Efficient frequency-domain implementation of speech distortion weighted multi-channel Wiener filtering for noise reduction

Simon Doclo **Ann Spriet** **Marc Moonen**

Katholieke Universiteit Leuven - Dept. of Electrical Engineering (ESAT - SCD)

Kasteelpark Arenberg 10, B-3001 Heverlee (Leuven), Belgium

1 Multi-microphone noise reduction techniques

- reduction of noise wrt useful speech signal in different acoustic environments
- exploit spatial + spectral information of speech and noise sources
- for **small-size microphone arrays** (e.g. hearing instruments) → increased sensitivity to signal model errors (e.g. microphone mismatch)

Adaptive beamforming (GSC) : not robust against signal model errors

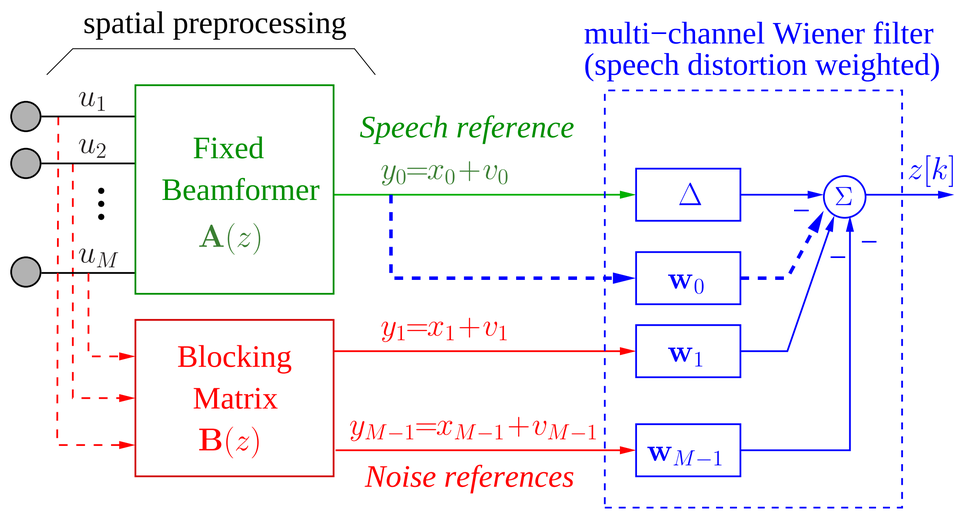
⇓

Robust generalised multi-microphone noise reduction scheme

⇓

Efficient FD implementation using stochastic gradient algorithm

2 Spatially pre-processed SDW-MWF



- **Structure of SP-SDW-MWF** resembles Generalised Sidelobe Canceller (GSC):
  - *spatial pre-processor* → speech reference and noise references
  - *adaptive stage*: estimation of noise component in speech reference $v_0[k-\Delta]$
- **Standard GSC** minimises output noise power : $\min \mathcal{E}\left\{\left|v_0[k-\Delta]-\mathbf{w}^T[k]\mathbf{v}[k]\right|^2\right\}$
- Fixed + adaptive stage rely on assumptions (no mismatch, no reverberation), but in practice these assumptions are not satisfied ⇒ **speech distortion**
  - distortion of speech component $x_0[k]$ in speech reference
  - speech leakage $\mathbf{x}[k]$ into noise references

$$z_x[k]=x_0[k-\Delta]-\mathbf{w}^T[k]\mathbf{x}[k]$$

- Design of robust noise reduction algorithm:
  - robust spatial pre-processor : calibration, microphone statistics, WNG
  - robust adaptive stage : limit effect of speech leakage $\mathbf{w}^T[k]\mathbf{x}[k]$
    - Quadratic Inequality Constraint (QIC): $||\mathbf{w}[k]|| < \beta$ → conservative
    - **SDW-MWF**: take speech distortion into account in optimisation criterion

$$\min_{\mathbf{w}[k]} \mathcal{E}\left\{\left|v_0[k-\Delta]-\mathbf{w}^T[k]\mathbf{v}[k]\right|^2\right\}+\frac{1}{\mu}\mathcal{E}\left\{\left|\mathbf{w}^T[k]\mathbf{x}[k]\right|^2\right\}$$

      - $1/\mu$ trades off noise reduction and speech distortion ($1/\mu=0 \rightarrow$ GSC)
      - regularisation term ∼ amount of speech leakage
- Wiener solution (using $\mathcal{E}\left\{\mathbf{x}[k]\mathbf{x}^T[k]\right\}=\mathcal{E}\left\{\mathbf{y}[k]\mathbf{y}^T[k]\right\}-\mathcal{E}\left\{\mathbf{v}[k]\mathbf{v}^T[k]\right\}$)

$$\mathbf{w}[k]=\left[\frac{1}{\mu}\mathcal{E}\left\{\mathbf{y}[k]\mathbf{y}^T[k]\right\}+\left(1-\frac{1}{\mu}\right)\mathcal{E}\left\{\mathbf{v}[k]\mathbf{v}^T[k]\right\}\right]^{-1}\mathcal{E}\left\{\mathbf{v}[k]v_0[k-\Delta]\right\}$$

- **Generalised scheme** → different algorithms, depending on $1/\mu$ and $\mathbf{w}_0$:
  - Without $\mathbf{w}_0$: **Speech Distortion Regularised GSC** (SDR-GSC), i.e. standard ANC criterion is supplemented with regularisation term
  - With $\mathbf{w}_0$: **Spatially pre-processed SDW-MWF** (SP-SDW-MWF)

Compared to QIC-GSC, the SDR-GSC and SP-SDW-MWF obtain a better noise reduction performance for a given maximum speech distortion level

3 Stochastic Gradient algorithms

3.1 Time-Domain implementation

- LMS-based updating formula (operated during noise-only periods)

$$\mathbf{w}[k+1]=\underbrace{\mathbf{w}[k]+\rho\left[\mathbf{v}[k](v_0[k-\Delta]-\mathbf{v}^T[k]\mathbf{w}[k])\right.}_{\text{Classical GSC}}\underbrace{\left.-\frac{1}{\mu}\mathbf{x}[k]\mathbf{x}^T[k]\mathbf{w}[k]\right]}_{\mathbf{r}[k]}$$

- Regularisation term is unknown (instantaneous correlation matrix $\mathbf{x}[k]\mathbf{x}^T[k]$)
  - Store samples in **circular buffer** $\mathbf{B}_y$ during speech-periods and approximate $\mathbf{r}[k]$ by using average clean speech correlation matrix $\mathcal{E}\{\mathbf{x}[k]\mathbf{x}^T[k]\}$

$$\mathbf{r}[k]=\frac{1}{\mu}(1-\bar{\lambda})\sum_{l=0}^{k}\bar{\lambda}^{k-l}\left[\mathbf{y}_{B_y}[l]\mathbf{y}_{B_y}^T[l]-\mathbf{v}[l]\mathbf{v}^T[l]\right]\cdot\mathbf{w}[k]$$

  - Avoid expensive matrix operations: assume $\mathbf{w}[k]\approx\mathbf{w}[l]$

$$\mathbf{r}[k]=\bar{\lambda}\mathbf{r}[k-1]+(1-\bar{\lambda})\frac{1}{\mu}\left[\mathbf{y}_{B_y}[k]\mathbf{y}_{B_y}^T[k]-\mathbf{v}[k]\mathbf{v}^T[k]\right]\mathbf{w}[k]$$

3.2 Frequency-Domain implementation

- Block-based implementation: fast convolution and correlation (FFT)
  - ⊕ Complexity reduction + tuning of ρ and $1/\mu$ per frequency
  - ⊖ Large memory requirement due to data buffers
- Substantial memory (+ computational complexity) reduction:
  - Use (non-diagonal) **correlation matrices** for computing regularisation term

$$\mathbf{S}_y^{ij}[k]=(1-\lambda)\sum_{l=0}^{k}\lambda^{k-l}\mathbf{Y}_i^H[l]\mathbf{F}\mathbf{k}^T\mathbf{k}\mathbf{F}^{-1}\mathbf{Y}_j[l],\ \mathbf{S}_v^{ij}[k]=(1-\lambda)\sum_{l=0}^{k}\lambda^{k-l}\mathbf{V}_i^H[l]\mathbf{F}\mathbf{k}^T\mathbf{k}\mathbf{F}^{-1}\mathbf{V}_j[l]$$

$$\mathbf{R}_i[k]=\frac{1}{\mu}\sum_{j=M-N}^{M-1}\left[\mathbf{S}_y^{ij}[k]-\mathbf{S}_v^{ij}[k]\right]\mathbf{W}_j[k],\quad i,j=M-N\ldots M-1$$

  - Approximation: $\boxed{\mathbf{F}\mathbf{k}^T\mathbf{k}\mathbf{F}^{-1}\approx\mathbf{I}_{2L}/2}$ → diagonal correlation matrices

$$\mathbf{S}_y^{ij}[k]=\lambda\mathbf{S}_y^{ij}[k-1]+(1-\lambda)\mathbf{Y}_i^H[k]\mathbf{Y}_j[k]/2$$

3.3 Computational complexity and memory

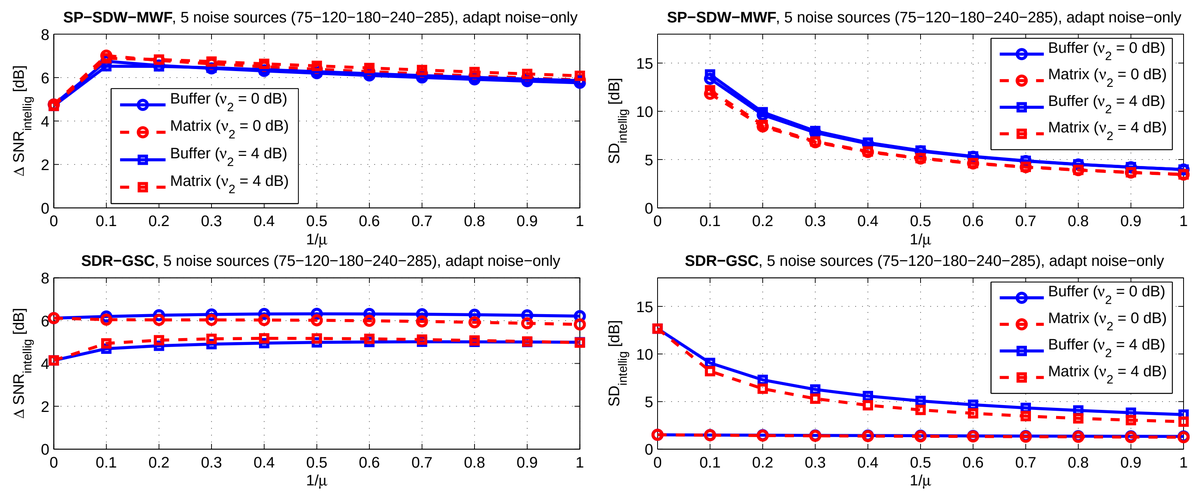
- Computational complexity of buffer-based and matrix-based implementation are comparable (for small number of input channels N)
- Considerable **reduction of memory usage** using matrix-based implementation

Algorithm	**Complexity (MAC)**	**MIPS**
QIC-GSC	$(3M-1)$FFT $+16M-9$	2.16
SP-SDW-MWF (Buffer)	$(3N+5)$FFT $+30N+10$	$3.22^{(a)}$, $4.27^{(b)}$
SP-SDW-MWF (Matrix)	$(3N+2)$FFT $+10N^2+15N+4$	$2.71^{(a)}$, $4.31^{(b)}$
	Memory	**kWords**
QIC-GSC	$4(M-1)L+6L$	0.45
SP-SDW-MWF (Buffer)	$2NL_y+6LN+7L$	$40.61^{(a)}$, $60.80^{(b)}$
SP-SDW-MWF (Buffer)	$4LN^2+6LN+7L$	$1.12^{(a)}$, $1.95^{(b)}$

($M=3$, $L=32$, $f_s=16$ kHz, $L_y=10000$, (a) $N=M-1$, (b) $N=M$)

4 Simulations

- **Set-up**: 3-microphone BTE hearing aid ($d=1$ cm, 1.5 cm) on dummy head
  - speech in front ($0°$) + 5 speech-like noise sources ($75°$, $120°$, $180°$, $240°$, $285°$)
  - Microphone gain mismatch $\nu_2=4$ dB at the second microphone
- **Performance measures**: Intelligibility weighted signal-to-noise ratio improvement $\Delta\text{SNR}_{\text{intellig}}$ and spectral distortion $\text{SD}_{\text{intellig}}$
  - GSC ($1/\mu=0$, no $\mathbf{w}_0$) : degraded performance if significant leakage
  - SDR-GSC : $1/\mu>0$ increases robustness (speech distortion ↔ noise reduction)
  - SP-SD-MWF ($\mathbf{w}_0$) : performance practically not degraded by mismatch



Small performance difference between buffer-based and matrix-based implementation (different assumptions)