Localization cues preservation in hearing aids by combining noise reduction and dynamic range compression

Adrien Llave, Simon Leglaive, Renaud Seguier

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

Adrien Llave, Simon Leglaive and Renaud Séguier
CentraleSupélec, IETR, France
firstname.lastname@centralesupelec.fr

***Abstract*—Dynamic range compression (DRC) and noise reduction algorithms are commonly used in hearing aids. They are known to have opposite objectives concerning the Signal-to-Noise Ratio (SNR) and to affect negatively the localization performance. Yet, the study of their interaction received few attention. In this work, we improve an existing combined approach of DRC and noise reduction to bridge the gap between the algorithms proposed independently in their respective communities. The proposed solution is then compared to state-of-the-art algorithms thanks to objective criteria assessing the spatial fidelity preservation, the SNR improvement and the output dynamic range reduction. Experimental results show that the standard serial concatenation of noise reduction and DRC stages is unable to improve the SNR and preserve the noise component acoustic characteristics. They suggest that the proposed design restores the noise localization cues and manages to improve the output SNR.**

***Index Terms*—Hearing aids, spatial hearing, beamforming, dynamic range compression**

I. Introduction

Dynamic range compression (DRC) is the main processing in hearing aids for compensating the listener hearing loss. It consists in amplifying quiet sounds and/or attenuating loud ones. To do so, the gain function defining this amplification/attenuation is split into two zones: (i) below a given threshold the input signal is linearly amplified; (ii) above the same threshold the gain follows a linear decreasing function in the dB domain.

Usually, hearing aids compute the left and right DRC gains independently. It has been showed that the difference between the left and right DRC gain distorts the interaural level difference (ILD) which is an important localization cue. It results in a worsening of the localization performance for the listener, which also affects speech understanding [27], [35].

The straightforward strategy to solve this issue is to apply the same DRC gain to both ears. It restores the ILD as well as the localization performance in anechoic environment [36]. However, other experiments failed to replicate this result [14], [16]. Particularly, it has been recently highlighted that the linked DRC fails at preserving the localization performance in reverberant conditions [13]. The authors showed that, in a reverberant environment, the preservation of the ILD is not sufficient to ensure the preservation of the localization cues. Indeed, usually, the DRC acts quickly in order to follow the short-term speech level fluctuations. Therefore, the reverberation tail is considered as a soft speech period which has to be reinforced. The consequence is a lowering of the direct-to-reverberant energy ratio (DRR), leading to more internalization, image source diffusion for the listener as well as more front-back confusion and source localization splitting. The authors further showed by means of a perceptual evaluation that the Interaural Coherence (IC) [11] is an objective criterion which better correlates with the localization performance of the listener.

To address this issue, it has been proposed [12] to prolong the DRC gain computed from the last direct sound period to the reverberation tail in order to preserve the DRR. To do so, in direct sound period, the DRC acts quickly whereas in the segments dominated by reverberant speech, the DRC acts slowly in order to prolong DRC attenuation from the last direct sound dominated period. The authors showed that such a direct-sound-driven DRC is able to restore both IC and localization performance on the horizontal plane in presence of reverberation. However, the design of this algorithm is limited to a specific auditory scenario composed of one speaker with reverberation but without background noise or other interfering speakers. In noisy environment, the DRC is unable to correct the speech dynamic range accurately [29]. Moreover, it reduces the output SNR [6], [29].

This algorithm has been extended to a scenario including background noise [24]. In this work, the authors use a speech presence detector to drive the DRC time constant, in the same way as in [12], where a direct sound detector was used to drive the DRC in noise-free reverberant conditions. The authors also defined explicitly the objectives of an ideal hearing aids system: (i) amplifying the soft-speech segments; (ii) keeping loud-speech segments below the pain level; (iii) avoiding amplification of the noise in speech absence; and (iv) preserving the original noise dynamic range. However, this solution answers only partially to these objectives. Indeed, the DRC gain applied on a noise-only segment is set according to the one computed on the last frame where speech was detected. As a consequence, the DRC gain highly differs from one noise-only segment to another. This effect is illustrated in Figure 1, where we can observe that the DRC gain on noise-only segments varies from -20 dB at 1.8 s to -12 dB at 3 s,

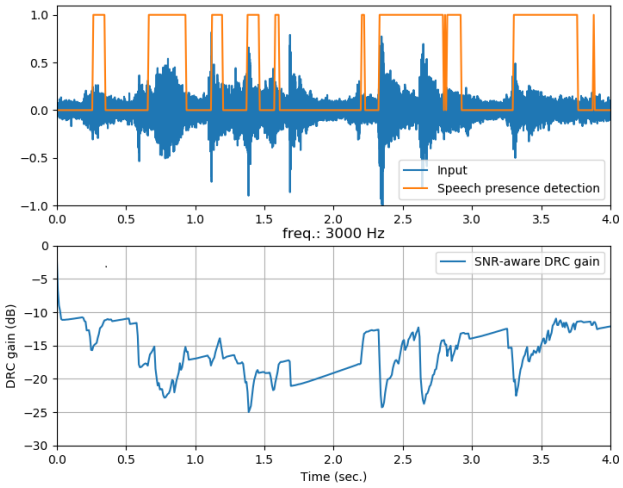

Fig. 1. Noisy speech and speech presence detection (top) and reduction gain at 3 kHz of the SNR-aware DRC proposed by May *et al.* [24].

depending on the DRC at preceding speech frames. It should be noted that this gain is frequency dependent. Therefore, at each speech-absence period, the frequency response of the filter applied to the noise segment differs in an uncontrolled way.

So far, we presented studies considering the DRC as the unique processing in hearing aids. However, an extensive work has been done in the last two decades to improve SNR in hearing aids [20], [30] before applying DRC. Speech enhancement based on beamforming techniques [8] have become the reference and are nowadays commonly used in commercial hearing aids. State-of-the-art noise reduction algorithms such as the Minimum Variance Distortionless Response (MVDR) beamformer and the Multichannel Wiener Filter (MWF) have the particularity of collapsing all the localization cues present in the auditory scene toward the steering direction. Therefore, the preservation of both speech and noise localization cues through the noise reduction stage has attracted a lot of attention in the last decade [18], [21], [22], [34] and is still an open issue. The IC has also been introduced in this research field both as an assessment criterion of the noise localization cues preservation and as a penalty term in beamforming optimization problems [23]. A common strategy to preserve localization cues consists in mixing the beamformer output with the original noisy signal at a reference microphone. It is called multichannel Wiener filtering with partial noise estimation (MWF-N) [33]. Localization performance in the horizontal plane can be restored with an appropriate gain for mixing the enhanced speech signal with the noisy one [22], [32], [33].

In the literature, we observed that noise reduction with beamforming and DRC are treated as two separate research topics, even though they correspond to two key elements of hearing aids. Only very few works studied their interaction and the consequence of their combination on localization and speech understanding performance [5], [17], [26]. Yet, it has been pointed out that noise reduction and DRC have antagonistic objectives [26]: the noise reduction stage aims to reduce the background noise while the DRC amplifies soft sounds and keeps loud sounds below the pain level. In other words, after lowering the noise level thanks to the noise reduction stage, the DRC increases it further. Therefore, it was proposed in [26] to extract both speech and noise signals from the input mixture using beamforming, then apply a DRC independently to each signal before mixing them back to obtain the final enhanced signal. The noise reintroduction level in this final mixing operation is quite low (about -10 dB), therefore residual interfering noise in the estimated speech signal from the beamformer can greatly affect the final noise level in the output mixture.

In this study, the authors only considered a monaural hearing aid system (a two microphones Behind-The-Ear (BTE) device), thus limiting the SNR improvement. Moreover, they did not make the link between their method and the MWF-N beamformer, and they did not assess their algorithm with respect to the localization cues preservation. Finally, their algorithm does not allow for the use of different time constants to process the noise and the speech, even though the above-mentioned studies suggest that it is relevant.

In this study, we propose to adapt the combined beamforming and DRC system from [26] with respect to the recent findings [12], [24]. Then, we add to the four aims of an ideal hearing aids system from [24] another one addressing the localization cues preservation. We rewrite them as follows:

1) reducing the speech dynamic range;
2) preserving the original noise dynamic range;
3) improving the output SNR;
4) preserving the localization cues of both speech and noise.

Each objectives will be assessed thanks to standard objective criteria.

The document is structured as follows: first, the data model is depicted in section II and the proposed algorithm is presented in section III. Then, it will be assessed and compared to the aforementioned algorithms in the section IV.

II. Data model

We consider an auditory scene composed of one speech source of interest denoted $s(t)$ and a spatially diffuse noise denoted $n_m(t)$, where $m \in \{1, ..., M\}$ is the microphone index, and t is the discrete-time index. The signal received at microphone m is expressed as follows:

$$x_m(t) = (h_m \star s)(t) + n_m(t), \tag{1}$$

where $\star$ denotes the convolution operator, and $h_m(t)$ is the impulse response of the acoustic channel from the speaker to m^{th} microphone. It can be expressed in the Short-Term Fourier Transform (STFT) domain. Assuming that $h_m(t)$ is short compared with the STFT analysis window, convolution in the time domain becomes a simple product in the STFT domain [2]:

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

where k and ℓ denote the frequency and frame indices, respectively. It is convenient, for the following, to consider this equation in the vector form by concatenating the terms related to the microphones:

$$\mathbf{x}(k,\ell) = \mathbf{h}(k)\ s(k,\ell) + \mathbf{n}(k,\ell), \tag{3}$$

where $\mathbf{x}(k,\ell) = [x_1(k,\ell), x_2(k,\ell), ..., x_M(k,\ell)]^T \in \mathbb{C}^M$, $\mathbf{h}(k) \in \mathbb{C}^M$ and $\mathbf{n}(k,\ell) \in \mathbb{C}^M$ are defined similarly.

Both speech and noise DFT coefficients are modeled as random variables following a zero mean isotropic complex Gaussian distribution of variances $\phi_s(k,\ell)$ and $\phi_n(k,\ell)$, respectively, and the noise is assumed to be spatially diffuse:

$$s(k,\ell) \sim \mathcal{N}_c(0, \phi_s(k,\ell)) \tag{4}$$

$$\mathbf{n}(k,\ell) \sim \mathcal{N}_c(0, \phi_n(k,\ell)\mathbf{\Gamma}_{\text{diff}}(k)) \tag{5}$$

where $\mathbf{\Gamma}_{\text{diff}}(k) \in \mathbb{C}^{M\times M}$ is the spatial coherence matrix corresponding to a diffuse noise field.

In addition, let us assume that the noise is always present while the speech can be present or absent. It leads to the following two mutually exclusive hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$:

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

Moreover, we consider an anechoic scenario (*i.e.* without reverberation) where the acoustic transfer functions $\mathbf{h}(k)$ are assumed to be known.

III. Proposed algorithm

Our objective is to apply an independent DRC on the speech and noise signals, while improving the SNR and preserving the localization cues of both speech and noise. We derive an algorithm close to the proposition of Ngo *et al.* [26]. Our approach differs in some points because the data model is not exactly the same. These differences will be detailed in the following. A schematic overview of the proposed algorithm is provided in Figure 2. It incorporates three main stages that we are going to detail in this section: (i) target speech and noise separation with beamforming; (ii) speech presence probability estimation; and (iii) dynamic range compression.

A. Source separation

The source separation algorithm is based on a MWF. It consists in estimating the target speech signal at the left ear reference microphone $s_L(k,\ell) = h_L(k)\ s(k,\ell)$ by a linear combination of the microphone signals in the STFT domain:

$$\hat{s}_L(k,\ell) = \hat{\mathbf{w}}_L(k,\ell)^H \mathbf{x}(k,\ell), \tag{6}$$

where $\hat{s}_L(k,\ell)$ is the speech estimate at the left-ear reference microphone, $\hat{\mathbf{w}}_L(k,l) \in \mathbb{C}^M$ is an estimate of the unknown coefficients (or weights) of the beamformer, and $\cdot^H$ denotes the Hermitian transpose. The speech signal estimate at the right ear $\hat{s}_R(k,\ell)$ is defined similarly and for the sake of brevity, only the left expression will be derived in the following.

The beamformer weights are estimated by solving an optimization problem [8]. For the MWF beamformer, it consists in minimizing the mean square error between the true signal and its estimate:

$$\hat{\mathbf{w}}_L(k,\ell) = \underset{\mathbf{w}}{\text{argmin}}\{J_1(\mathbf{w})\}, \tag{7}$$

where

$$J_1(\mathbf{w}) = \mathbb{E}\left[\left|s_L(k,\ell) - \mathbf{w}^H\mathbf{x}(k,\ell)\right|^2\right]. \tag{8}$$

We use an alternative formulation from [26] called MWF-Flex, which leverages an estimate of the speech presence probability (SPP). The corresponding cost function $J_1(\mathbf{w}_L(k,\ell))$ can be split into several terms, allowing us to set a speech distortion weight depending on whether the speech is present or not:

$$\begin{aligned} J_2(\mathbf{w}) =& P(\ell)\left[p(k,\ell)\mathbb{E}\left[|s_L(k,\ell) - \mathbf{w}^H\mathbf{x}(k,\ell)|^2 \,\middle|\, \mathcal{H}_1\right]\right. \\ &\left.+(1-p(k,\ell))\mathbb{E}\left[|\mathbf{w}^H\mathbf{x}(k,\ell)|^2 \,\middle|\, \mathcal{H}_0\right]\right] \\ &+ (1-P(\ell))\left[\frac{1}{\mu_{\mathcal{H}_0}}\mathbb{E}\left[|s_L(k,\ell) - \mathbf{w}^H\mathbf{h}(k)s(k,\ell)|^2\right]\right. \\ &\left.+\mathbb{E}\left[|\mathbf{w}^H\mathbf{x}(k,\ell)|^2\right]\right], \end{aligned} \tag{9}$$

where $P(\ell) \in [0,1]$ is the broadband speech presence probability, $p(k,\ell) \in [0,1]$ the narrowband speech presence probability and $\mu_{\mathcal{H}_0} \in \mathbb{R}$ is the attenuation to be applied of the speech component during the speech absence period.

The minimizer of (9) is given by [26]:

$$\begin{aligned} \hat{\mathbf{w}}_L(k,\ell) = &\left(\phi_s(k,\ell)\mathbf{h}(k)\mathbf{h}(k)^H + \mu(k,\ell)\mathbf{\Phi}_{\mathbf{nn}}(k,\ell)\right)^{-1} \\ &\times \mathbf{h}(k)\phi_s(k,\ell)h_L(k)^*, \end{aligned} \tag{10}$$

where $\mathbf{\Phi}_{\mathbf{nn}}(k,\ell) = \mathbb{E}\left[\mathbf{n}(k,\ell)\mathbf{n}(k,\ell)^H\right] = \phi_n(k,\ell)\mathbf{\Gamma}_{\text{diff}}(k)$, the operator $\cdot^*$ denotes the complex conjugate, and

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

This solution can be decomposed into a Minimum Variance Distorsionless (MVDR) beamforming filter $\mathbf{w}_{\text{MVDR}}$, a parametric Wiener filter w_{WF} and a spatialization filter h_L^* [3]:

$$\hat{\mathbf{w}}_L(k,\ell) = w_{\text{WF}}(k,\ell)\ \mathbf{w}_{\text{MVDR}}(k)\ h_L(k)^*, \tag{12}$$

where

$$w_{\text{WF}}(k,\ell) = \frac{\xi(k,\ell)}{\mu(k,\ell) + \xi(k,\ell)} \tag{13}$$

and $\xi(k,\ell)$ is the *a priori* SNR at the MVDR beamformer output, defined by:

$$\xi(k,\ell) = \frac{\phi_s(k,\ell)}{\phi_n(k,\ell)\mathbf{w}_{\text{MVDR}}(k)^H\mathbf{\Gamma}_{\text{diff}}(k)\mathbf{w}_{\text{MVDR}}(k)}. \tag{14}$$

As we assume a spatially diffuse noise, the MVDR beamformer turns into a maximum directivity index one [31]:

$$\mathbf{w}_{\text{MVDR}}(k) = \frac{\mathbf{\Gamma}_{\text{diff}}(k)^{-1}\mathbf{h}(k)}{\mathbf{h}(k)^H\mathbf{\Gamma}_{\text{diff}}(k)^{-1}\mathbf{h}(k)}. \tag{15}$$

Assuming a time-independent noise spatial coherence matrix makes the beamformer no longer time dependent. The diffuse

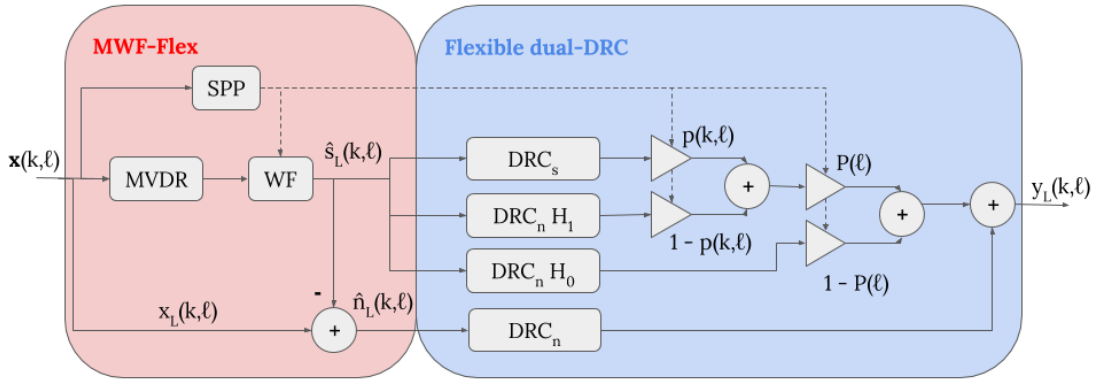

Fig. 2. Schematic overview of the proposed algorithm.

noise coherence matrix is defined as the mean of the contribution of D acoustic transfer functions uniformly spaced on the horizontal planes:

$$\mathbf{\Gamma}_{\text{diff}}(k) = \frac{1}{D}\sum_{d=1}^{D}\mathbf{h}_d(k)\mathbf{h}_d(k)^H \tag{16}$$

where $\mathbf{h}_d$ is the anechoic transfer functions from the d^{th} direction.

The left ear noise estimate, $\hat{n}_L(k,\ell)$, can be derived in a similar way leading to:

$$\hat{n}_L(k,\ell) = x_L(k,\ell) - \hat{s}_L(k,\ell), \tag{17}$$

where $x_L(k,\ell)$ is the microphone signal used as a reference for the left ear.

Our implementation of the MWF-Flex differs from [26] because the data model is not exactly the same. In this study, the authors do not assume the *a priori* knowledge of $\mathbf{h}$ and that the noise is spatially diffuse. Therefore, they have to estimate the speech and noise covariance matrices which is difficult and not robust in presence of reverberation. Moreover, it allows us to steer in an arbitrary direction *i.e.* selecting the desired target rather than steering in an uncontrolled direction based on the speech covariance matrix estimation. This is also an advantage for the speech presence probability estimation presented in the following.

B. Dynamic range compression

In this subsection, we start by describing the general operation of the DRC used in this work and then we will describe how the DRCs are arranged in our algorithm.

The DRC consists in filtering the input signal by a gain $g(k,\ell) \in \mathbb{R}$ depending on its input level. To do so, the input signal is passed through a rectangular filterbank and the instantaneous power of each band is computed. The resulting power is filtered with a first order recursive low-pass filter with time constants depending on whether the input signal is in a rising period (attack) or in a falling one (release). The release time constant is usually greater than the attack one. The resulting signal envelop is expressed in dB and is denoted by $\tilde{P}_b^{\text{dB}}(\ell)$ in the following, where b denotes the frequency bands associated with the above-mentioned filter bank.

The DRC gain expressed in dB is defined as follows for the b^{th} frequency band:

$$G_b(\ell) = \begin{cases} G_0 + \left(\tilde{P}_b^{\text{dB}}(\ell) - T\right)\left(\frac{1}{R} - 1\right) & \text{if } \tilde{P}_b^{\text{dB}}(\ell) > T \\ G_0 & \text{otherwise,} \end{cases} \tag{18}$$

where T is the compression threshold, G_0 a constant gain, and R is the compression ratio. Then, the gain is brought back into the linear domain and passed through the inverse filterbank.

The drawback of the serial concatenation of the noise reduction stage and the fast-acting DRC is that the latter reduces the SNR improvement achieved by the former. Therefore, it has been proposed in [26] to use three DRCs in parallel, and to switch from one to another depending on whether the speech is active or not:

- "DRC$_s$" refers to the DRC applied when the speech is present at this time-frequency point,
- "DRC$_n$ H$_1$" refers to the DRC applied when the speech is active in the frame but not at this time-frequency point,
- and, "DRC$_n$ H$_0$" refers to the DRC applied when the speech is absent of the frame.

A drawback of the original implementation of a such DRC combination in [26] is the impossibility of using different smoothing time constants for each DRC. Indeed, in the authors' implementation the attack/release smoothing is performed on the final DRC gain in the dB domain. Maybe even more problematic, the attack time constant is used to smooth the gain both for the rising speech periods and for the transition between an active speech period and a noise-dominated one. The issue is similar for the release time constant.

In this work, we propose a different implementation by combining the DRC gains in the linear domain rather than in the dB domain, and by low-pass filtering of the input power envelop rather than of the output gain in dB:

$$\begin{aligned} g(k,\ell) =& P(\ell)\left[p(k,\ell)g_{\text{s}}(k,\ell) + (1-p(k,\ell))g_{\text{H1}}(k,\ell)\right] \\ &+ (1-P(\ell))\; g_{\text{H0}}(k,\ell) \end{aligned} \tag{19}$$

where $g_s(k,\ell)$, $g_{\text{H1}}(k,\ell)$ and $g_{\text{H0}}(k,\ell)$ are the gains associated to "DRC$_s$", "DRC$_n$ H$_1$" and "DRC$_n$ H$_0$", respectively. It gives us more flexibility and consistency with the standard DRC designs used in hearing aids [13], [15], [24].

C. Speech presence probability estimation

The source separation algorithm, as well as the DRCs presented previously, involve the knowledge of the narrowband and broadband speech presence probabilities $p(k,\ell)$ and $P(\ell)$. In this subsection, we detail how the speech presence parameters are estimated. To do so, we use the two hypotheses, $\mathcal{H}_0$ and $\mathcal{H}_1$, presented in Section II.

Using the Bayes theorem, the *a posteriori* speech presence probability, denoted by $p(k,\ell) = P(\mathcal{H}_1|\mathbf{x}(k,\ell))$, can be expressed as follows [28]:

$$p(k,\ell) = \left(1 + \frac{q(k,\ell)}{1-q(k,\ell)}(1+\zeta(k,\ell))e^{-\frac{\beta(k,\ell)}{1+\zeta(k,\ell)}}\right)^{-1} \quad (20)$$

where $q(k,\ell) = P(\mathcal{H}_0)$ is the *a priori* speech absence probability, estimated recursively as in [4], $\zeta(k,\ell)$ is similar to the SNR:

$$\zeta(k,\ell) = \frac{\phi_s(k,\ell)}{\phi_n(k,\ell)}\text{Tr}\left\{\mathbf{\Gamma}_{\text{diff}}(k)^{-1}\mathbf{h}(k)\mathbf{h}(k)^H\right\}, \quad (21)$$

with Tr$\{.\}$ is the trace operator, and finally

$$\beta(k,\ell) = \mathbf{x}(k,\ell)^H\mathbf{\Gamma}_{\text{diff}}(k)^{-1}\mathbf{h}(k)\mathbf{h}(k)^H\mathbf{\Gamma}_{\text{diff}}(k)^{-1}\mathbf{x}(k,\ell) \times \frac{\phi_s(k,\ell)}{\phi_n(k,\ell)^2}. \quad (22)$$

The speech and noise variances estimates, denoted by $\hat{\phi}_s(k,\ell)$ and $\hat{\phi}_n(k,\ell)$ respectively, are defined as follows [10]:

$$\hat{\phi}_s(k,\ell) = \mathbf{w}_{\text{MVDR}}(k)^H\left(\hat{\mathbf{\Phi}}_{\mathbf{xx}}(k,\ell) - \hat{\phi}_n(k,\ell)\mathbf{\Gamma}_{\text{diff}}(k)\right) \times \mathbf{w}_{\text{MVDR}}(k) \quad (23)$$

$$\hat{\phi}_n(k,\ell) = \frac{1}{M-1}\text{Tr}\left\{\mathbf{P}(k)\hat{\mathbf{\Phi}}_{\mathbf{xx}}(k,\ell)\mathbf{\Gamma}_{\text{diff}}(k)^{-1}\right\} \quad (24)$$

where $\mathbf{P}(k) = \mathbf{I} - \mathbf{h}(k)\mathbf{w}_{\text{MVDR}}(k)^H$ with $\mathbf{I}$ the identity matrix of size M. The input data covariance estimate $\hat{\mathbf{\Phi}}_{\mathbf{xx}}(k,\ell)$ is computed thanks to a recursive filter:

$$\hat{\mathbf{\Phi}}_{\mathbf{xx}}(k,\ell) = \alpha\ \mathbf{x}(k,\ell)\ \mathbf{x}(k,\ell)^H + (1-\alpha)\hat{\mathbf{\Phi}}_{\mathbf{xx}}(k,\ell-1) \quad (25)$$

where α is the smoothing factor.

The broadband speech presence detection, $P(\ell)$, is computed thanks to a hysteresis comparator with adaptive thresholds:

$$P(\ell) = \begin{cases} 1 & \text{if } \sum_k p(k,\ell) > t_{high} \text{ and } P(\ell-1) = 0 \\ 0 & \text{if } \sum_k p(k,\ell) > t_{low} \text{ and } P(\ell-1) = 1 \\ P(\ell-1) & \text{otherwise,} \end{cases} \quad (26)$$

where t_{low} and t_{high} are computed by means of a one-dimensional 2-means clustering algorithm over the $p(k,\ell)$ for the L last frames. Finally, $P(\ell)$ is smoothed with a recursive first order low-pass filter with attack and release time constants to avoid the gate effect due to the DRC gain between speech+noise and noise-only segments.

IV. Evaluation

In this section, we assess the proposed algorithm according to the different objectives defined in the introduction. For each objective, a standard objective criterion is associated and the performance are compared to reference algorithms we already presented in the introduction. First, we will depict the reference algorithms implementation. Second, we will present the experimental set-up and, third, the assessment criteria will be detailed. Finally, the results will be presented. The audio examples for the first subject are available online[1].

A. Reference algorithms

Firstly, the unprocessed signal is considered as a reference and is called the *linear* condition. The *independent* and *linked* conditions refer to a DRC applied to the left and right ears microphones, independently or not, respectively. For the *linked* condition, the DRC gains are computed independently for each ear and the minimum gain is applied to both. The filterbank is a rectangular octave-spaced one from 125 to 8 kHz. The SNR-aware DRC proposed in [24] is also tested and called *May18*. This algorithm consists in using a different set of attack/release time constants depending on whether the speech is active or not in the frame. In [24], the authors use another data model leading to a different SPP estimation algorithm than the one presented in section III-C. For the sake of data model consistency, we use the latter. The *MWF-N+DRC* consists in the serial concatenation of a MWF-N [33] and a DRC. The MWF-N consists in a weighted sum between the MWF output and the left (or right) reference microphone. It achieves a trade-off between noise reduction and preservation of the noise localization cues [34]. The MWF is implemented as a concatenation of a fixed MVDR beamformer (see (15)) and a Wiener filter [3]. The trade-off parameter is set to 0.3 (noise reintroduction gain of -10 dB) in order to maximize the IC [22]. Finally, the *ideal* condition consists in applying the DRC to the speech sentence before spatialization [13] and mixed the DRC output signals with a gain difference of 10 dB.

Frequency (Hz)	125	250	500	1k	2k	4k	8k
Threshold (dB$_{\text{SPL}}$)	31	36	40	32	34	31	9
Ratio	2.2	2.2	1.8	1.9	2.2	2.9	2.6

TABLE I
DRC PARAMETERS

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

TABLE II
DRC OVERALL GAIN AND TIME CONSTANT PARAMETERS

[1]Audio examples repository URL: https://a-llave.github.io/demo_apsipa2020

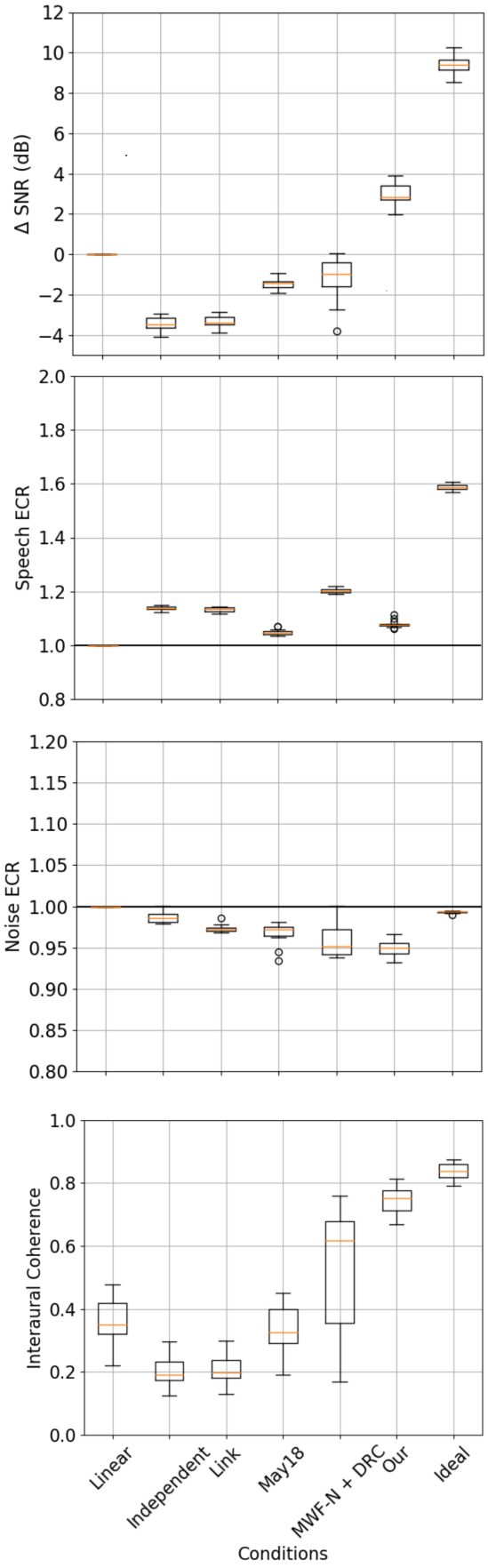

Fig. 3. The SNR improvement (ΔSNR), speech and noise Effective Compression Ratio (ECR) and Interaural Coherence (IC) for each conditions. The box limits represent the 1^{st} and 3^{rd} quartiles of the subjects results, the orange line indicates the median, the Whiskers show the data range up to 1.5 times the interquartile distance, and circles show outliers.

B. Experiment set-up

The auditory scene is simulated thanks to a 4 s-long male speech sentence recording from the France Culture radio station and a mono cafeteria recording. We use the first fourteen BTE hearing aids anechoic HRTF from the OTIMP database [25] (front and back microphones, $M = 4$) resulting in as many stimuli. The noise stimulus is built from the mono cafeteria recording. It is cut into 4 s pieces and spatialized through virtual loudspeakers thanks to the BTE HRIR. The loudspeakers are arranged in two rings at $\pm 45°$ of elevation with a horizontal resolution of $22.5°$. The speech stimulus is placed at the frontal direction and presented at 75 $\text{dB}_{\text{SPL}}$and the SNR is equal to 5 dB. The tested algorithms are embedded into an overlap-add framework with a 128 samples (8 ms) long frame and 50 % overlap. The analysis and synthesis windows are the square root of a Hann function and the sampling frequency is 16 kHz. The DRC settings used in this work are the same as in [13] and are summarized in the table I. The low-pass filter time constant used in (25) for the estimation of $\mathbf{\Phi_{xx}}$ is 9 ms and the broadband speech presence detection $P(\ell)$ is smoothed with a similar recursive filter with attack and release time constants of 2 ms and 20 ms, respectively. The DRC thresholds and ratios for each frequency bands are set as in [13] and are summarized in Tab. I. The overall gain for each DRC as well as the time constants used for the power envelop estimation are gathered in Tab. II.

C. Assessment criteria

1) Signal-to-Noise Ratio improvement: For a source $s(t)$ and a noise $n(t)$, the SNR, in dB, is defined as follows:

$$\text{SNR} = 10\log_{10}\sum_{t=0}^{N-1} s(t)^2 - 10\log_{10}\sum_{t=0}^{N-1} n(t)^2 \qquad (27)$$

where N is the total number of samples in the time domain. Only the periods where the speech is active are considered. The SNR improvement (ΔSNR) is defined as the difference between the output SNR and the *linear* condition one.

For each condition, the required processing parameters are computed from the noisy speech recording. The resulting filters are applied to the mixed $\mathbf{x}(k,\ell)$ as well as the speech and noise components separately. It allows us to compute the true output SNR. This method is sometimes called the shadow-filtering [24] in the literature.

2) Interaural Coherence: The IC is used to assess the perception of the width of the auditory scene [19] and is known to be important to access to the localization cues [7]. It is defined as the absolute maximum value of the normalized cross-correlation between the left and right band-pass filtered output signals, denoted by $\tilde{y}_L$ and $\tilde{y}_R$ with $|\tau| \leq 1$ ms [11], [13]:

$$\text{IC} = \max_{\tau}\left|\frac{\sum_t \tilde{y}_L(t+\tau)\tilde{y}_R(t)}{\sqrt{\sum_t |\tilde{y}_L(t)|^2 \sum_t |\tilde{y}_R(t)|^2}}\right|. \qquad (28)$$

The hearing aids output signals $y_L(t)$ and $y_R(t)$ are passed through a filterbank modeling the auditory system composed

of fourth order gammatone filters with equivalent rectangular bandwidth spacing [9].

3) Effective Compression Ratio: The ECR is the ratio between the dynamic range in input and output. The dynamic range is computed by slicing the signal in short frames of 10 ms, computing the power (in dB) of each ones, and then, taking the levels standard deviation [1]. This metric is computed for each frequency channel and finally averaged across the frequencies.

D. Results and discussion

In this section, we analyze the results of the experiments. The performance of each algorithm is summarized in the Figure 3.

1) SNR improvement: On one hand, as expected, the *independent* and *linked* conditions show an SNR lowering greater than 3 dB. This result is consistent with [6], [29]. As for the SNR aware DRC from [24], it manages to contain this effect but still reduces the SNR by 1.5 dB. On the other hand, in the conditions including noise reduction techniques, the SNR is lowered by 1.5 dB for the MWF-N+DRC condition and improved by 3 dB for our proposition. This result suggests that the parallel DRC combination overcomes the antagonistic objectives dichotomy highlighted by Ngo *et al*. [26] and allows us to reach the target SNR. However, we can see it remains a large room of improvement to reach the ideal performance. This is due to the noise component leakage into the speech branch. This may be improved thanks to a more complex beamformer.

2) Speech and noise ECR: An ECR greater than 1 indicates a reduction of the dynamic range and *vice versa*. First, we show that all the conditions without noise reduction stage lead to a speech ECR lower than the *ideal* condition. Taking advantage of beamforming, *MWF-N+DRC* and our proposition achieve a better compression because the low-level speech segments are no longer mixed up with the background noise but a gap remains compared to the *ideal* performance, possibly due to speech estimation errors (in high frequency particularly). However, our algorithm failed to improve the speech ECR probably because of speech presence probability estimation errors implying an excessive attenuation of the speech after a silent period. Second, we show that the noise dynamic range is increased (ECR < 1) for all the conditions. Our proposition failed to improve the criterion. However, a part of the noise component is highly correlated with the speech in such a way that it is masked by the latter one. Informal tests suggest that it is not perceived to be as problematic.

3) Interaural Coherence: Firstly, we have to note that the input (*linear*) and the *ideal* IC are not equal because the input and *ideal* SNR are not the same in the experiments. Similarly to the ΔSNR, the conditions without noise reduction stage reduces the IC. As for the *MWF-N+DRC*, due to the serial concatenation of beamforming and DRC, it fails to increase the IC up to the *ideal* one. Moreover, the inter-subject variability is particularly important. Finally, our proposition manages to get close to the IC of the *ideal* condition. More investigation is needed to show if the remaining gap is perceptually important or not.

V. Conclusion

In this study, we considered an auditory scene composed of one speech source and a cafeteria noise (spatially diffuse noise). Firstly, we pointed out that the serial concatenation of beamforming algorithm and DRC fails at improving the SNR, consistently with [26]. Moreover, it fails at preserving the Interaural Coherence (IC) as well as they make the noise comodulates with the speech envelop. Secondly, we proposed an amelioration of an existing algorithm combining DRC and noise reduction in order to increase both the SNR improvement and the spatial fidelity of the auditory scene. We showed that the proposed combined approach reaches the IC target performance and manages to improve the SNR. However, some estimation errors in our algorithm prevent the speech and noise ECR from being improved. These outcomes have to be confirmed thanks to perceptual tests.

References

[1] Joshua M. Alexander and Varsha Rallapalli. Acoustic and perceptual effects of amplitude and frequency compression on high-frequency speech. *Journ. of the Acoustical Society of America*, 142(2):908–923, August 2017.

[2] Y. Avargel and I. Cohen. On Multiplicative Transfer Function Approximation in the Short-Time Fourier Transform Domain. *IEEE Signal Processing Letters*, 14(5):337–340, May 2007.

[3] Lowell Brooks and Irving Reed. Equivalence of the Likelihood Ratio Processor, the Maximum Signal-to-Noise Ratio Filter, and the Wiener Filter. *IEEE Transactions on Aerospace and Electronic Systems*, AES-8(5):690–692, September 1972.

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

[5] Ryan M. Corey and Andrew C. Singer. Dynamic range compression for noisy mixtures using source separation and beamforming. In *2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 289–293, New Paltz, NY, October 2017. IEEE.

[6] Ryan Michael Corey. *Microphone Array Processing for Augmented Listening*. PhD thesis, University of Illinois, Urbana-Champaign, 2019.

[7] Christof Faller and Juha Merimaa. Source localization in complex listening situations: Selection of binaural cues based on interaural coherence. *Journ. of the Acoustical Society of America*, 116(5):3075–3089, November 2004.

[8] S. Gannot, Emmanuel Vincent, Shmulik Markovich-Golan, and Alexey Ozerov. A Consolidated Perspective on Multimicrophone Speech Enhancement and Source Separation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 25(4):692–730, April 2017.

[9] Brian R Glasberg and Brian C.J Moore. Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, 47(1-2):103–138, August 1990.

[10] Hao Ye and D. DeGroat. Maximum likelihood DOA estimation and asymptotic Cramer-Rao bounds for additive unknown colored noise. *IEEE Transactions on Signal Processing*, 43(4):938–949, April 1995.

[11] William M Hartmann, Brad Rakerd, and Aaron Koller. Binaural Coherence in Rooms. *ACTA ACUSTICA UNITED WITH ACUSTICA*, 91:12, 2005.

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

[13] Henrik Gert Hassager, Alan Wiinberg, and Torsten Dau. 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, April 2017.

[14] Iman Ibrahim, Vijay Parsa, Ewan Macpherson, and Margaret Cheesman. Evaluation of speech intelligibility and sound localization abilities with hearing aids using binaural wireless technology. *Audiology Research*, 3(1):1, December 2012.

[15] James M. Kates. Principles of Digital Dynamic-Range Compression. *Trends in Amplification*, 9(2):45–76, March 2005.

[16] Gitte Keidser, Kristin Rohrseitz, Harvey Dillon, Volkmar Hamacher, Lyndal Carter, Uwe Rass, and Elizabeth Convery. The effect of multi-channel wide dynamic range compression, noise reduction, and the directional microphone on horizontal localization performance in hearing aid wearers. *Int. Journal of Audiology*, 45(10):563–579, January 2006.

[17] Petri Korhonen, Chi Lau, Francis Kuk, Denise Keenan, and Jennifer Schumacher. Effects of Coordinated Compression and Pinna Compensation Features on Horizontal Localization Performance in Hearing Aid Users. *Journ. of the American Academy of Audiology*, 26(1):80–92, 2015.

[18] A. Koutrouvelis. *Multi-Microphone Noise Reduction for Hearing Assistive Devices*. PhD thesis, Delft University of Technology, 2018.

[19] Kohichi Kurozumi and Kengo Ohgushi. The relationship between the cross-correlation coefficient of two-channel acoustic signals and sound image quality. *Journ. of the Acoustical Society of America*, 74(6):1726–1733, December 1983.

[20] Fa-Long Luo, Jun Yang, Chaslav Pavlovic, and Arye Nehorai. Adaptive null-forming scheme in digital hearing aids. *IEEE Transactions on signal processing*, 50(7):1583–1590, 2002.

[21] D. Marquardt. *Development and evaluation of psychoacoustically motivated binaural noise reduction and cue preservation techniques*. PhD thesis, Von der Fakultät für Medizin und Gesundheitswissenschaften der Carl von Ossietzky Universität Oldenburg, Oldenburg, November 2015.

[22] D. Marquardt and S. Doclo. 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, July 2018.

[23] D. Marquardt, Volker Hohmann, and S. Doclo. Coherence preservation in multi-channel Wiener filtering based noise reduction for binaural hearing aids. In *2013 IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, pages 8648–8652, Vancouver, BC, Canada, May 2013. IEEE.

[24] Tobias May, Borys Kowalewski, and Torsten Dau. Signal-to-Noise-Ratio-Aware Dynamic Range Compression in Hearing Aids. *Trends in Hearing*, 22:233121651879090, January 2018.

[25] Alastair H. Moore, Jan Mark de Haan, Michael Syskind Pedersen, Patrick A. Naylor, Mike Brookes, and Jesper Jensen. Personalized signal-independent beamforming for binaural hearing aids. *Journ. of the Acoustical Society of America*, 145(5):2971–2981, May 2019.

[26] Kim Ngo, Ann Spriet, Marc Moonen, Jan Wouters, and Søren Holdt Jensen. A combined multi-channel Wiener filter-based noise reduction and dynamic range compression in hearing aids. *Signal Processing*, 92(2):417–426, February 2012.

[27] Andrew H. Schwartz and Barbara G. Shinn-Cunningham. 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, 2013.

[28] Mehrez Souden, Jingdong Chen, Jacob Benesty, and Sofiène Affes. Gaussian Model-Based Multichannel Speech Presence Probability. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(5):1072–1077, July 2010.

[29] Pamela E. Souza, Lorienne M. Jenstad, and Kumiko T. Boike. Measuring the acoustic effects of compression amplification on speech in noise. *Journ. of the Acoustical Society of America*, 119(1):41–44, January 2006.

[30] Ann Spriet, Marc Moonen, and Jan Wouters. Spatially Pre-Processed Speech Distortion Weighted Multi-Channel Wiener Filtering For Noise Reduction In Hearing Aids. In *Int. Workshop on Acoustic Echo and Noise Control*, volume 84, pages 2367–2387, Kyoto, Japan, September 2003. Signal Processing.

[31] R. W. Stadler and William M. Rabinowitz. On the potential of fixed arrays for hearing aids. *Journ. of Acoustical Society of America*, 94(3):1332–1342, September 1993.

[32] Joachim Thiemann, Menno Müller, D. Marquardt, S. Doclo, and Steven van de Par. Speech enhancement for multimicrophone binaural hearing aids aiming to preserve the spatial auditory scene. *EURASIP Journal on Advances in Signal Processing*, 2016(1), December 2016.

[33] Tim Van den Bogaert, S. Doclo, Jan Wouters, and Marc Moonen. The effect of multimicrophone noise reduction systems on sound source localization by users of binaural hearing aids. *Journ. of the Acoustical Society of America*, 124(1):484–497, July 2008.

[34] Tim Van den Bogaert, S. Doclo, Jan Wouters, and Marc Moonen. Speech enhancement with multichannel Wiener filter techniques in multimicrophone binaural hearing aids. *Journ. of the Acoustical Society of America*, 125(1):360–371, January 2009.

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

[36] I. M. Wiggins and B. U. Seeber. 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, 2013.