Unification of multi-microphone noise reduction systems

Ann Spriet, Simon Doclo, Marc Moonen, Jan Wouters

7th February 2006

1 General cost function

1.1 Signal model

Let $X_i(f)$, $i = 1,\ldots,M$ denote the frequency-domain microphone signals. Each microphone signal $X_i(f)$, $i = 1,\ldots,M$ can be decomposed into a speech component $X_i^s(f)$ and an additive noise component $X_i^n(f)$ as:

$$X_i(f) = X_i^s(f) + X_i^n(f). \tag{1}$$

Defining $H_i^s(f)$ as the acoustic transfer function from the speech source $S(f)$ to the i-th microphone, the speech component $X_i^s(f)$ equals:

$$X_i^s(f) = H_i^s(f)S(f) = \frac{H_i^s(f)}{H_1^s(f)}X_1^s(f), \tag{2}$$

were $\tilde{H}_i^s(f) = \frac{H_i^s(f)}{H_1^s(f)}$ denotes the relative transfer function ratio of the i-th to the first microphone.

Let $\mathbf{X}(f) \in \mathbb{C}^{M\times 1}$ be defined as the stacked vector

$$\mathbf{X}(f) = \begin{bmatrix} X_1(f) & X_2(f) & \cdots & X_M(f) \end{bmatrix}^T. \tag{3}$$

Then, (1) can be written as

$$\mathbf{X}(f) = \mathbf{X}^s(f) + \mathbf{X}^n(f) = \mathbf{H}^s(f)S(f) + X^n(f) = \tilde{\mathbf{H}}^s(f)X_1^s(f) + \mathbf{X}^n(f), \tag{4}$$

with $\mathbf{X}^s(f)$ and $\mathbf{X}^n(f)$ defined similarly as in (3) and

$$\tilde{\mathbf{H}}^s(f) = \mathbf{H}^s(f)/H_1^s(f) = \begin{bmatrix} 1 & \frac{H_2^s(f)}{H_1^s(f)} & \ldots & \frac{H_M^s(f)}{H_1^s(f)} \end{bmatrix}^T. \tag{5}$$

To simplify notation, we define the power spectral density (PSD) of the speech and noise component in the i-th microphone signal as

$$P_{X_i}^s(f) = \varepsilon\{X_i^s(f)X_i^{*,s}(f)\} \tag{6}$$
$$P_{X_i}^n(f) = \varepsilon\{X_i^n(f)X_i^{*,n}(f)\} \tag{7}$$

and the PSD of the speech source $S(f)$ as

$$P_S(f) = \varepsilon\{S(f)S^*(f)\}. \tag{8}$$

In addition, we define the noise and speech correlation matrix as:

$$\mathbf{R}^n(f) = \varepsilon\{\mathbf{X}^n(f)\mathbf{X}^{n,H}(f)\}, \tag{9}$$
$$\mathbf{R}^s(f) = \varepsilon\{\mathbf{X}^s(f)\mathbf{X}^{s,H}(f)\} = P_{X_1}^s(f)\tilde{\mathbf{H}}^s(f)\tilde{\mathbf{H}}^{s,H}(f). \tag{10}$$

1.2 Free-field propagation model

Single point source

Assuming free-field propagation, mathematical expressions can be derived for the acoustic tranfer function from a point source $S(f,\mathbf{p})$ to the M microphones. Let $S(f,\mathbf{p})$ be defined as a point source at location $\mathbf{p}$ with cartesian coordinates $\mathbf{p} = (x,y,z)$ and spherical coordinates (R,θ,ϕ) (where R is the distance, θ the azimuth, ϕ the elevation) as defined in Figure 1. Without loss of generality, we define the origin of the coordinates at the position of the first microphone of the microphone array. The contribution $X_i(f,\mathbf{p})$ of the point source $S(f,\mathbf{p})$ in the i-th microphone signal (with coordinates $\mathbf{p}_i$) equals

$$X_i(f,\mathbf{p}) \quad = \quad A_i(f,\mathbf{p})a_i(\mathbf{p})e^{-j2\pi f\tau_i(\mathbf{p})}S(f,\mathbf{p}), \tag{11}$$

where $A_i(f,\mathbf{p}) = A_i(f,\theta,\phi)$ includes the microphone characteristics of the i-th microphone (and in the case of a hearing aid, the head related transfer function to the i-th microphone), $a_i(\mathbf{p})$ is the attenuation of the point source $S(f,\mathbf{p})$ at the position of the i-th microphone (near-field effect) and

$$\tau_i(\mathbf{p}) = \frac{\|\mathbf{p}-\mathbf{p}_i\|}{c} \tag{12}$$

with c the speed of sound (340 m/s), is the propagation delay from the point source $S(f,\mathbf{p})$ to the i-th microphone. Defining the first microphone signal $X_1(f,\mathbf{p})$ as reference signal, $\mathbf{X}(f,\mathbf{p})$ can be written as:

$$\mathbf{X}(f,\mathbf{p}) = \tilde{\mathbf{d}}(f,\mathbf{p})X_1(f,\mathbf{p}) \tag{13}$$

where $\tilde{\mathbf{d}}(f,\mathbf{p})$ is the steering vector

$$\tilde{\mathbf{d}}(f,\mathbf{p}) = \left[\begin{array}{cccc} 1 & \frac{A_2(f,\mathbf{p})}{A_1(f,\mathbf{p})}\frac{a_2(\mathbf{p})}{a_1(\mathbf{p})}e^{-j2\pi f(\tau_2(\mathbf{p})-\tau_1(\mathbf{p}))} & \dots & \frac{A_M(f,\mathbf{p})}{A_1(f,\mathbf{p})}\frac{a_M(\mathbf{p})}{a_1(\mathbf{p})}e^{-j2\pi f(\tau_M(\mathbf{p})-\tau_1(\mathbf{p}))} \end{array}\right]^T. \tag{14}$$

Using (28), it can be shown that the PSD $P_{X_1(f,\mathbf{p})}(f)$ of the first microphone signal $X_1(f,\mathbf{p})$ (i.e., the reference signal) equals:

$$P_{X_1(f,\mathbf{p})}(f) = |A_1(f,\mathbf{p})a_1(\mathbf{p})|^2 P_{S(f,\mathbf{p})}(f), \tag{15}$$

where $P_{S(f,\mathbf{p})}(f)$ is the PSD of the source $S(f,\mathbf{p})$. If the first microphone is omnidirectional (i.e., $A_1(f,\mathbf{p}) = 1$), the PSD of the first microphone signal equals the PSD of the source signal $S(f,\mathbf{p})$ up to a scalar $|a_1(\mathbf{p})|^2$. An estimate of the first microphone signal is then a scaled and delayed version of the estimate of the source signal.

Multiple point sources

If several point sources $S(f,\mathbf{p})$ at positions $\mathbf{p} \in \mathbf{P}$ are propagating, the microphone signals $\mathbf{X}(f)$ can be modeled as:

$$\mathbf{X}(f) = \int_{\mathbf{p}\in\mathbf{P}} \tilde{\mathbf{d}}(f,\mathbf{p})X_1(f,\mathbf{p}), \tag{16}$$

with $X_1(f,\mathbf{p})$ defined by (28).

For uncorrelated point sources

$$\varepsilon\{X_1(f,\mathbf{p}_k)X_1(f,\mathbf{p}_l)\} = P_{X_1}(f)\delta_{kl}. \tag{17}$$

The model (16)-(17) can be used when the speech/noise sources cover a certain known region in space or when an approximate position of the speech source is known.

Remark: The free-field propagation model assumes that there is no reverberation and that the microphone characteristics and positions, the HRTFs (in case of hearing aids) are known. In practice, these assumptions will often be violated (e.g., microphone mismatch, reverberation) such that the true model (4) deviates from the free-field model (13)-(14), i.e.,

$$\tilde{\mathbf{H}}(f) = \tilde{\mathbf{d}}(f) + \delta\tilde{\mathbf{d}}(f). \tag{18}$$

Because of this deviation, techniques assuming a free-field propagation model may have a performance degradation in practice. The amount of degradation depends on the deviation $\delta\tilde{\mathbf{d}}(f)$.

1.2.1 Far-field propagation

For far-field propagation, (14) equals

$$\tilde{\mathbf{d}}(f,\mathbf{p}) = \left[\begin{array}{cccc} 1 & \frac{A_2(f,\mathbf{p})}{A_1(f,\mathbf{p})}e^{-j2\pi f(\tau_2(\mathbf{p})-\tau_1(\mathbf{p}))} & \dots & \frac{A_M(f,\mathbf{p})}{A_1(f,\mathbf{p})}e^{-j2\pi f(\tau_M(\mathbf{p})-\tau_1(\mathbf{p}))} \end{array}\right]^T, \tag{19}$$

where

$$\tau_j(\mathbf{p}) - \tau_1(\mathbf{p}) = \frac{-x_j\sin\phi\cos\theta - y_j\sin\phi\sin\theta - z_j\cos\phi}{c} \tag{20}$$

For the special case of a linear array (i.e., $\phi = 90^\circ$, $y = 0$), $\tau_j(\mathbf{p}) - \tau_1(\mathbf{p})$ reduces to:

$$\tau_j(\mathbf{p}) - \tau_1(\mathbf{p}) = \frac{-x_j\cos\theta}{c}. \tag{21}$$

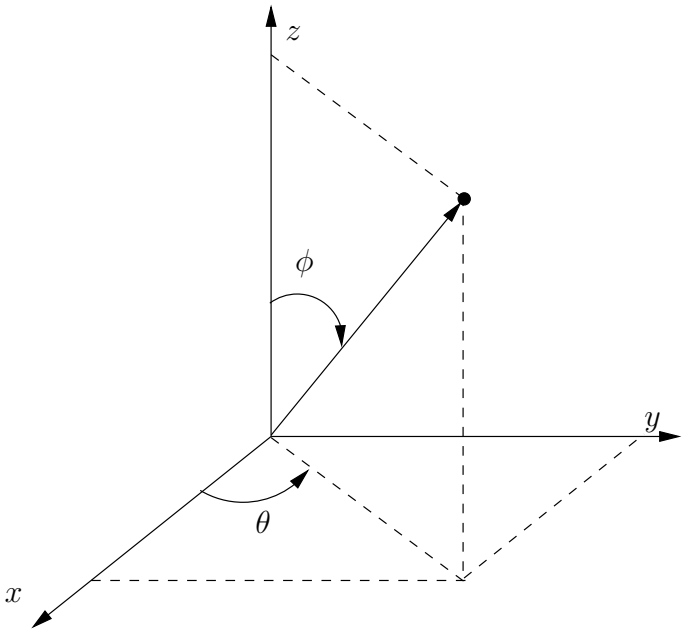

Figure 1: Coordinate system with cartesian and spherical coordinates. The origin corresponds to the position of the first microphone.

1.2.2 Near-field propagation

For near-field propagation, the steering vector $\tilde{\mathbf{d}}(f,\mathbf{p})$ equals

$$\tilde{\mathbf{d}}(f,\mathbf{p}) = \left[\begin{array}{cccc} 1 & \frac{A_2(f,\mathbf{p})}{A_1(f,\mathbf{p})}\frac{a_2(\mathbf{p})}{a_1(\mathbf{p})}e^{-j2\pi f(\tau_2(\mathbf{p})-\tau_1(\mathbf{p}))} & \cdots & \frac{A_M(f,\mathbf{p})}{A_1(f,\mathbf{p})}\frac{a_M(\mathbf{p})}{a_1(\mathbf{p})}e^{-j2\pi f(\tau_M(\mathbf{p})-\tau_1(\mathbf{p}))} \end{array} \right]^T, \tag{22}$$

with

$$a_i(\mathbf{p}) = \frac{1}{\|\mathbf{p}-\mathbf{p}_i\|}. \tag{23}$$

1.3 Multi-microphone noise reduction

In a multi-microphone noise reduction system, the microphone signals $X_i(f)$ are filtered by (adaptive or fixed) filters $W_i(f)$ and combined in order to obtain an enhanced speech signal $Z(f)$. To simplify the formulation of the noise reduction algorithms, we define

$$\mathbf{W}(f) = \left[\begin{array}{cccc} W_1(f) & W_2(f) & \cdots & W_M(f) \end{array} \right]^H, \tag{24}$$

with $W_l(f) = \sum_{m=0}^{L-1} w_{l,m} e^{-j2\pi\frac{f}{f_s}m}$. The output $Z(f)$ of the multi-channel noise reduction algorithm can then be expressed as

$$Z(f) \quad = \quad \mathbf{W}^H(f)\mathbf{X}(f) = \underbrace{\mathbf{W}^H(f)\,\mathbf{X}^s(f)}_{Z^s(f)} + \underbrace{\mathbf{W}^H(f)\mathbf{X}^n(f)}_{Z^n(f)}. \tag{25}$$

The goal of the filter $\mathbf{W}(f)$ is to minimize the output noise energy as much as possible without severely distorting the speech signal. The amount of speech distortion is measured with respect to a reference speech signal $D^s(f)$. This reference signal can be the speech component $X_1^s(f)$ in the first microphone, the speech source signal $S(f)$ or the speech component in the output of a fixed beamformer (e.g., the speech reference in the spatially pre-processed SDW-MWF **[?]**).
A general cost function $J(\mathbf{W}(f))$ for the filter $\mathbf{W}(f)$ is:

$$\begin{array}{l} J(\mathbf{W}) \;=\; (1-\lambda)\mathbf{W}^H(f)\mathbf{R}^n(f)\mathbf{W}(f) + \lambda\mathbf{W}^H(f)\mathbf{R}_m^n(f)\mathbf{W}(f)))^H\} \\ +\mu_1\varepsilon\{(D^s(f)-\mathbf{W}^H(f)\mathbf{X}^s(f))(D^s(f)-\mathbf{W}^H(f)\mathbf{X}^s(f+\mu_2\varepsilon\{(D_m^s(f)-\mathbf{W}^H(f)\mathbf{X}_m^s(f))(D_m^s(f)-\mathbf{W}^H(f)\mathbf{X}_m^s(f))^H\}. \end{array} \tag{26}$$

The first two terms in $J(\mathbf{W})$ correspond to output noise power. This output noise power can be:

- estimated online (i.e., the term $\mathbf{W}^H\mathbf{R}^n(f)\mathbf{W}(f)$) or

Speech model	Noise model	Hard/Soft constraint on speech distortion		Technique
Fixed beamforming				
A-priori (Section 2)	A-priori	$\mu_1 = 0$	$\mu_2 = \infty$	LCMV-based
		$\mu_1 = 0$	$\mu_2 \neq \infty$	Weighted LS criterion
None (Section 2)	A-priori	$\mu_1 = 0$	$\mu_2 = 0$	Differential microphone array (Constraint on $\mathbf{W}(f)$)
Adaptive beamforming				
A-priori (Section 3)	Online	$\mu_1 = 0$	$\mu_2 = \infty$	GSC
		$\mu_1 = 0$	$\mu_2 \neq \infty$	Soft-constrained MWF (Nordholm)
	Combination	$\mu_1 = 0$	$\mu_2 = \infty$	Sensitivity constrained GSC
		$\mu_1 = 0$	$\mu_2 \neq \infty$	Soft-constrained with partial noise model (E.g. Nordholm calibration data)
Online (Section 4)	Online	$\mu_1 = \infty$	$\mu_2 = 0$	TF-LCMV/ABM
		$\mu_1 \neq \infty$	$\mu_2 = 0$	SDW-MWF
	Combination	$\mu_1 = \infty$	$\mu_2 = 0$	TF-LCMV with (partial) noise model
		$\mu_1 \neq \infty$	$\mu_2 = 0$	SDW-MWF with (partial) noise model
	Fixed	$\mu_1 = \infty$	$\mu_2 = 0$	TF-LCMV with noise model
		$\mu_1 \neq \infty$	$\mu_2 = 0$	SDW-MWF with noise model (not useful)
Combination (Section 5)	Online	$\mu_1 \neq \infty$	$\mu_2 = \infty$	SDR-GSC
		$\mu_1 \neq \infty$	$\mu_2 \neq \infty$	Combination SDW-MWF/soft-constrained
		$\mu_1 = \infty$	$\mu_2 \neq \infty$	Combination TF-LCMV/GSC (cf. Kates)
	Combination	$\mu_1 \neq \infty$	$\mu_2 = \infty$	SDR-GSC with partial noise model
		$\mu_1 \neq \infty$	$\mu_2 \neq \infty$	SDW-MWF/Soft-constrained + partial noise model
		$\mu_1 = \infty$	$\mu_2 \neq \infty$	TF-LCMV/GSC with partial noise model
	Fixed	$\mu_1 = \infty$	$\mu_2 \neq \infty$	TF-LCMV/GSC with noise model

Table 1: Classification of multi-microphone noise reduction techniques.

- based on a pre-defined model $\mathbf{R}_m^n(f)$ of the noise correlation matrix, which is constructed through calibration measurements or mathematical models.

The last two terms in $J(\mathbf{W})$ denote the distortion energy between the output speech component $\mathbf{W}^H(f)\mathbf{X}^s(f)$ (or $\mathbf{W}^H(f)\mathbf{X}_m^s(f)$) and a reference speech signal $D^s(f)$ (or $D_m^s(f)$). Again, the output speech distortion energy may be

- estimated online (i.e., as $\varepsilon\{(D^s(f)-\mathbf{W}^H(f)\mathbf{X}^s(f))(D^s(f)-\mathbf{W}^H(f)\mathbf{X}^s(f))^H\}$) or
- based on a pre-defined model $\mathbf{X}_m^s(f)$ for the microphone signals (i.e., as $\varepsilon\{(D_m^s(f)-\mathbf{W}^H(f)\mathbf{X}_m^s(f))(D_m^s(f)-\mathbf{W}^H(f)\mathbf{X}_m^s(f))^H\}$). Again, this model can be constructed based on calibration data or based on mathematical models.

Depending on the use of a-priori knowledge of the speech and/or noise correlation matrix and the use of a hard constraint on the speech distortion term (i.e., $\mu_{1,2} = \infty$ or $\mu_{1,2} \neq \infty$), different existing multi-microphone noise reduction techniques can be obtained, as indicated in Table 1. When using a hard constraint ($\mu_1 = \infty$ or $\mu_2 = \infty$), noise suppression is only achieved in the subspace orthogonal to the defined or actual subspace of speech. Signals in the (defined or actual) speech subspace are passed through undistorted by the noise reduction algorithm. The use of a soft-constraint ($\mu_1 \neq \infty$ or $\mu_2 \neq \infty$) typically results in a spectral filtering of the desired speech component $D^s(f)$ since the speech and noise subspace are generally not orthogonal (often, the noise subspace spans the complete space).

Below, the different techniques are explained in more detail.

2 A-priori speech and noise model: fixed beamforming

2.1 Hard constraint on speech distortion ($\mu_2 = \infty$)

Cost function

$$J(\mathbf{W}) \quad = \quad \mathbf{W}^H(f)\mathbf{R}_m^n(f)\mathbf{W}(f) + \mu_2\varepsilon\{(D_m^s(f)-\mathbf{W}^H(f)\mathbf{X}_m^s(f))(D_m^s(f)-\mathbf{W}^H(f)\mathbf{X}_m^s(f))^H\} \tag{27}$$

with $\mu_2 = \infty$.

Assumed speech model The free-field propagation model in (13)-(14) is assumed:

$$\mathbf{X}_m^s(f) = \tilde{\mathbf{d}}^s(f,\mathbf{p}^s)X_{m,1}^s(f), \tag{28}$$

where $\mathbf{p}^s$ refers to the position of the speech source. The reference signal $D_m^s(f) = X_{m,1}^s(f)$.

Assumed noise model Different noise models have been used in the literature. The most well-known are:

- Delay-and-sum beamformer: a homogeneous, spatially uncorrelated noise field is assumed, i.e.,

$$\mathbf{R}_m^n(f) = P_N(f)\mathbf{I}_M, \tag{29}$$

  with $P_{X_i}^n(f) = P_N(f)$, $i = 1,\ldots,M$.

- Superdirective beamformer [**?**]: a homogeneous, diffuse (spherical isotropic) noise field is assumed, i.e.,

$$\mathbf{R}_m^n(f) = P_N(f)\Gamma^n(f) \tag{30}$$

  with $P_{X_i}^n(f) = P_N(f)$, $i = 1,\ldots,M$ and $\Gamma^n(f)$ the coherence matrix of diffuse noise, i.e.,

$$\Gamma(f) = \begin{bmatrix} 1 & \Gamma_{12}(f) & \cdots & \Gamma_{1M}(f) \\ \Gamma_{21}(f) & 1 & \cdots & \Gamma_{2M}(f) \\ \vdots & \vdots & \vdots & \vdots \\ \Gamma_{M1}(f) & \Gamma_{M2}(f) & \cdots & 1 \end{bmatrix} \tag{31}$$

$$\Gamma_{kl}(f) = \frac{\sin(2\pi f\|\mathbf{p}_l - \mathbf{p}_k\|/c)}{2\pi f\|\mathbf{p}_l - \mathbf{p}_k\|/c} = \mathrm{sinc}(2\pi f\|\mathbf{p}_l - \mathbf{p}_k\|/c) \tag{32}$$

  with $\|\mathbf{p}_l - \mathbf{p}_k\|$ the interspacing between microphone l and k.

- Combination of diffuse and spatially uncorrelated noise field: sensitivity constrained superdirective beamformer [**?**, **?**]

$$\mathbf{R}_m^n(f) = P_N(f)\left(\Gamma_m^n(f) + \eta(f)\mathbf{I}_M\right),$$

  with $\Gamma_m^n(f)$ the coherence matrix of diffuse noise and $\eta(f)$ a (frequency-dependent) weighting factor.

Solution The filter that minimizes (27) equals

$$\mathbf{W}(f) = \left(\mathbf{R}_m^n(f) + \mu_2 P_{X_{m,1}}^s(f)\tilde{\mathbf{d}}^s(f,\mathbf{p}^s)\tilde{\mathbf{d}}^{s,H}(f,\mathbf{p}^s)\right)^{-1}\mu_2 P_{X_{m,1}}^s(f)\tilde{\mathbf{d}}^s(f,\mathbf{p}^s). \tag{33}$$

Since $\tilde{\mathbf{d}}^s(f,\mathbf{p}^s)\tilde{\mathbf{d}}^{s,H}(f,\mathbf{p}^s)$ equals a rank-one matrix and $\left(\mathbf{R}_m^n(f) + \mu_2 P_{X_{m,1}}^s(f)\tilde{\mathbf{d}}^s(f,\mathbf{p}^s)\tilde{\mathbf{d}}^{s,H}(f,\mathbf{p}^s)\right)$ is assumed to be full-rank, the matrix inversion lemma can be applied to $\left(\mathbf{R}_m^n(f) + \mu_2 P_{X_{m,1}}^s(f)\tilde{\mathbf{d}}^s(f,\mathbf{p}^s)\tilde{\mathbf{d}}^{s,H}(f,\mathbf{p}^s)\right)^{-1}$:

$$\left(\mathbf{R}_m^n(f) + \mu_2 P_{X_{m,1}}^s(f)\tilde{\mathbf{d}}^s(f,\mathbf{p}^s)\tilde{\mathbf{d}}^{s,H}(f,\mathbf{p}^s)\right)^{-1} = \mathbf{R}_m^{n^{-1}}(f) - \frac{\mathbf{R}_m^{n^{-1}}(f)\mu_2 P_{X_{m,1}}^s(f)\tilde{\mathbf{d}}^s(f,\mathbf{p}^s)\tilde{\mathbf{d}}^{s,H}(f,\mathbf{p}^s)\mathbf{R}_m^{n^{-1}}(f)}{1+\mu_2 P_{X_{m,1}}^s(f)\tilde{\mathbf{d}}^{s,H}(f,\mathbf{p}^s)\mathbf{R}_m^{n^{-1}}(f)\tilde{\mathbf{d}}^s(f,\mathbf{p}^s)} \tag{34}$$

such that

$$\mathbf{W}(f) = \frac{\mathbf{R}_m^{n^{-1}}(f)}{1+\mu_2 P_{X_{m,1}}^s(f)\tilde{\mathbf{d}}^{s,H}(f,\mathbf{p}^s)\mathbf{R}_m^{n^{-1}}(f)\tilde{\mathbf{d}}^s(f,\mathbf{p}^s)}\mu_2 P_{X_{m,1}}^s(f)\tilde{\mathbf{d}}^s(f,\mathbf{p}^s). \tag{35}$$

Using $\mu_2 = \infty$ (i.e., hard constraint on the speech distortion term),

$$\mathbf{W}(f) = \frac{\mathbf{R}_m^{n^{-1}}(f)\tilde{\mathbf{d}}^s(f,\mathbf{p}^s)}{\tilde{\mathbf{d}}^{s,H}(f,\mathbf{p}^s)\mathbf{R}_m^{n^{-1}}(f)\tilde{\mathbf{d}}^s(f,\mathbf{p}^s)} = \frac{\Gamma_m^{n^{-1}}(f)\tilde{\mathbf{d}}^s(f,\mathbf{p}^s)}{\tilde{\mathbf{d}}^{s,H}(f,\mathbf{p}^s)\Gamma_m^{n^{-1}}(f)\tilde{\mathbf{d}}^s(f,\mathbf{p}^s)}. \tag{36}$$

2.2 Soft constraint on speech distortion ($\mu_2 \neq \infty$)

Cost function

$$J(\mathbf{W}) \quad = \quad \mathbf{W}^H(f)\mathbf{R}_m^n(f)\mathbf{W}(f) + \mu_2\varepsilon\{(D_m^s(f) - \mathbf{W}^H(f)\mathbf{X}_m^s(f))(D_m^s(f) - \mathbf{W}^H(f)\mathbf{X}_m^s(f))^H\} \tag{37}$$

with $\mu_2 \neq \infty$.

Example The weighted Least-Squares (wLS) criterion [**?**, **?**, **?**, **?**, **?**, **?**]

$$J_{WLS}(\mathbf{W}(f)) \quad = \quad \int_{\mathbf{p}\in\mathbf{P}} L^2(f,\mathbf{p}) \left[\mathbf{W}^H(f)\tilde{\mathbf{d}}(f,\mathbf{p}) - D(f,\mathbf{p})\right]^2 d\mathbf{p}, \tag{38}$$

with $D(f,\mathbf{p})$ a desired directivity pattern, can be transformed to the cost function (37), namely when the desired directivity pattern $D(f,\mathbf{p})$ is defined as:

$$\begin{aligned} D(f,\mathbf{p}) \quad &\neq \quad 0 \text{ for } \mathbf{p} \in \mathbf{P}_{pass} \\ &= \quad 0 \text{ for } \mathbf{p} \in \mathbf{P}_{stop}. \end{aligned} \tag{39}$$

$$J(\mathbf{W}(f)) = \int_{\mathbf{p}\in\mathbf{P}_{stop}} \mathbf{W}^H(f)\tilde{\mathbf{d}}(f,\mathbf{p})\tilde{\mathbf{d}}^H(f,\mathbf{p})\mathbf{W}(f)d\mathbf{p} + \mu_2 \int_{\mathbf{p}\in\mathbf{P}_{pass}} (1-\mathbf{W}^H(f)\tilde{\mathbf{d}}(f,\mathbf{p}))(1-\mathbf{W}^H(f)\tilde{\mathbf{d}}(f,\mathbf{p}))^H d\mathbf{p}. \tag{40}$$

This corresponds to the following speech and noise model:

- Speech model: The speech source is modeled as an infinite number of (uncorrelated) point sources with PSD $L^2(f,\mathbf{p})$ in the angular region $\mathbf{P}_{pass}$:

$$\mathbf{X}_m^s(f) = \int_{\mathbf{p}\in\mathbf{P}_{pass}} \tilde{\mathbf{d}}^s(f,\mathbf{p}) X_1^s(f,\mathbf{p}) d\mathbf{p}, \tag{41}$$

  with

$$\varepsilon\{X_1^s(f,\mathbf{p}_k) X_1^{s,*}(f,\mathbf{p}_l)\} = L^2(f,\mathbf{p}_k)\delta_{kl} \quad \text{for } \mathbf{p}_k,\mathbf{p}_l \in \mathbf{P}_{pass}. \tag{42}$$

  The reference signal $D_m^s(f)$equals:

$$D_m^s(f) = \int_{\mathbf{p}\in\mathbf{P}_{pass}} D(f,\mathbf{p}) X_1^s(f,\mathbf{p}) d\mathbf{p}. \tag{43}$$

- Noise model: The noise source is modeled as an infinite number of (uncorrelated) point sources with PSD $L^2(f,\mathbf{p})$ in the angular region $\mathbf{P}_{stop}$

$$\mathbf{X}_m^n(f) = \int_{\mathbf{p}\in\mathbf{P}_{stop}} \tilde{\mathbf{d}}^n(f,\mathbf{p}) X_1^n(f,\mathbf{p}) d\mathbf{p}, \tag{44}$$

  with

$$\varepsilon\{X_1^n(f,\mathbf{p}_k) X_1^{n,*}(f,\mathbf{p}_l)\} = L^2(f,\mathbf{p}_k)\delta_{kl} \quad \text{for } \mathbf{p}_k,\mathbf{p}_l \in \mathbf{P}_{stop}. \tag{45}$$

  From (44)-(45), it follows that:

$$\mathbf{R}_m^n = \int_{\mathbf{p}\in\mathbf{P}_{stop}} L^2(f,\mathbf{p})\tilde{\mathbf{d}}^n(f,\mathbf{p})\tilde{\mathbf{d}}^{n,H}(f,\mathbf{p}) d\mathbf{p}. \tag{46}$$

2.3 No constraint on speech distortion: differential microphone arrays [?, ?]

Cost function

$$J(\mathbf{W}(f)) = \mathbf{W}^H(f)\mathbf{R}_m^n\mathbf{W}(f) \tag{47}$$

with $\mathbf{W}(f) = \begin{bmatrix} 1 & \alpha \end{bmatrix}^T$ to avoid the trivial solution $\mathbf{W}(f) = 0$. The noise is modelled as $M-1$ uncorrelated noise sources, i.e,

$$\mathbf{X}_m^n(f) = \sum_{i=1}^{M-1} X_1^n(f,\mathbf{p}_i)\tilde{\mathbf{d}}(f,\mathbf{p}_i) \tag{48}$$

with

$$\varepsilon\{X_1^n(f,\mathbf{p}_i) X_1^n(f,\mathbf{p}_j)\} = P_{X_1}^n(f,\mathbf{p}_i)\delta_{ij} \tag{49}$$

Using (48)-(49), the noise correlation matrix $\mathbf{R}_m^n$ equals

$$\mathbf{R}_m^n(f) = \sum_{i=1}^{M-1} P_{X_1}^n(f,\mathbf{p}_i)\tilde{\mathbf{d}}(f,\mathbf{p}_i)\tilde{\mathbf{d}}^H(f,\mathbf{p}_i) \tag{50}$$

where $\mathbf{p}_i$ are the coordinates of the noise sources. Typically a linear array and far-field propagation is assumed such that $\mathbf{p}_i$ is characterized by the azimuth θ_i^n of the noise source (cf. Section 1.2.1). Depending on θ_i^n, different directivity patterns are obtained (e.g., cardioid ($\theta_i^n = 180°$), hypercardioid ($\theta_i^n = 90°$, ...).

3 A-priori speech model

3.1 Online estimated noise model

3.1.1 Hard constraint ($\mu_2 = \infty$): GSC [?, ?]

Cost function

$$J(\mathbf{W}) \quad = \quad \mathbf{W}^H(f)\mathbf{R}^n(f)\mathbf{W}(f) + \mu_2 \varepsilon\{(D_m^s(f) - \mathbf{W}^H(f)\mathbf{X}_m^s(f))(D_m^s(f) - \mathbf{W}^H(f)\mathbf{X}_m^s(f))^H\} \tag{51}$$

with $\mu_2 = \infty$.

Assumed speech model The free-field propagation model in (13)-(14) is assumed:

$$\mathbf{X}_m^s(f) = \tilde{\mathbf{d}}^s(f,\mathbf{p}^s) X_{m,1}^s(f), \tag{52}$$

where $\mathbf{p}^s$ refers to the position of the speech source. The reference signal $D_m^s(f)$ equals $X_{m,1}^s(f)$.

Noise model The noise model is estimated online.

Solution The filter $\mathbf{W}(f)$ equals

$$\mathbf{W}(f) = \left(\mathbf{R}^n(f) + \mu_2 P_{X_1}^s(f)\tilde{\mathbf{d}}^s(f,\mathbf{p}^s)\tilde{\mathbf{d}}^{s,H}(f,\mathbf{p}^s)\right)^{-1} \mu_2 P_{X_1}^s(f)\tilde{\mathbf{d}}^s(f,\mathbf{p}^s). \tag{53}$$

Since $\tilde{\mathbf{d}}^s(f,\mathbf{p}^s)\tilde{\mathbf{d}}^{s,H}(f,\mathbf{p}^s)$ equals a rank-one matrix and $\left(\mathbf{R}^n(f) + \mu_2 P_{X_1}^s(f)\tilde{\mathbf{d}}^s(f,\mathbf{p}^s)\tilde{\mathbf{d}}^{s,H}(f,\mathbf{p}^s)\right)$ is assumed to be full-rank, the matrix inversion lemma can be applied, resulting in (cf. Section 2.1):

$$\mathbf{W}(f) = \frac{\mathbf{R}^{n^{-1}}(f)}{1 + \mu_2 P_{X_1}^s(f)\tilde{\mathbf{d}}^{s,H}(f,\mathbf{p}^s)\mathbf{R}^{n^{-1}}(f)\tilde{\mathbf{d}}^s(f,\mathbf{p}^s)} \mu_2 P_{X_1}^s(f)\tilde{\mathbf{d}}^s(f,\mathbf{p}^s). \tag{54}$$

Using $\mu_2 = \infty$ (i.e., hard constraint on the speech distortion term),

$$\mathbf{W}(f) = \frac{\mathbf{R}^{n^{-1}}(f)\tilde{\mathbf{d}}^s(f,\mathbf{p}^s)}{\tilde{\mathbf{d}}^{s,H}(f,\mathbf{p}^s)\mathbf{R}^{n^{-1}}(f)\tilde{\mathbf{d}}^s(f,\mathbf{p}^s)}. \tag{55}$$

In a GSC-scheme, the hard constraint $\mathbf{W}^H(f)\tilde{\mathbf{d}}^s(f,\mathbf{p}^s) = 1$ is imposed through the fixed beamformer and blocking matrix. The filter $\mathbf{W}(f)$ is then decomposed into a fixed filter $\mathbf{W}_q(f)$ (i.e., the so-called quiescent vector) and an adaptive filter $\mathbf{W}_a(f)$:

$$\mathbf{W}(f) = \mathbf{W}_q + \mathbf{B}(f)\mathbf{W}_a(f) \tag{56}$$

with $\mathbf{W}_q(f) = \frac{1}{M}\tilde{\mathbf{d}}^s(f,\mathbf{p}^s)$ and $\mathbf{B}^H(f)\tilde{\mathbf{d}}^s(f,\mathbf{p}^s) = 0$.

3.1.2 Soft constraint ($\mu_2 \neq \infty$): soft-constrained MWF techniques by Nordholm et al.

Cost function

$$J(\mathbf{W}) \quad = \quad \mathbf{W}^H(f)\mathbf{R}^n(f)\mathbf{W}(f) + \mu_2 \varepsilon\{(D_m^s(f) - \mathbf{W}^H(f)\mathbf{X}_m^s(f))(D_m^s(f) - \mathbf{W}^H(f)\mathbf{X}_m^s(f))^H\} \tag{57}$$

with $\mu_2 \neq \infty$.

Assumed speech model In [**?**] a fixed model is used for the spatial characteristics $\tilde{\mathbf{H}}^s(f)$ of the speech while the speech PSD $P_{X_1}^s(f)$ is estimated online. The speech source is modeled as an infinite number of (uncorrelated) point sources with true PSD $P_{X_1}^s(f)$ clustered closely in space within a pre-defined area $\mathbf{P}$:

$$\mathbf{X}_m^s(f) \quad = \quad \int_{\mathbf{p}\in\mathbf{P}} X_{m,1}^s(f,\mathbf{p})\tilde{\mathbf{d}}^s(f,\mathbf{p})d\mathbf{p} \tag{58}$$

$$D_m^s(f) \quad = \quad \int_{\mathbf{p}\in\mathbf{P}} X_{m,1}^s(f,\mathbf{p})d\mathbf{p} \tag{59}$$

with

$$\varepsilon\{X_{m,1}^s(f,\mathbf{p}_k)X_{m,1}^{s,*}(f,\mathbf{p}_l)\} = P_{X_1}^s(f)\delta_{kl} \quad \forall \mathbf{p}_k,\mathbf{p}_l \in \mathbf{P}_{pass}. \tag{60}$$

The speech PSD $P_{X_1}^s(f)$ is estimated online. To separate the estimation of the spectral and spatial characteristics, the technique is implemented in the frequency-domain.

Noise model The noise model is estimated online.

Solution The filter $\mathbf{W}(f)$ equals

$$W(f) = (\mu_2 \mathbf{R}^s_m(f) + \mathbf{R}^n(f))^{-1} \mu_2 \varepsilon\{\mathbf{X}^s_m(f) D^{s,*}_m(f)\}. \tag{61}$$

In the soft-constrained MWF techniques by Nordholm et al., μ_2 is set to 1[**?**].
Assuming uncorrelated point sources, $\mathbf{R}^s_m$ and $\varepsilon\{\mathbf{X}^s_m(f)D^s_m(f)\}$ in (61) can be computed as:

$$\begin{aligned}
\mathbf{R}^s_m &= \varepsilon\{\int_{\mathbf{p}\in\mathbf{P}} X^s_{m,1}(f,\mathbf{p})\tilde{\mathbf{d}}^s(f,\mathbf{p})d\mathbf{p} \int_{\mathbf{p}\in\mathbf{P}} X^{s,*}_{m,1}(f,\mathbf{p})\tilde{\mathbf{d}}^{s,H}(f,\mathbf{p})d\mathbf{p}\}, \\
&= \int_{\mathbf{p}\in\mathbf{P}} \tilde{\mathbf{d}}^s(f,\mathbf{p})\tilde{\mathbf{d}}^{s,H}(f,\mathbf{p})\varepsilon\{X^s_{m,1}(f,\mathbf{p})X^{s,*}_{m,1}(f,\mathbf{p})\}d\mathbf{p}, \\
&= P^s_{X_1}(f)\int_{\mathbf{p}\in\mathbf{P}} \tilde{\mathbf{d}}^s(f,\mathbf{p})\tilde{\mathbf{d}}^{s,H}(f,\mathbf{p})d\mathbf{p}. \qquad (62) \\
\varepsilon\{\mathbf{X}^s_m(f)D^s_m(f)\} &= P^s_{X_1}(f)\int_{\mathbf{p}\in\mathbf{P}} \tilde{\mathbf{d}}^s(f,\mathbf{p})d\mathbf{p}, \qquad (63)
\end{aligned}$$

where $P^s_{X_1}(f)$ is estimated online.
Instead of using a mathematical speech model, the speech correlation matrix $\mathbf{R}^s_m(f)$ and the cross-correlation $\varepsilon\{\mathbf{X}^s_m(f)D^{s,*}_m(f)\}$ can also be computed based on calibration data [**?**]. This can for example be useful in hearing aid applications where the head shadow effect should be taken into account.

3.2 Combination of online and fixed noise model

Instead of using only an online estimate of the noise model, the online estimated noise model can be combined with a predefined fixed noise model. This may be useful to increase robustness of the noise reduction algorithm against model errors (e.g., sensitivity-constrained GSC [**?**]) and/or VAD failures or when the location of some interfering sources is known a-prior (e.g., echo or feedback in case of a set-up with fixed loudspeaker and microphone positions [**?**]).

Cost function

$$\begin{aligned}
J(\mathbf{W}) &= (1-\lambda)\mathbf{W}^H(f)\mathbf{R}^n(f)\mathbf{W}(f) + \lambda\mathbf{W}^H(f)\mathbf{R}^n_m(f)\mathbf{W}(f) \\
&\quad +\mu_2\{(D^s_m(f) - \mathbf{W}^H(f)\mathbf{X}^s_m(f))(D^s_m(f) - \mathbf{W}^H(f)\mathbf{X}^s_m(f))^H\},
\end{aligned} \tag{64}$$

with $\lambda > 0$.

3.2.1 Hard constraint ($\mu_2 = \infty$)

Examples

- Sensitivity-constrained GSC [**?**, **?**]

  In [**?**, **?**] the robustness of the GSC against model errors is increased by inserting spatially incorrelated noise. This corresponds to a fixed noise model for spatially uncorrelated noise

  $$\mathbf{R}^n_m(f) = P_N(f)\mathbf{I}_m, \tag{65}$$

  with $P^n_{X_{m,i}}(f) = P_N(f)$, $i = 1,\ldots,M$.

Alternatives

- Alternatively, the online noise estimate of the GSC can be combined with a diffuse noise model or a model of noise in the back hemisphere to prevent amplification of sounds coming from the back. In addition, noise sources with a fixed, known location can be included (e.g., echo or feedback in the case of a fixed loudspeaker-microphone position).

3.2.2 Soft constraint ($\mu_2 \neq \infty$)

Examples

- In [**?**], a model based on calibration signals is used for the speech signal and the echo signal, while the noise statistics are estimated online.

Alternatives

- Combination of an online noise estimate and a pre-defined model of noise (e.g., noise in the back hemisphere or diffuse noise). This technique may improve the performance of the soft-constrained MWF in the presence of VAD failures.
- Combination of an online noise estimate with a pre-defined model for echo or feedback. This approach may be useful when the transfer function from the loudspeaker to the microphone can assumed to be fixed.

4 Online speech model

In the above mentioned techniques, a speech model is always used in the computation of the filters $\mathbf{W}(f)$. In this speech model, assumptions are made about the microphones (microphone characteristics, positions), the location of the desired speaker and the room acoustics (e.g., no reverberation). These assumptions are often violated in practice so that the performance may be suboptimal. In this section, techniques that use an online estimate of both the speech statistics are discussed. Since the source signal $S(f)$ is unknown, these techniques estimate the speech component in one of the microphones (e.g., the first microphone), i.e., $D^s(f) = X_1^s(f)$ or the speech component in the output of a fixed beamformer. These techniques typically exploit a VAD mechanism and assume the noise statistics to be more stationary than the speech statistics. VAD errors or highly non-stationary noise may then affect the performance.

4.1 Online noise model

4.1.1 Hard constraint ($\mu_1 = \infty$): Transfer Function LCMV (TF-LCMV)[?]

Cost function

$$J(\mathbf{W}) \quad = \quad \mathbf{W}^H(f)\mathbf{R}^n(f)\mathbf{W}(f) + \mu_1 \varepsilon\{(D^s(f) - \mathbf{W}^H(f)\mathbf{X}^s(f))(D^s(f) - \mathbf{W}^H(f)\mathbf{X}^s(f))^H\} \tag{66}$$

with $\mu_1 = \infty$.

Solution Minimizing (66) with $\mu_1 = \infty$ results in (cf. the derivation in Section 3.1.1)

$$\mathbf{W}(f) = \frac{\mathbf{R}^{n^{-1}}(f)\tilde{\mathbf{H}}^s(f)}{\tilde{\mathbf{H}}^{s,H}(f)\mathbf{R}^{n^{-1}}(f)\tilde{\mathbf{H}}^s(f)}, \tag{67}$$

with $\tilde{\mathbf{H}}^s(f)$ is the relative transfer function ratio vector defined in (5).

Like in the GSC, the hard constraint $D^s(f) = X_1^s(f) = \mathbf{W}^H\mathbf{X}^s(f)$ can be imposed through the use of a spatial pre-processor. The fixed beamformer $\mathbf{W}_q(f)$ equals the matched filter

$$\mathbf{W}_q(f) = \frac{\tilde{\mathbf{H}}^s(f)}{\left\|\tilde{\mathbf{H}}^s(f)\right\|^2} \tag{68}$$

The blocking matrix $\mathbf{B}(f)$ is chosen such that $\mathbf{B}^H(f)\mathbf{W}_q(f) = 0$, e.g.,

$$\mathbf{B}(f) = \begin{bmatrix} -\frac{H_2^s(f)}{H_1^s(f)} & -\frac{H_3^s(f)}{H_1^s(f)} & \cdots & -\frac{H_M^s(f)}{H_1^s(f)} \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{69}$$

Remark: The GSC with switching adaptive filters [?] and the GSC with adaptive blocking matrix [?, ?] also belong to this class. In [?, ?], the output of a fixed beamformer is used as a reference signal instead of the first microphone signal $X_1^s(f)$.

4.1.2 Soft constraint ($\mu_1 \neq \infty$): speech distortion weighted MWF (SDW-MWF)[?]

Cost function

$$J(\mathbf{W}) \quad = \quad \mathbf{W}^H(f)\mathbf{R}^n(f)\mathbf{W}(f) + \mu_1 \varepsilon\{(D^s(f) - \mathbf{W}^H(f)\mathbf{X}^s(f))(D^s(f) - \mathbf{W}^H(f)\mathbf{X}^s(f))^H\} \tag{70}$$

with $\mu_1 \neq \infty$ and $D^s(f) = X_1^s(f)$.

Solution Minimizing (70) results in

$$\mathbf{W}(f) = (\mathbf{R}^n(f) + \mu_1 \mathbf{R}^s(f))^{-1} \mu_1 \varepsilon\{\mathbf{X}^s(f) X_1^{s,H}(f)\}. \tag{71}$$

Comparison with TF-LCMV Assuming a rank-one model for $\mathbf{R}^s(f)$, i.e,

$$\mathbf{R}^s(f) = P_{X_1}^s(f) \tilde{\mathbf{H}}^s(f) \tilde{\mathbf{H}}^{s,H}(f), \tag{72}$$

$\mathbf{W}(f)$ can be decomposed into a TF-LCMV with a single-channel post-filter:

$$\mathbf{W}(f) = \underbrace{\frac{1}{\tilde{\mathbf{H}}^s(f)\mathbf{R}^{n^{-1}}(f)\tilde{\mathbf{H}}^s(f)} \mathbf{R}^{n^{-1}}(f)\tilde{\mathbf{H}}^s(f)}_{TF-LCMV} \underbrace{\left(\frac{\mu_1 P_{X_1}^s(f)}{\mu_1 P_{X_1}^s(f) + \frac{1}{\tilde{\mathbf{H}}^{s,H}(f)\mathbf{R}^{n^{-1}}\tilde{\mathbf{H}}^s(f)}} \right)}_{post-filter}. \tag{73}$$

Hence, the soft constraint on the speech distortion term introduces spectral filtering of the speech component $X_1^s(f)$ (unless the speech and the noise subspace are orthogonal such that $\frac{1}{\tilde{\mathbf{H}}^{s,H}(f)\mathbf{R}^{n^{-1}}\tilde{\mathbf{H}}^s(f)} = 0$).

Remark:In the SP-SDW-MWF [**?**], the output of a fixed beamformer is estimated instead of the first microphone signal.

4.2 Combination of online and fixed noise model

Like in Section 3.2, a combination of an online and fixed noise model can be used to increase robustness or to include a-priori known noise sources. The cost function then becomes:

$$J(\mathbf{W}) = (1-\lambda)\mathbf{W}^H(f)\mathbf{R}^n(f)\mathbf{W}(f) + \lambda \mathbf{W}^H(f)\mathbf{R}_m^n(f)\mathbf{W}(f) + \mu_1 \varepsilon\{(D^s(f) - \mathbf{W}^H(f)\mathbf{X}^s(f))(D^s(f) - \mathbf{W}^H(f)\mathbf{X}^s(f))^H\} \tag{74}$$

with $\lambda > 0$ and $D^s(f) = X_1^s(f)$.

4.3 Fixed noise model

If λ in (74) is set to 1, only a fixed noise model is used. In scenarios with a moving speech source and a lot of noise sources or reverberation (e.g., cocktail-party), a fixed noise model (e.g., a diffuse noise model) can be useful and more robust (e.g., in case of interfering speakers) than an online noise estimate. The use of a fixed noise model is especially useful in combination with the TF-LCMV (i.e., $\mu_1 = \infty$): the speech source is then preserved, while diffuse or background noise is attenuated as much as possible. In free-field/an anechoic room, the TF-LCMV then corresponds to a fixed LCMV-beamformer with an adaptive steering vector $\tilde{\mathbf{d}}^s(f, \mathbf{p}^s)$ (cf. Section 2.1).

5 Combination of an online and a-priori speech model

So far, either an a-priori model or an online estimate was used for the speech correlation matrix. Like for the noise correlation matrix, also a combination of both, a-priori knowledge and online estimation can be used for $\mathbf{R}^s(f)$. The cost function $J(\mathbf{W})$ then becomes

$$\begin{aligned} J(\mathbf{W}) &= (1-\lambda)\mathbf{W}^H(f)\mathbf{R}^n(f)\mathbf{W}(f) + \lambda \mathbf{W}^H(f)\mathbf{R}_m^n(f)\mathbf{W}(f) \\ &+ \mu_1 \varepsilon\{(D^s(f) - \mathbf{W}^H(f)\mathbf{X}^s(f))(D^s(f) - \mathbf{W}^H(f)\mathbf{X}^s(f))^H\} + \mu_2 \varepsilon\{(D_m^s(f) - \mathbf{W}^H(f)\mathbf{X}_m^s(f))(D_m^s(f) - \mathbf{W}^H(f)\mathbf{X}_m^s(f))^H\} \end{aligned} \tag{75}$$

5.1 Online noise model

Cost function

$$\begin{aligned} J(\mathbf{W}) &= \mathbf{W}^H(f)\mathbf{R}^n(f)\mathbf{W}(f) + \mu_1 \varepsilon\{(D^s(f) - \mathbf{W}^H(f)\mathbf{X}^s(f))(D^s(f) - \mathbf{W}^H(f)\mathbf{X}^s(f))^H\} \\ &+ \mu_2 \varepsilon\{(D_m^s(f) - \mathbf{W}^H(f)\mathbf{X}_m^s(f))(D_m^s(f) - \mathbf{W}^H(f)\mathbf{X}_m^s(f))^H\}. \end{aligned} \tag{76}$$

5.1.1 Hard constraint on (constrained) online speech model ($\mu_1 = \infty$)

- **Special case: combination TF-LCMV and GSC**

  To increase the robustness of the online estimation of the relative transfer function ratios $\tilde{\mathbf{H}}^s(f)$in TF-LCMV, the estimated relative transfer function ratios can be constrained such that they do not deviate too much from an a-priori transfer function ratio $\tilde{\mathbf{d}}^s(f)$.

  E.g., assume an a-priori speech model and only allow small deviations from this model, i.e.,

$$\mathbf{X}^s(f) = (\underbrace{\tilde{\mathbf{d}}^s(f)}_{\text{fixed}} + \underbrace{\delta\tilde{\mathbf{d}}^s(f)}_{\text{adaptive}})X_1^s(f) \tag{77}$$

  with a constraint on $\delta\tilde{\mathbf{d}}^s(f)$ (e.g., $\delta\tilde{\mathbf{d}}^{s,H}(f)\delta\tilde{\mathbf{d}}^s(f) \leq \zeta^2$).

  A possible realization of this approach is to use constrained adaptation of the adaptive blocking matrix and/or fixed beamformer proposed in [**?**, **?**, **?**] (cf. the constrained adaptation techniques proposed by Kates for feedback cancellation [**?**]).

5.1.2 Soft Constraint on online speech model ($\mu_1 \neq \infty$)

- **Hard constraint on fixed speech model: speech distortion regularized GSC (SDR-GSC)**

  In the SDR-GSC, the speech model of the GSC is used, i.e.,

$$\mathbf{X}_m^s(f) = \tilde{\mathbf{d}}^s(f,\mathbf{p}^s)X_{m,1}^s(f) \tag{78}$$
$$D_m^s(f) = X_{m,1}^s(f) \tag{79}$$

  and $\mu_2 = \infty$. The hard constraint of GSC is imposed by the fixed beamformer $\mathbf{W}_q(f) = \frac{\mathbf{d}(f)}{M}$ and the blocking matrix $\mathbf{B}(f)$ with $\mathbf{B}^H(f)\mathbf{W}_q(f) = 0$, i.e.,

$$\mathbf{W}(f) = \mathbf{W}_q(f) + \mathbf{B}(f)\mathbf{W}_a(f). \tag{80}$$

  In addition to the pre-defined speech model, an online estimate of the speech component in the speech reference $D^s(f) = \mathbf{W}_q^H\mathbf{X}^s(f)$ is made ($\mu_1 \neq \infty$).

  Using (80), the online estimated speech distortion term in (76) can be written as:

$$\varepsilon\{(D^s(f) - \mathbf{W}^H(f)\mathbf{X}^s(f))(D^s(f) - \mathbf{W}^H(f)\mathbf{X}^s(f))^H\} = \varepsilon\{\mathbf{W}_a^H(f)\mathbf{B}^H(f)\mathbf{X}^s(f)\mathbf{X}^{s,H}(f)\mathbf{B}(f)\mathbf{W}_a(f)\}. \tag{81}$$

  Using (81) in (76), results in the SDR-GSC cost function proposed in [**?**].

- **Soft constraint on fixed speech model: combination of SDW-MWF [?] and soft constrained MWF [?]**

  The online estimate of the speech correlation matrix (as used in the SDW-MWF) can be combined with a pre-defined (or pre-computed) speech model as proposed by Nordholm (cf. Section 3.1.2). By doing so, a priori knowledge is combined with an online estimation of the speech correlation matrix based on incoming data. The pre-defined model acts as a regularization of the online speech correlation matrix. This method allows for a (partial) update of the speech correlation matrix while it is expected to increase robustness of the SDW-MWF to an erroneous estimation of the speech correlation matrix (e.g., due to VAD failures).

5.2 Combination of online and fixed noise model or fixed noise model

The techniques described in Section 4.2 and Section 4.3 can also be used in combination with a (partially) fixed speech model. This leads to the most general cost function $J(\mathbf{W})$:

$$\begin{aligned} J(\mathbf{W}) &= (1-\lambda)\mathbf{W}^H(f)\mathbf{R}^n(f)\mathbf{W}(f) + \lambda\mathbf{W}^H(f)\mathbf{R}_m^n(f)\mathbf{W}(f) \\ &+\mu_1\varepsilon\{(D^s(f) - \mathbf{W}^H(f)\mathbf{X}^s(f))(D^s(f) - \mathbf{W}^H(f)\mathbf{X}^s(f))^H\} + \mu_2\varepsilon\{(D_m^s(f) - \mathbf{W}^H(f)\mathbf{X}_m^s(f))(D_m^s(f) - \mathbf{W}^H(f)\mathbf{X}_m^s(f))^H\}. \end{aligned} \tag{82}$$

6 Conclusions

In this paper, a general cost function for multi-microphone noise reduction has been proposed. From this cost function, most existing multi-microphone noise reduction techniques (e.g., fixed beamforming, GSC, TF-LCMV, SDW-MWF) can be derived. In addition, some novel noise reduction techniques have been suggested, the most interesting ones being:

- Combination of soft-constrained MWF and SDW-MWF;
- Combination of TF-LCMV and GSC;
- TF-LCMV with a fixed noise model;
- Combination of fixed and adaptive noise model.

Basically, the noise reduction techniques differ from each other in the use of pre-defined or online estimated speech and/or noise models and the use of a soft or hard constraint on the amount of speech distortion.

References

[1] J. Bitzer and K. U. Simmer. *Superdirective Microphone Arrays*, chapter 2 in “Microphone Arrays: Signal Processing Techniques and Applications” (Brandstein, M. S. and Ward, D. B., Eds.), pages 19–38. Springer-Verlag, May 2001.

[2] K. M. Buckley. Broad-band beamforming and the Generalized Sidelobe Canceller. 34(5):1322–1323, October 1986.

[3] H. Cox, R. M. Zeskind, and T. Kooij. Practical Supergain. 34(3):393–398, June 1986.

[4] H. Cox, R. M. Zeskind, and M. M. Owen. Robust Adaptive Beamforming. 35(10):1365–1376, 1987.

[5] S. Doclo and M. Moonen. Design of far-field and near-field broadband beamformers using eigenfilters. 83(12):2641–2673, December 2003.

[6] G. W. Elko. *Superdirectional microphone arrays*, chapter 10 in "Acoustic Signal Processing for Telecommunications" (Gay, S. L. and Benesty, J., Eds.), pages 181–237. Kluwer Academic Publishers, 2000.

[7] S. Gannot, D. Burshtein, and E. Weinstein. Signal enhancement using beamforming and non-stationarity with applications to speech. 49(8):1614–1626, August 2001.

[8] L. J. Griffiths and C. W. Jim. An alternative approach to linearly constrained adaptive beamforming. 30(1):27–34, January 1982.

[9] W. Herbordt and W. Kellermann. *Adaptive Beamforming for Audio Signal Acquisition*, chapter 6 in "Adaptive Signal Processing: Applications to Real-World Problems" (Benesty, J. and Huang, Y., Eds.), pages 155–188. Springer-Verlag, 2003.

[10] O. Hoshuyama, A. Sugiyama, and A. Hirano. A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters. 47:2677–2683, 1999.

[11] N. K. Jablon. Adaptive beamforming with the Generalized Sidelobe Canceller in the presence of array imperfections. 34:996–1012, August 1986.

[12] M. Kajala and M. Hämäläinen. Broadband beamforming optimization for speech enhancement in noisy environements. pages 19–22, New Paltz, NY, USA, October 1999.

[13] J. M. Kates. Superdirective arrays for hearing aids. 94(4):1930–1933, October 1993.

[14] J. M. Kates. *Adaptive feedback cancellation in hearing aids*, chapter 2 in "Adaptive signal processing: Applications to real-world problems" (Benesty, J. and Huang, Y.), pages 23–55. Springer-Verlag, 2003.

[15] D. Korompis, K. Yao, and F. Lorenzelli. Broadband maximum energy array with user imposed spatial and frequency constraints. pages 529–532, Adelaide, Australia, April 1994.

[16] S. Nordebo, I. Claesson, and S. Nordholm. Weighted Chebyshev approximation for the design of broadband beamformers using quadratic programming. 1(7):103–105, July 1994.

[17] S. Nordholm, I. Claesson, and M. Dahl. Adaptive microphone array employing calibration signals: an analytical evaluation. 7(3):241–252, May 1999.

[18] S. Nordholm, H.Q. Dam, N. Grbíc, and S. Y. Low. *Adaptive microphone array employing spatial quadratic soft constraints and spectral shaping*, chapter 10 in “Speech Enhancement” (Benesty, J., Makino, S. and Chen, J.,Eds.), pages 229–246. Springer-Verlag, 2005.

[19] S. Nordholm, K. Rehbock, and S. Nordebo. Chebyshev optimization for the design of broadband beamformers in the near field. 45(1):141–143, January 1998.

[20] A. Spriet, M. Moonen, and J. Wouters. Spatially pre-processed speech distortion weighted multi-channel Wiener filtering for noise reduction. 84(12):2367–2387, December 2004.

[21] S. C. Thompson. Directional patterns obtained from two or three microphones. Technical report, Knowles Electronics, Illinois, USA, September 2000.

[22] D. Van Compernolle. Switching adaptive filters for enhancing noisy and reverberant speech from microphone array recordings. volume 2, pages 833–836, Albuquerque, April 1990.

[23] K.F.C. Yiu, N. Grbíc, K.-L. Teo, and S. Nordholm. A new design method for broadband microphone arrays for speech input in automobiles. 9(7):222–224, July 2002.