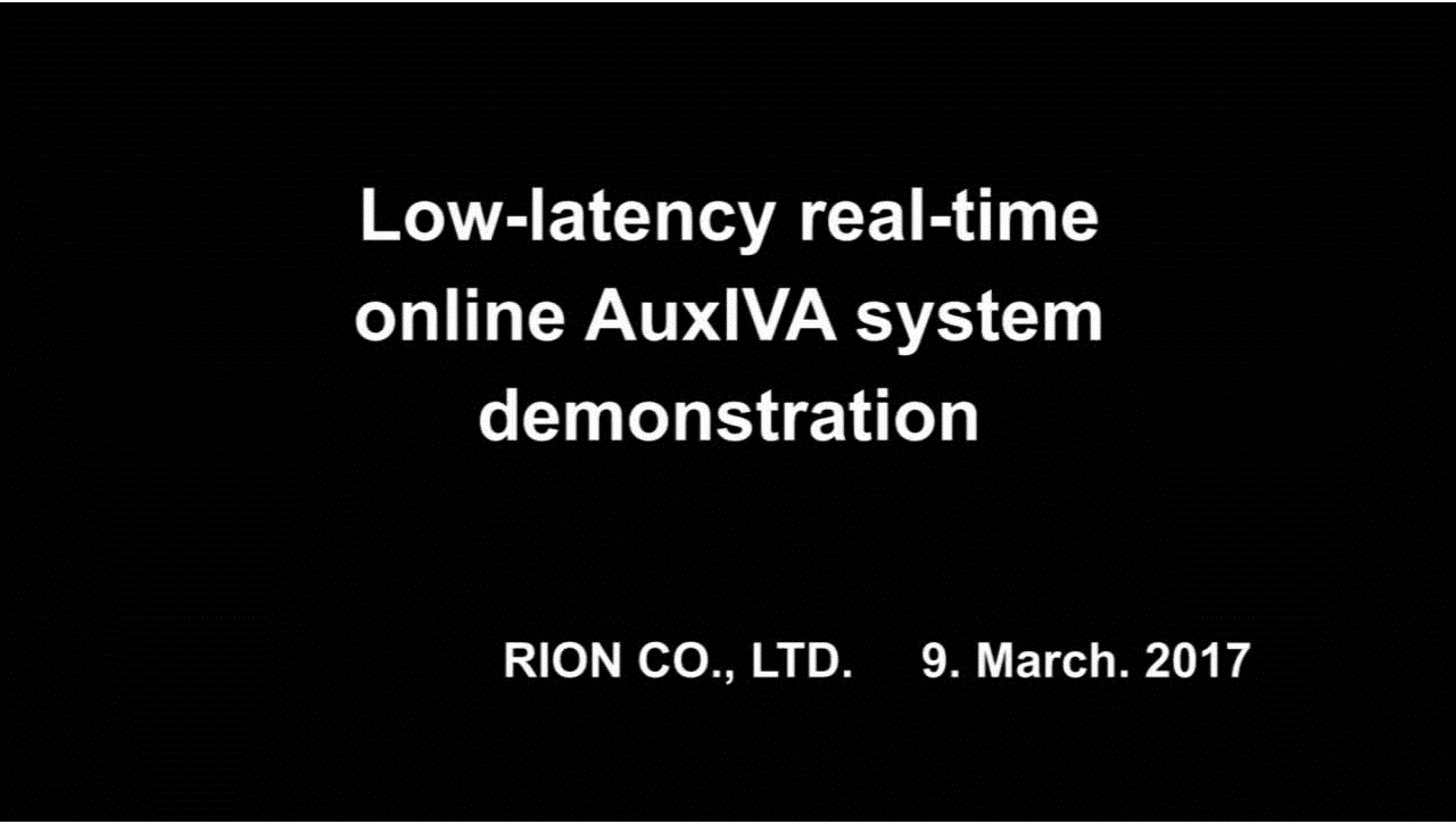
- Algorithmic delay: 10 ms
- Frame delay: 1 ms
- ASIO delay: about 5 ms

About **16 ms**

Audio in the demo video is only output of the real-time system.



Demonstration of a real-time system



Conclusion

- A real-time BSS algorithm with low latency based on online IVA for hearing aids were proposed.
- The proposed algorithm can significantly shorten the algorithmic delay by the time-domain implementation of demixing matrices as FIR filters and the truncation of part of their non-causal components.
- From an analysis of the causality, it is found that the ideal separation filters are causal if an earlier channel is larger.
- The algorithmic delay in the proposed system was within 10 ms and the average SIR was 7.7 dB.
- The proposed algorithm can be used for real-time audio devices such as hearing aids.

Thank you for your attention.