Paper ID: OC121

Speech enhancement using binary estimator selection applied to hearing aids with a remote microphone

Vasudha Sathyapriyan*[1, 2] Michael Syskind Pedersen[1] Mike Brookes[3]
Jan Østergaard[2] Patrick A. Naylor[3] Jesper Jensen[1, 2]

[1]Demant A/S, Denmark
[2]Aalborg University, Denmark
[3]Imperial College London, United Kingdom

24-10-2023

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 956369.

Speech enhancement using binary estimator selection applied to hearing aids with a remote microphone

Vasudha Sathyapriyan*, Michael Syskind Pedersen, Mike Brookes, Jan Østergaard, Patrick A. Naylor, Jesper Jensen

24-10-2023

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 956369.

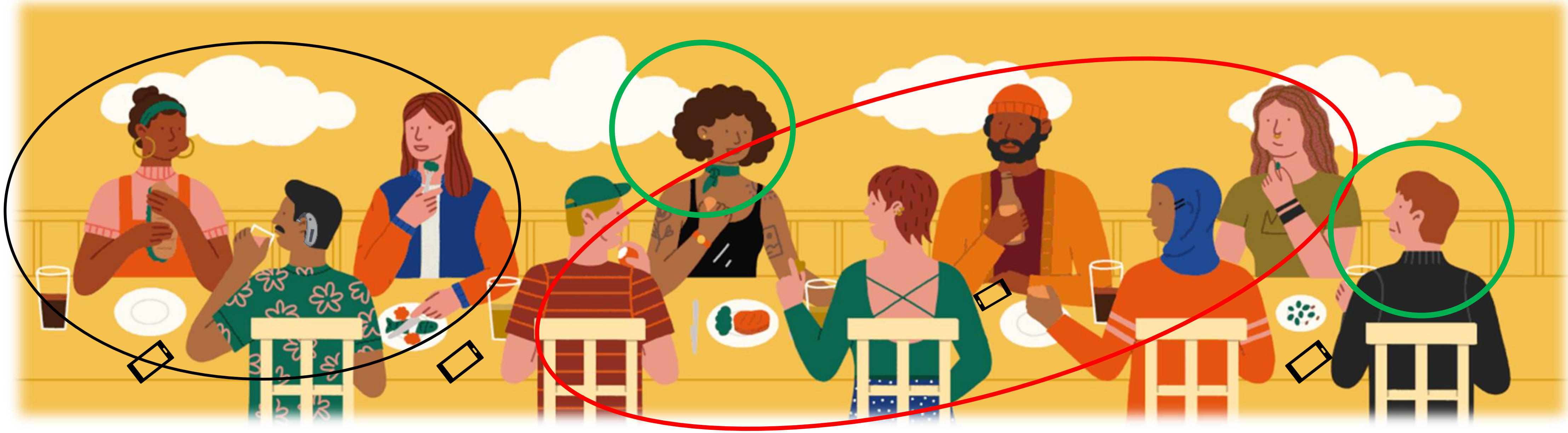
Overview

- Introduction
- Proposed Method
- Conclusion

Overview

- Introduction
  - Motivation
- Proposed Method
  - Signal model
  - Binary Estimator Selection
- Conclusion
  - Results

Motivation

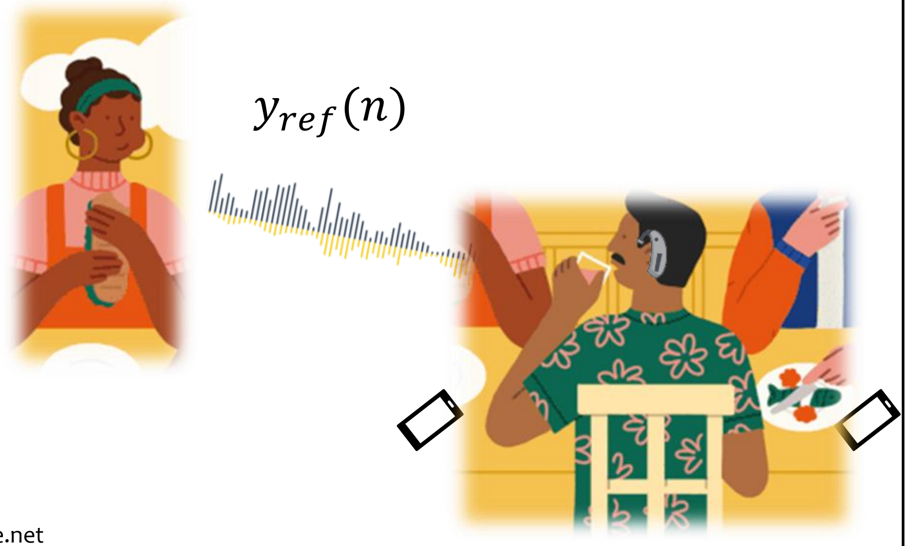


Source: Behance.net

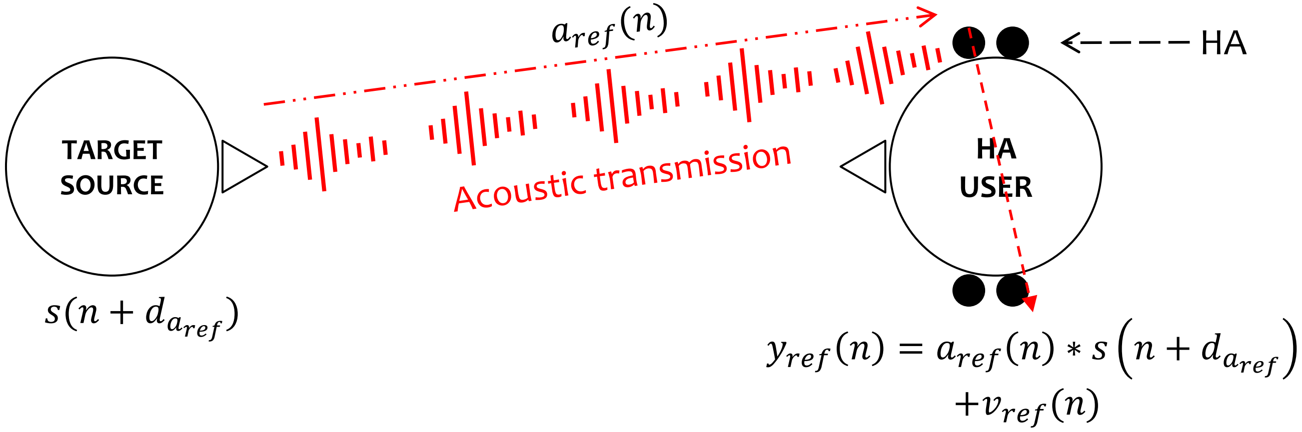
Motivation

Source: Behance.net

Motivation



Source: Behance.net

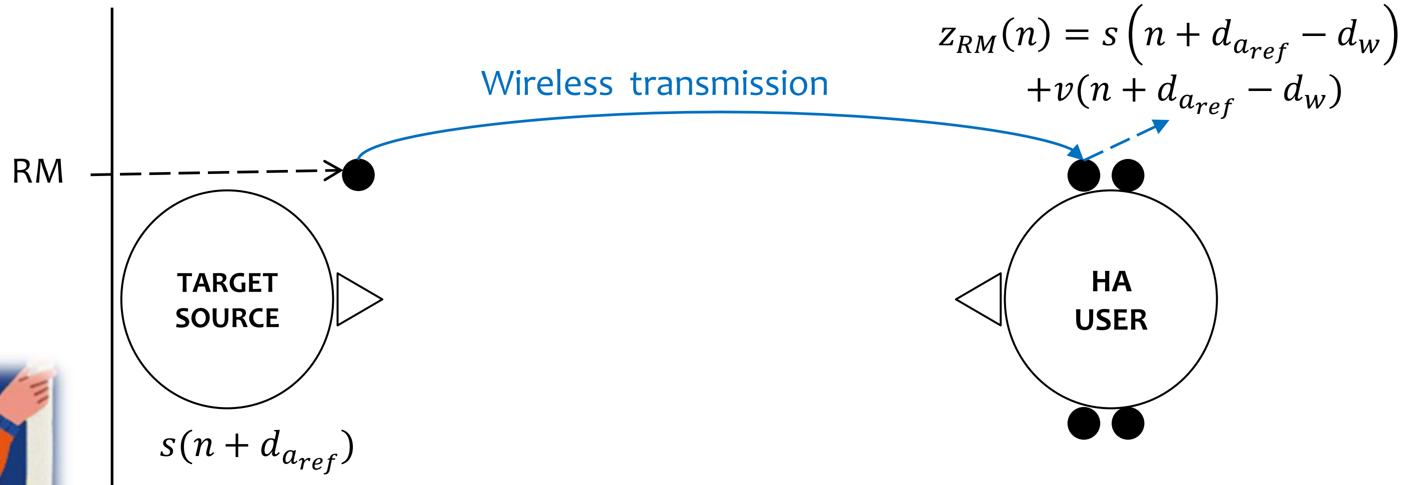


Acoustic transmission delay = $d_{a_{ref}} = \tau_a f_s$

Motivation

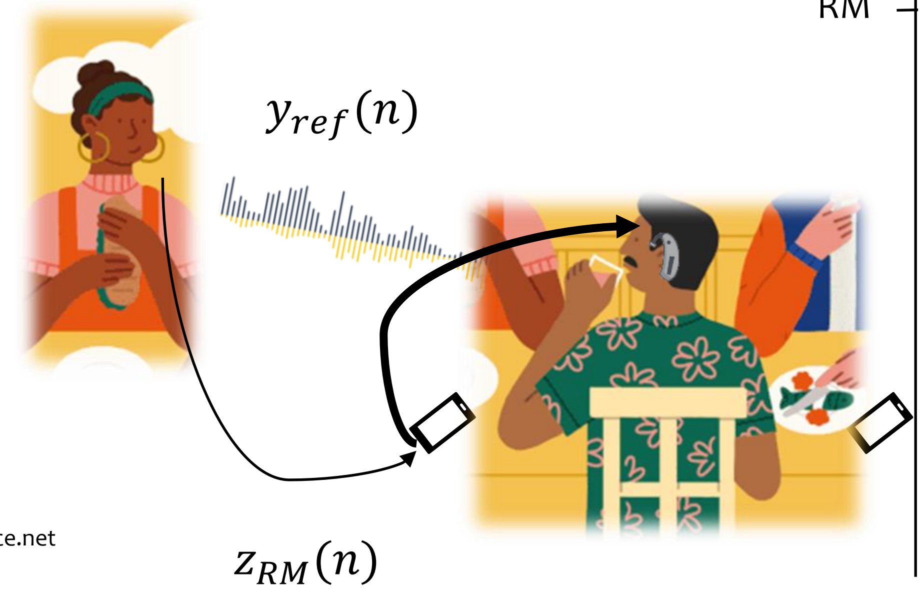
Source: Behance.net

$z_{RM}(n)$

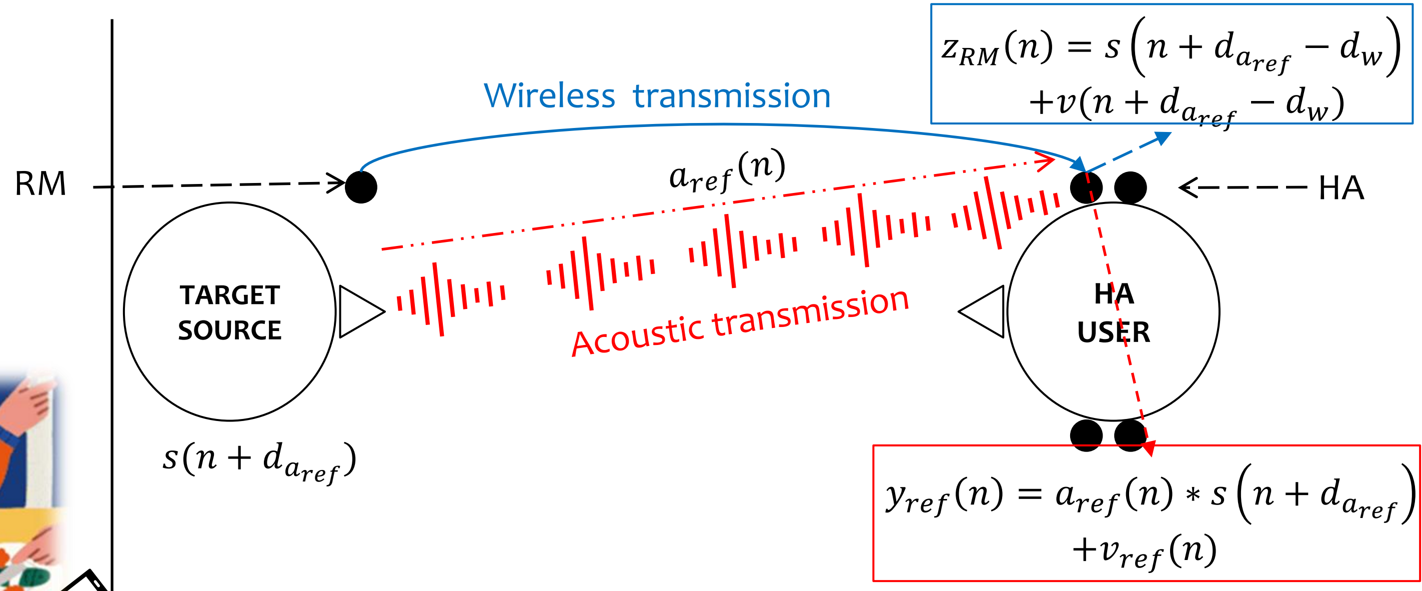


Wireless transmission delay = $d_w = \tau_w f_s$

Motivation



Source: Behance.net



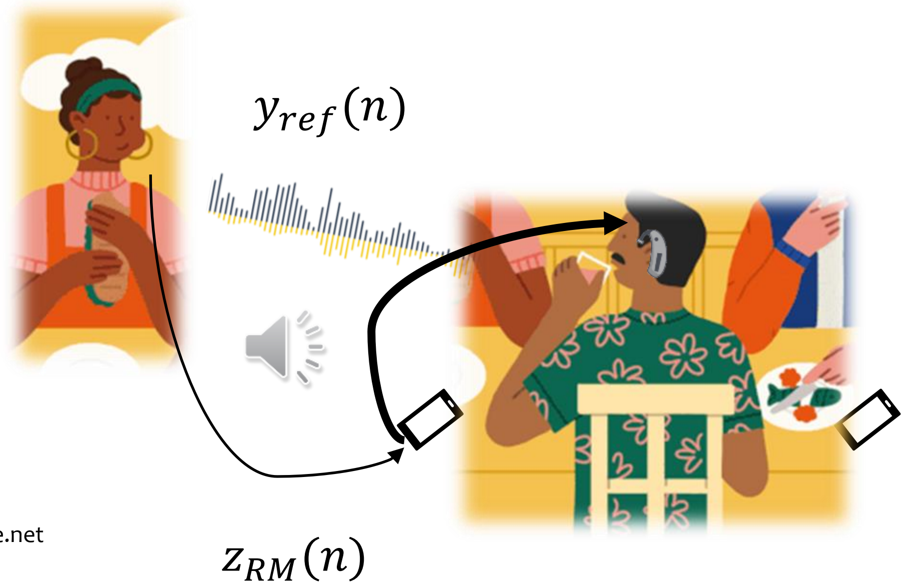
Acoustic transmission delay = $d_{a_{ref}} = \tau_a\, f_s$

Wireless transmission delay = $d_w = \tau_w\, f_s$

TDOA: $\boldsymbol{\tau} = \boldsymbol{\tau_w} - \boldsymbol{\tau_a}$

$$d_\tau = d_w - d_{a_{ref}}$$

Motivation



Source: Behance.net

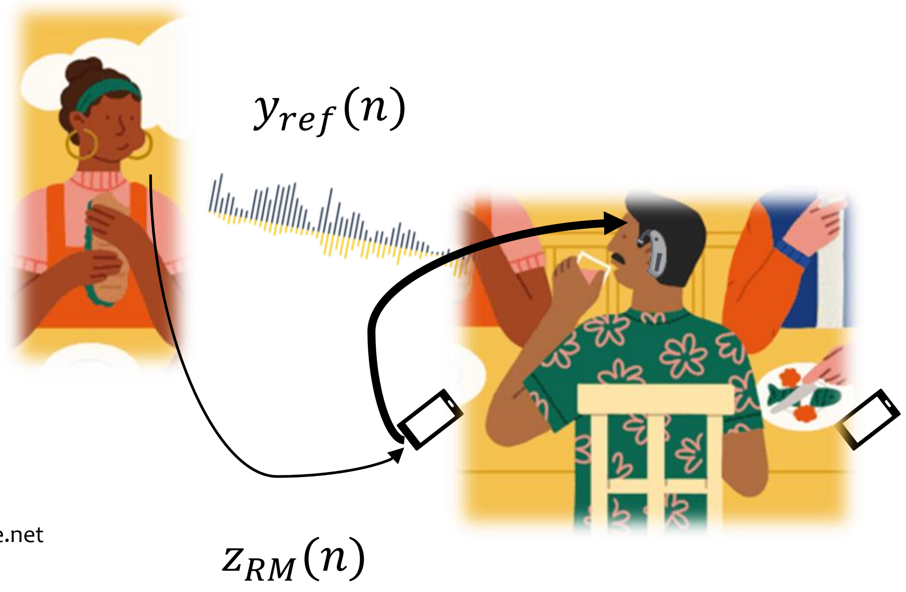
Where τ is the time difference of arrival (TDOA) between the airborne acoustic signal, $y(n)$, and the wirelessly transmitted signal, $y(n - d_\tau)$.

1. $\underbrace{\text{Short TDOAs}}_{1-10\text{ ms}}$ lead to comb-filtering effect [1].
2. $\underbrace{\text{Long TDOAs}}_{>10\text{ ms}}$ lead to echoes [1].

Motivation

1. TDOA, τ, can be positive i.e., $\tau \geq 0$ or negative i.e., $\tau < 0$.
   a) But for $\tau < 0$, the RM signal can be stored in a buffer.
   b) $0 \leq \tau \leq 10\ \mathrm{ms}$, to avoid undesirable perceptual audio effects [2].

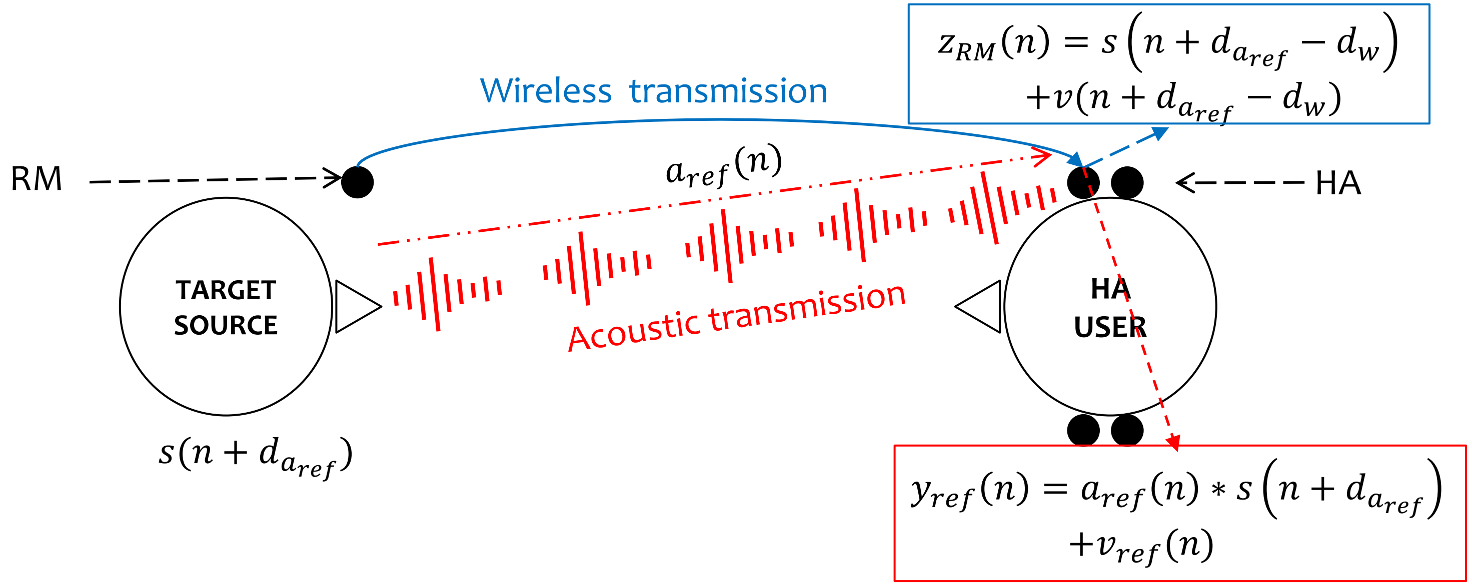
2. **Direct RM signal playback can be detrimental in the presence of TDOAs.**



Source: Behance.net

Use a combination of HA microphone and RM signals !

Signal Model



Acoustic transmission delay = $d_{a_{ref}} = \tau_a \, f_s$

Wireless transmission delay = $d_w = \tau_w \, f_s$

TDOA: $\boldsymbol{\tau} = \boldsymbol{\tau_w} - \boldsymbol{\tau_a}$

Signals available for processing :

From the HA microphones:

$$y_{ref}(n) = x_{ref}(n) + v_{ref}(n)$$

$$\underline{y}_{HA}(n) = [y_1(n) \quad \dots \quad y_M(n)]^T$$

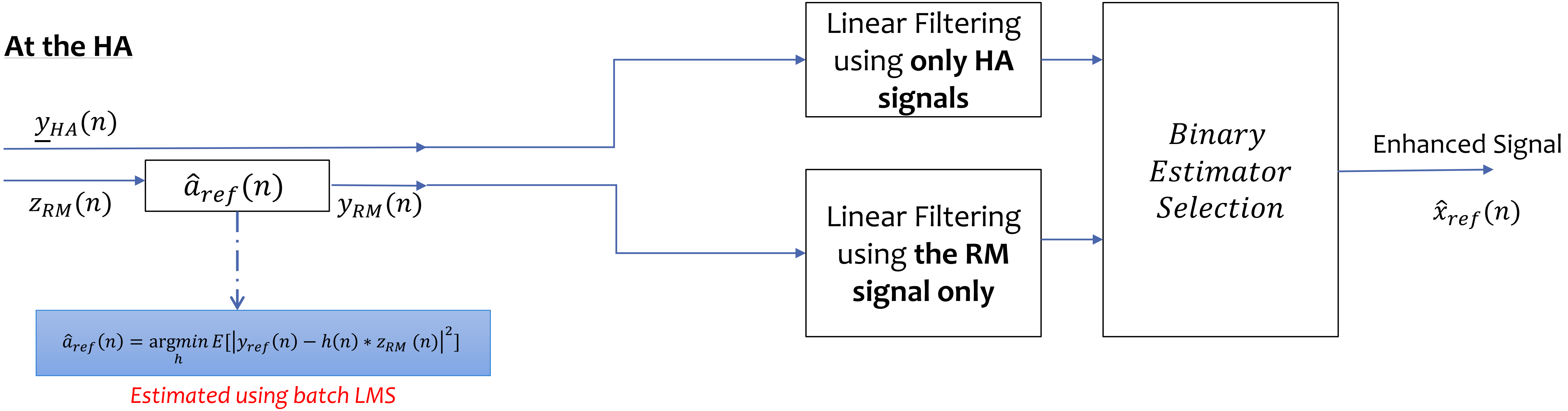
From the RMs (received at the HA unit):

$$y_{RM}(n) = \hat{a}_{ref}(n) * z_{RM}(n)$$

Now, TDOA $\{y_{ref}(n), y_{RM}(n)\}$: $\boldsymbol{\tau} = \boldsymbol{\tau_w}$

Binary Estimator Selection Algorithm

At the HA



Signal Model

Signals available for processing (in STFT domain):

From the HA microphones:

$$\underline{\boldsymbol{y}}_{HA}(k,l) = \begin{bmatrix} Y_{HA,1}(k,l) \\ Y_{HA,2}(k,l) \\ \vdots \\ Y_{HA,M}(k,l) \end{bmatrix} \in \mathbb{C}^{M\times 1}$$

From the RMs (received at the HA unit):

$$\underline{\boldsymbol{y}}_{RM}(k,l) = \begin{bmatrix} Y_{RM}(k,l) \\ Y_{RM}(k,l-1) \\ \vdots \\ Y_{RM}(k,l-L+1) \end{bmatrix} \in \mathbb{C}^{L\times 1}$$

$Y_{HA,m}(k,l)$ *– STFT coefficient of the m^th^ noisy HA microphone signal,* $y_m(n)$.

$Y_{RM}(k,l))$ *– STFT coefficient of the noisy RM signal received at the HA,* $y_{RM}(n)$.

$\underline{\boldsymbol{y}}_{HA}(k,l)$ *– multi-channel HA STFT coefficient vector.*

$\underline{\boldsymbol{y}}_{RM}(k,l)$ *– multi-frame RM STFT coefficient vector.*

M *– number of HA microphones*

L– number of past-frames .

Signal Model

Signals available for processing (in STFT domain):

From the HA microphones:

$$\underline{\boldsymbol{y}}_{HA}(k,l) = \begin{bmatrix} Y_{HA,1}(k,l) \\ Y_{HA,2}(k,l-1) \\ \vdots \\ Y_{HA,M}(k,l-L+1) \end{bmatrix} \in \mathbb{C}^{M\times 1}$$

From the RMs (received at the HA unit):

$$\underline{\boldsymbol{y}}_{RM}(k,l) = \begin{bmatrix} Y_{RM}(k,l) \\ Y_{RM}(k,l-1) \\ \vdots \\ Y_{RM}(k,l-L+1) \end{bmatrix} \in \mathbb{C}^{L\times 1}$$

Rank-1 Multi-channel Wiener Filter[3]:

$$\underline{\boldsymbol{w}}_{MWF} = \underset{w}{argmin}\ E[||\underline{\boldsymbol{w}}^H \underline{\boldsymbol{y}}_{HA} - X_{ref}||^2]$$

$$\hat{X}_{MWF} = \underline{\boldsymbol{w}}_{MWF}{}^H \underline{\boldsymbol{y}}_{HA}$$

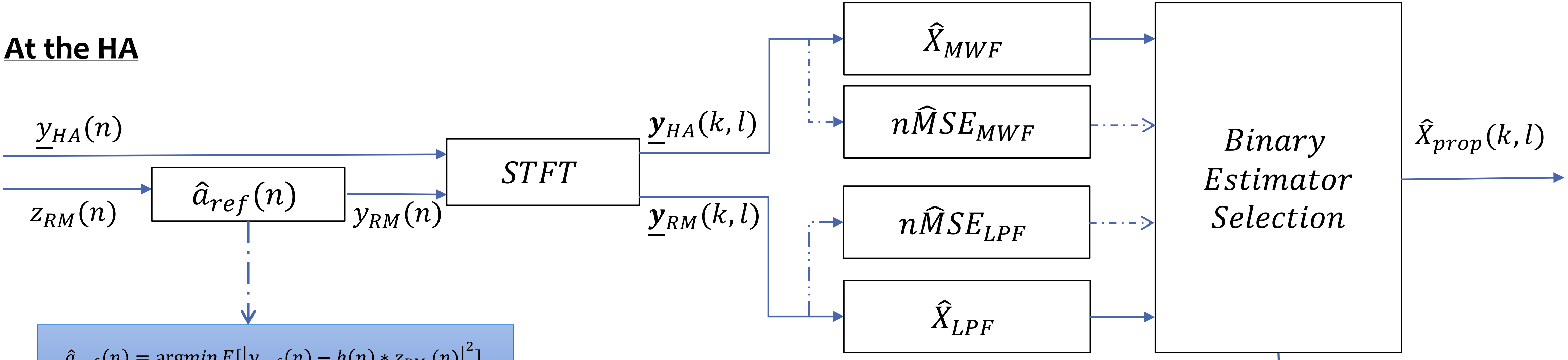
Single Channel Multi-Frame Wiener Filter [4], [5]:

$$\underline{\boldsymbol{w}}_{LPF} = \underset{w}{argmin}\ E[||\underline{\boldsymbol{w}}^H \underline{\boldsymbol{y}}_{RM} - X_{ref}||^2]$$

$$\hat{X}_{LPF} = \underline{\boldsymbol{w}}_{LPF}{}^H \underline{\boldsymbol{y}}_{RM}$$

Binary Estimator Selection Algorithm

At the HA



$$\hat{a}_{ref}(n) = \underset{h}{\operatorname{argmin}}\, E\left[\left|y_{ref}(n) - h(n) * z_{RM}(n)\right|^2\right]$$

Estimated using batch LMS

$$\hat{X}_{prop}(k,l) = \hat{G}.\hat{X}_{MWF}(k,l) + \left(1 - \hat{G}\right).\hat{X}_{LPF}(k,l)$$

$$\hat{G}(k,l) = \begin{cases} 1, & \dfrac{nMSE_{MWF}(k,l)}{nMSE_{LPF}(k,l)} < 1 \\ 0, & otherwise \end{cases}$$

Results

Simulation Parameters:

1. ***Speech signal:*** CSTR VCTK corpus [6] .
2. ***HAHRIR:*** *Monaural Left (M=2), anechoic hearing aid head-related impulse responses from [7].*
3. ***Noisy Types:*** *isotropic SSN, cafeteria noise recordings [7].*
4. ***SNR at the HA :*** *0 dB.*
5. ***SNR at the RM****: 20 dB.*
6. ***TDOAs*** *: 0-40 ms.*
7. ***Performance metrics:*** ESTOI *(speech intelligibility), MSE (speech quality).*

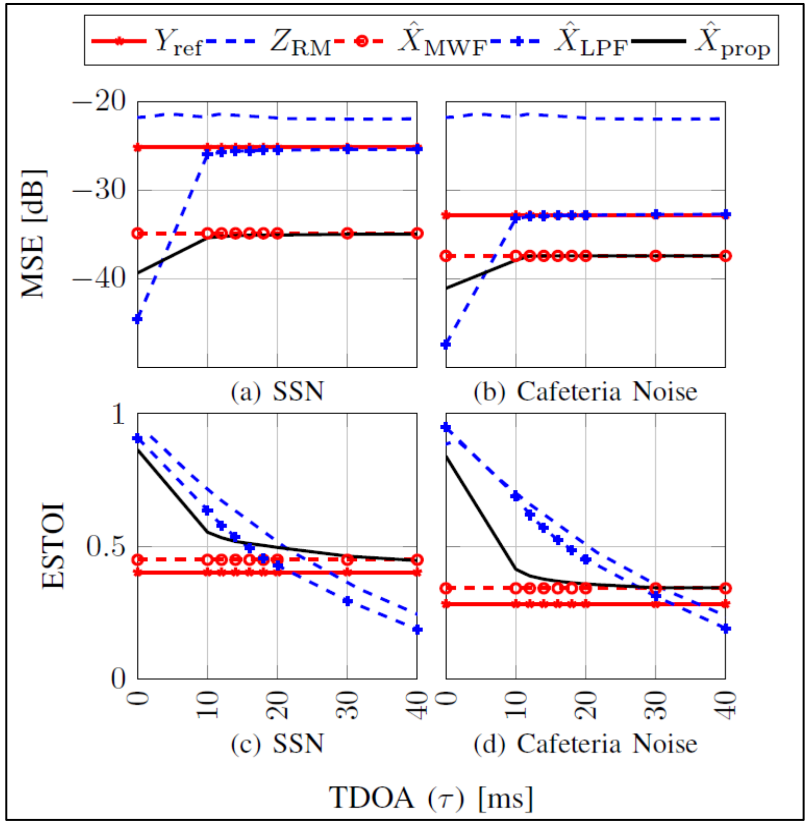


Figure 1 :Performance in terms of (a), (b) MSE and (c), (d) ESTOI for different TDOAs (τ), for SSN and Cafeteria Noise at 0 dB SNR.

Audio Examples

Speech enhancement using binary estimator selection applied to hearing aids with a remote microphone

The demos consist of monoaural left signals for different noise and SNRs, headphones are recommended for listening.

- The monaural left signals are generated using measured HRIRs from the Kayser Database [3]
- The speech signals are taken from the VCTK Corpus [2]

This page was generated using **trackswitch.js** in [1].

Isotropic Speech Shaped Noise

-10 dB	-5 dB	0 dB	5 dB

Source: https://vsathyapriyan.github.io/bes_complex/

Conclusion

- Remote microphone signal introduces **undesirable perceptual effects for $\tau > 10$ ms** .
- Binary estimator selection (BES) algorithm **uses both the HA and RM signals,** in the presence of a TDOA.
- Binary estimator selection (BES) algorithm**, choses between** an MMSE filter on the HA microphone signals, **MWF** and an MMSE filter on the RM signals, **LPF,** in the presence of TDOAs.
- The **BES algorithm provides a benefit in estimated speech intelligibility for $0 \leq \tau \leq 30$ ms,** over the MWF, that uses only the HA microphone signals.

References

[1] Stone, Michael A., Brian C. J. Moore, Katrin Meisenbacher, and Ralph P. Derleth. "Tolerable Hearing Aid Delays. V. Estimation of Limits for Open Canal Fittings." *Ear & Hearing* 29, no. 4 (August 2008): 601–17. https://doi.org/10.1097/AUD.0b013e3181734ef2.

[2] L. Bramsløw, "Preferred signal path delay and high-pass cut-off in open fittings," *International Journal of Audiology*, vol. 49, no. 9, pp. 634–644, 2010.

[3] K. Uwe Simmer, Joerg Bitzer, and Claude Marro, *Post-Filtering Techniques*, pp. 39–60, Springer Berlin Heidelberg, Berlin, Heidelberg, 2001.

[4] Benesty, Jacob, and Yiteng Huang. "A Single-Channel Noise Reduction MVDR Filter." In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 273–76. Prague, Czech Republic: IEEE, 2011. https://doi.org/10.1109/ICASSP.2011.5946393.

[5] V. Sathyapriyan, M. S. Pedersen, J. Østergaard, M. Brookes, P. A. Naylor, and J. Jensen, "A linear MMSE filter using delayed remote microphone signals for speech enhancement in hearing aid applications," in *International Workshop on Acoustic Signal Enhancement (IWAENC)*, 2022, pp. 1–5.

[6] J. Yamagishi, C. Veaux, and K. MacDonald, "CSTR VCTK corpus: English multi-speaker corpus for CSTR voice cloning toolkit," *University of Edinburgh. The Centre for Speech Technology Research (CSTR)*, 2017.

[7] H. Kayser, S. D. Ewert, J. Anemüller, T. Rohdenburg, V. Hohmann, and B. Kollmeier, "Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses," EURASIP Journal on advances in signal processing, vol. 2009, pp. 1–10, 2009.

Questions?

Speech enhancement using binary estimator selection applied to hearing aids with a remote microphone

Vasudha Sathyapriyan*, Michael Syskind Pedersen, Mike Brookes,
Jan Østergaard, Patrick A. Naylor, Jesper Jensen

24-10-2023

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 956369.