Localization Cues Preservation in Hearing Aids by Combining Noise Reduction and Dynamic Range Compression

Adrien Llave, Simon Leglaive, Renaud Séguier

7 December 2020

FAST team, IETR UMR CNRS 6164, CentraleSupélec

Introduction

Dynamic range compression (DRC)

Aim: amplify the soft sound without reaching the pain threshold

Dynamic Range Compression (DRC):

- Attenuate the output if the level exceeds a given threshold
- Number of frequency bands : 8 $\rightarrow$ 32

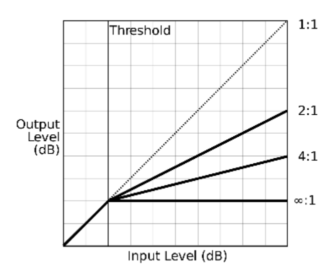


Figure 1: DRC broken-stick function

DRC - issue

Issue: left and right DRC gain are different:

- Hearing loss compensation distorts the localization cues
- Localization performance decreasing[1]
- Speech in noise understanding performance decreasing[2]

[1] [Hassager et al., 2017b, Wiggins and Seeber, 2011, Van den Bogaert et al., 2006]
[2] [Schwartz and Shinn-Cunningham, 2013]

SNR-aware DRC[3]

- **Aims** :
  - reducing the speech dynamic range
  - preserving the original noise dynamic range
  - improving the output SNR
- **Idea**: fast compression for the speech period, slowly otherwise

Advantage

improve the localization in presence of reverberation [Hassager et al., 2017a]

Drawbacks

The attenuation of the noise only period depends on the previous speech content!

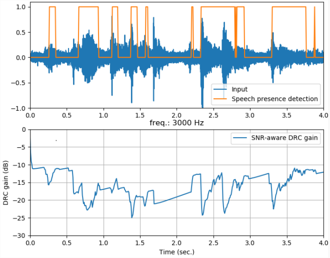


Figure 2: Noisy speech and speech presence detection (top), DRC gain at 3 kHz of the SNR-aware DRC

3[May et al., 2018]

Our approach

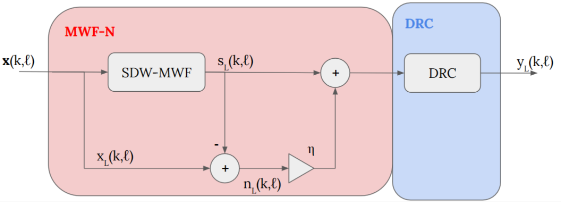


Figure 3: Standard association of noise reduction and DRC.

Our approach

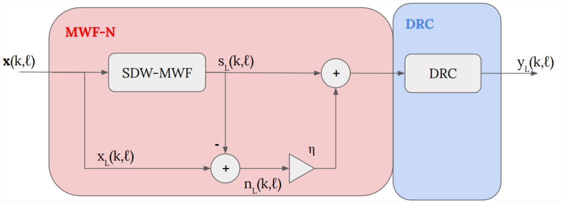


Figure 3: Standard association of noise reduction and DRC.

Objectives :

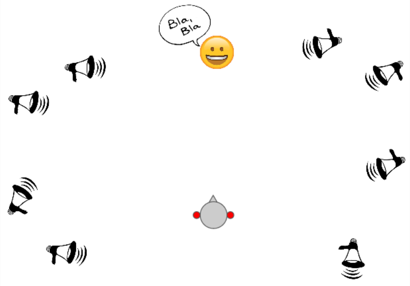
- Reducing the speech dynamic range
- Preserving the noise dynamic range
- Improving the SNR
- Preserving the localization cues of both components

Idea

Merge noise reduction and DRC

Data Model

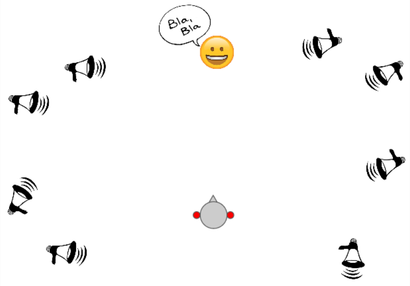
Acoustical scenario



At each microphone $x_m(t)$:

- one speech source (plane wave), $s(t)$, filtered by $h_m(t)$.
- a noise component (spatially diffuse), $n_m(t)$.

Acoustical scenario

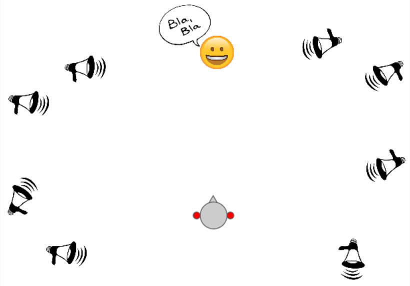


At each microphone $x_m(t)$:

- one speech source (plane wave), $s(t)$, filtered by $h_m(t)$.
- a noise component (spatially diffuse), $n_m(t)$.

$$\text{Time domain: } x_m(t) = (h_m \star s)(t) + n_m(t) \tag{1}$$

Acoustical scenario



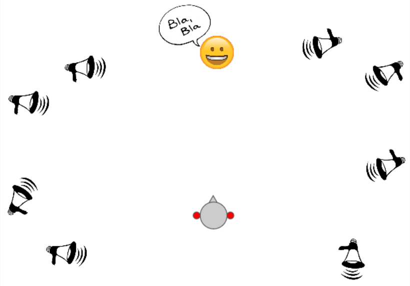
At each microphone $x_m(t)$:

- one speech source (plane wave), $s(t)$, filtered by $h_m(t)$.
- a noise component (spatially diffuse), $n_m(t)$.

$$\text{Time domain: } x_m(t) = (h_m \star s)(t) + n_m(t) \tag{1}$$

$$\text{STFT domain: } x_m(k,\ell) = h_m(k,\ell)s(k,\ell) + n_m(k,\ell) \tag{2}$$

Acoustical scenario



At each microphone $x_m(t)$:

- one speech source (plane wave), $s(t)$, filtered by $h_m(t)$.
- a noise component (spatially diffuse), $n_m(t)$.

$$\text{Time domain: } x_m(t) = (h_m \star s)(t) + n_m(t) \tag{1}$$

$$\text{STFT domain: } x_m(k,\ell) = h_m(k,\ell)s(k,\ell) + n_m(k,\ell) \tag{2}$$

$$\text{matrix notation: } \boldsymbol{x}(k,\ell) = \boldsymbol{h}(k)\ s(k,\ell) + \boldsymbol{n}(k,\ell). \tag{3}$$

$\boldsymbol{x}(k,\ell) \in \mathbb{C}^M$, $\boldsymbol{h}(k,\ell) \in \mathbb{C}^M$ and $\boldsymbol{n}(k,\ell) \in \mathbb{C}^M$.

Acoustical scenario

Speech sparsity assumption in the STFT domain:

- $\mathcal{H}_0$: $\boldsymbol{x}(k,\ell) = \boldsymbol{n}(k,\ell)$,
- $\mathcal{H}_1$: $\boldsymbol{x}(k,\ell) = \boldsymbol{h}(k)s(k,\ell) + \boldsymbol{n}(k,\ell)$

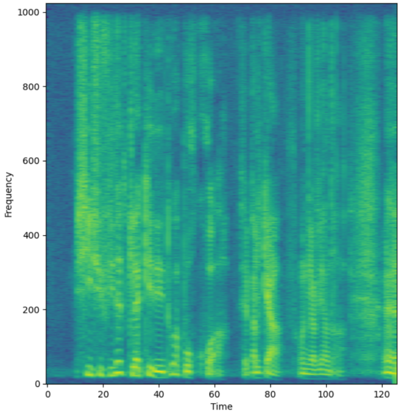


Figure 4: Speech spectrogram

Probabilistic model

Discrete Fourier coefficients are modeled as random variables:

$$s(k,\ell) \sim \mathcal{N}_{\mathbb{C}}\left(0, \phi_s(k,\ell)\right) \tag{4}$$

$$\boldsymbol{n}(k,\ell) \sim \mathcal{N}_{\mathbb{C}}\left(0, \Phi_{\boldsymbol{nn}}(k,\ell)\right) \tag{5}$$

Discrete Fourier coefficients are modeled as random variables:

$$s(k,\ell) \sim \mathcal{N}_{\mathbb{C}}\left(0, \phi_s(k,\ell)\right) \tag{4}$$

$$\boldsymbol{n}(k,\ell) \sim \mathcal{N}_{\mathbb{C}}\left(0, \Phi_{\boldsymbol{nn}}(k,\ell)\right) \tag{5}$$

Noise covariance matrix model (spatially diffuse):

$$\Phi_{\boldsymbol{nn}}(k,\ell) = \phi_n(k,\ell)\Gamma_{\mathrm{diff}}(k) \tag{6}$$

Proposed Algorithm

Source separation

$\hat{s}_L(k,\ell)$: speech source estimate at the left ear (same for right ear)

$$\hat{s}_L(k,\ell) = \boldsymbol{w}^H(k,\ell)\ \boldsymbol{x}(k,\ell) \tag{7}$$

with $\boldsymbol{w}(k,\ell) \in \mathbb{C}^M$.

Source separation

$\hat{s}_L(k,\ell)$: speech source estimate at the left ear (same for right ear)

$$\hat{s}_L(k,\ell) = \boldsymbol{w}^H(k,\ell)\ \boldsymbol{x}(k,\ell) \tag{7}$$

with $\boldsymbol{w}(k,\ell) \in \mathbb{C}^M$.

Optimization problem:

$$\hat{\boldsymbol{w}}_L(k,\ell) = \underset{\boldsymbol{w}}{\operatorname{argmin}}\{\mathbb{E}\left[\left|s_L(k,\ell) - \boldsymbol{w}^H\boldsymbol{x}(k,\ell)\right|^2\right]\} \tag{8}$$

Source separation

$\hat{s}_L(k,\ell)$: speech source estimate at the left ear (same for right ear)

$$\hat{s}_L(k,\ell) = \boldsymbol{w}^H(k,\ell)\ \boldsymbol{x}(k,\ell) \tag{7}$$

with $\boldsymbol{w}(k,\ell) \in \mathbb{C}^M$.

Optimization problem:

$$\hat{\boldsymbol{w}}_L(k,\ell) = \underset{\boldsymbol{w}}{\operatorname{argmin}}\{\mathbb{E}\left[\left|s_L(k,\ell) - \boldsymbol{w}^H\boldsymbol{x}(k,\ell)\right|^2\right]\} \tag{8}$$

Using the speech sparsity assumption

$$\begin{aligned}\mathbb{E}\left[\left|s_L(k,\ell) - \boldsymbol{w}^H\boldsymbol{x}(k,\ell)\right|^2\right] = P(\ell)\Big[&p(k,\ell)\mathbb{E}\left[|s_L(k,\ell) - \boldsymbol{w}^H\boldsymbol{x}(k,\ell)|^2 \mid \mathcal{H}_1\right] \\ &+(1-p(k,\ell))\mathbb{E}\left[|\boldsymbol{w}^H\boldsymbol{x}(k,\ell)|^2 \mid \mathcal{H}_0\right]\Big] \\ &+ (1-P(\ell))\left[\frac{1}{\mu_{\mathcal{H}_0}}\mathbb{E}\left[|s_L(k,\ell) - \boldsymbol{w}^H\boldsymbol{h}(k)s(k,\ell)|^2\right]\right. \\ &\left.+\mathbb{E}\left[|\boldsymbol{w}^H\boldsymbol{x}(k,\ell)|^2\right]\right]\end{aligned} \tag{9}$$

Informed Multichannel Wiener Filter (MWF) for the left ear:

$$\hat{\boldsymbol{w}}_L(k,\ell) = \left(\phi_s(k,\ell)\boldsymbol{h}(k)\boldsymbol{h}(k)^H + \mu(k,\ell)\Phi_{\boldsymbol{nn}}(k,\ell)\right)^{-1}\boldsymbol{h}(k)\phi_s(k,\ell)h_L(k)^*, \tag{10}$$

with

$$\mu(k,\ell) = P(\ell)\frac{1}{p(k,\ell)} + (1-P(\ell))\mu_{\mathcal{H}_0}. \tag{11}$$

- $p(k,\ell)$: narrowband speech presence probability estimation
- $P(\ell)$: broadband speech presence probability estimation

Dynamic range compression - i

Idea introduced by [Ngo et al., 2012] and similarly by [May et al., 2018]: use different DRC to process speech source and noise source

- DRC_s: to process the speech source when it is active at this T-F bin,
- DRC_n H_1: to process the speech source when it is NOT active at this T-F bin,
- DRC_n H_0: to process the speech source when it is NOT active at all bins,
- DRC_n: to process the noise component.

DRC	Attack (ms)	Release (ms)	Gain G_0 (dB)
DRC_s	10	60	0
DRC_n H_1	10	2000	-6
DRC_n H_0	10	2000	-10
DRC_n	2000	2000	-10

Multichannel estimator [Souden et al., 2010]

$P(\mathcal{H}_1|\boldsymbol{x}(k,\ell))$: *a posteriori* source presence presence probability, denoted $p(k,\ell)$

Bayes rule:

$$P(\mathcal{H}_1|\boldsymbol{x}(k,\ell)) = \frac{P(\boldsymbol{x}(k,\ell)|\mathcal{H}_1)P(\mathcal{H}_1)}{P(\boldsymbol{x}(k,\ell)|\mathcal{H}_1)P(\mathcal{H}_1) + P(\boldsymbol{x}(k,\ell)|\mathcal{H}_0)P(\mathcal{H}_0)} \tag{12}$$

$P(\boldsymbol{x}(k,\ell)|\mathcal{H}_1)$: data likelihood according to the Gaussian assumption

$P(\mathcal{H}_1)$: prior (adaptive [Cohen, 2002])

Source presence probability - ii

Broadband binary detector

$$\hat{P}(\ell) = \begin{cases} 1 & \text{if } \sum_k p(k,\ell) > t_{high} \text{ and } P(\ell-1) = 0 \\ 0 & \text{if } \sum_k p(k,\ell) > t_{low} \text{ and } P(\ell-1) = 1 \\ P(\ell-1) & \text{otherwise,} \end{cases} \tag{13}$$

Recursive filtering

$$P(\ell) = \alpha_P \hat{P}(\ell) + (1-\alpha_P) P(\ell-1) \tag{14}$$

Sum-up

- Similar to the proposition of [Ngo et al., 2012]
- Improvements :
  - DRC association more consistent with the literature
  - attack and release time constant decorrelation between DRC and broadband speech detection
  - binaural rather than monaural

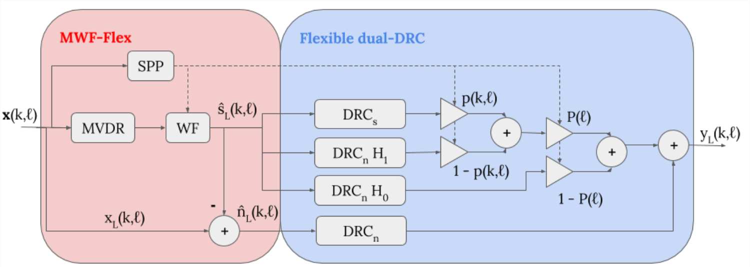
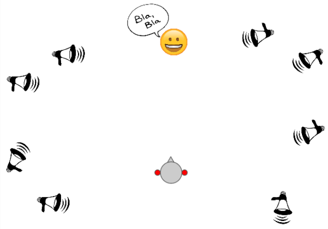


Figure 5: Our proposition

Experiments

Set-up

- Scenario:
  - speaker located in front of the listener
  - cafeteria noise
  - SNR: 5 dB
- Ideal scenario: SNR of 15 dB
- With 14 different HRTFs



Results - i

Interaural coherence

$$\mathrm{IC} = \max_{\tau} \left| \frac{\sum_t \tilde{y}_L(t+\tau)\tilde{y}_R(t)}{\sqrt{\sum_t |\tilde{y}_L(t)|^2 \sum_t |\tilde{y}_R(t)|^2}} \right| . \qquad (15)$$

Advantages

- Denoising across all the pipeline
- Interaural coherence closer to the ideal scenario

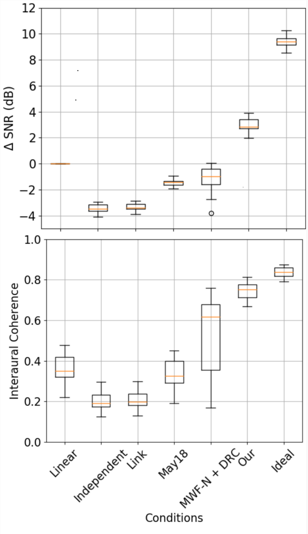


Figure 6: SNR improvement (top) and interaural coherence (bottom).

Effective Compression Ratio (ECR):

- < 1: more dynamic range
- > 1: less dynamic range

Drawbacks

- Speech cut $\rightarrow$ ECR deteriorated
- Residual noise component into the speech branch correlated with the speech $\rightarrow$ ECR < 1

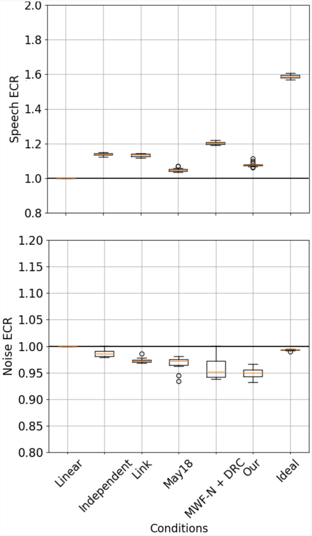


Figure 7: Speech ECR (top) and noise ECR (bottom).

Conclusion

- SNR improvement
- Better interaural coherence
- ECR deterioration due to estimation errors
- Future works: perceptive test

Thanks for your attention

Please share your comments and questions

References

Brimijoin, W. O., Whitmer, W. M., McShefferty, D., and Akeroyd, M. A. (2014).
The Effect of Hearing Aid Microphone Mode on Performance in an Auditory Orienting Task:.
Ear and Hearing, page 1.

Cohen, I. (2002).
Optimal speech enhancement under signal presence uncertainty using log-spectral amplitude estimator.
IEEE Signal Processing Letters, 9(4):113–116.

Hassager, H. G., May, T., Wiinberg, A., and Dau, T. (2017a).
Preserving spatial perception in rooms using direct-sound driven dynamic range compression.
Journ. of the Acoustical Society of America, 141(6):4556–4566.

Hassager, H. G., Wiinberg, A., and Dau, T. (2017b).
Effects of hearing-aid dynamic range compression on spatial perception in a reverberant environment.
Journ. of the Acoustical Society of America, 141(4):2556–2568.

Ibrahim, I., Parsa, V., Macpherson, E., and Cheesman, M. (2012).
Evaluation of speech intelligibility and sound localization abilities with hearing aids using binaural wireless technology.
Audiology Research, 3(1):1.

Koutrouvelis, A. I., Hendriks, R. C., Heusdens, R., and Jensen, J. (2017).
Relaxed Binaural LCMV Beamforming.
IEEE/ACM Transactions on Audio, Speech, and Language Processing, 25(1):137–152.

Luo, F.-L., Yang, J., Pavlovic, C., and Nehorai, A. (2002).
Adaptive null-forming scheme in digital hearing aids.
IEEE Transactions on signal processing, 50(7):1583–1590.

Marquardt, D. and Doclo, S. (2018).
Interaural Coherence Preservation for Binaural Noise Reduction Using Partial Noise Estimation and Spectral Postfiltering.
IEEE/ACM Transactions on Audio, Speech, and Language Processing, 26(7):1261–1274.

May, T., Kowalewski, B., and Dau, T. (2018).
Signal-to-Noise-Ratio-Aware Dynamic Range Compression in Hearing Aids.
Trends in Hearing, 22:233121651879090.

Ngo, K., Spriet, A., Moonen, M., Wouters, J., and Holdt Jensen, S. (2012).
A combined multi-channel Wiener filter-based noise reduction and dynamic range compression in hearing aids.
Signal Processing, 92(2):417–426.

Schwartz, A. H. and Shinn-Cunningham, B. G. (2013).
Effects of dynamic range compression on spatial selective auditory attention in normal-hearing listeners.
Journ. of the Acoustical Society of America, 133(4):2329–2339.

Souden, M., Jingdong Chen, Benesty, J., and Affes, S. (2010).
Gaussian Model-Based Multichannel Speech Presence Probability.
IEEE Transactions on Audio, Speech, and Language Processing, 18(5):1072–1077.

Van den Bogaert, T., Doclo, S., Wouters, J., and Moonen, M. (2009).
Speech enhancement with multichannel Wiener filter techniques in multimicrophone binaural hearing aids.
Journ. of the Acoustical Society of America, 125(1):360–371.

Van den Bogaert, T., Klasen, T. J., Moonen, M., Van Deun, L., and Wouters, J. (2006).
Horizontal localization with bilateral hearing aids: Without is better than with.
Journ. of the Acoustical Society of America, 119(1):515–526.

Wiggins, I. M. and Seeber, B. U. (2011).
Dynamic-range compression affects the lateral position of sounds.
Journ. of the Acoustical Society of America, 130(6):3939–3953.

Wiggins, I. M. and Seeber, B. U. (2013).
Linking dynamic-range compression across the ears can improve speech intelligibility in spatially separated noise.
Journ. of the Acoustical Society of America, 133(2):1004–1016.