Extension of the multi-channel Wiener filter with localization cues for noise reduction in binaural hearing aids

S. Doclo, R. Dong, T. Klasen, J. Wouters, S. Haykin, M. Moonen

Katholieke Universiteit Leuven, Belgium - Dept. Elec. Engineering (SCD), Lab. Exp. ORL

McMaster University, Canada - Adaptive Systems Laboratory

1 Binaural hearing aids

- Hearing impairment → reduction of speech intelligibility in background noise
  - signal processing to **selectively enhance/extract useful speech signal**
  - many hearing impaired are fitted with a hearing aid at both ears
- **Binaural auditory cues**:
  - Interaural Time Difference (ITD) - Interaural Level Difference (ILD)
  - binaural cues, in addition to spectral and temporal cues, play an important role in binaural noise reduction and sound localization
- Bilateral system: independent processing → binaural cues are not preserved
- **Binaural system: cooperation between left and right hearing aid**

Objectives/requirements for binaural processing:

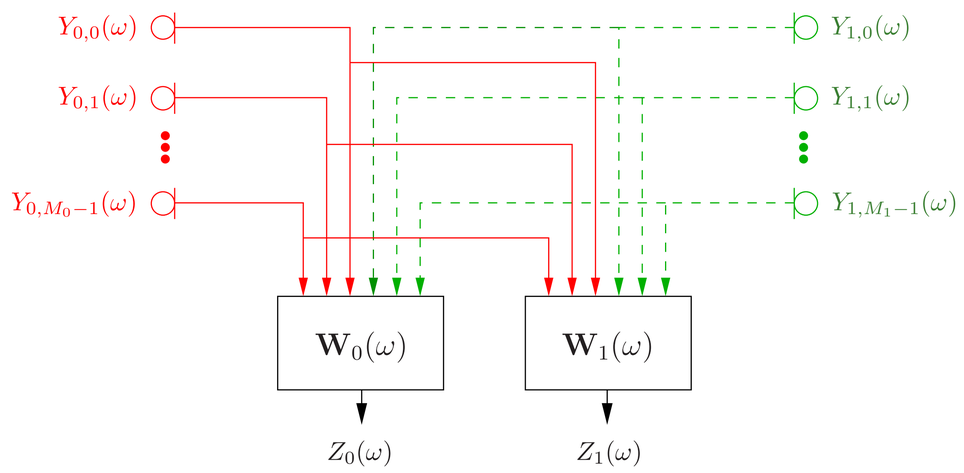
1. **SNR improvement: noise reduction, limit speech distortion**
2. **Preservation of binaural cues to exploit binaural hearing advantage**
3. **No assumptions about position of speech source and microphones**

2 Overview of binaural noise reduction techniques

- **Fixed beamforming**: spatial selectivity + preservation of binaural speech cues
  - maximize directivity index while restricting ITD error [Desloge, 1997]
  - superdirective beamformer using HRTFs [Lotter, 2004]
    - ⊖ broadside array, limited performance, assumptions about geometry
- **Adaptive beamforming**: based on Generalized Sidelobe Canceller structure
  - divide frequency spectrum: low-pass portion unaltered to preserve binaural cues (ITD), high-pass portion processed using GSC [Welker, 1997]
    - ⊖ low-pass: no noise reduction, high-pass: no preservation of binaural cues
  - TF-GSC: minimize output energy, constraint: speech component in output signal = speech component in reference mic signal (both ears) [Gannot, 2001]
    - ⊖ binaural noise cues may be distorted
- **Multi-channel Wiener filter** (MWF) [Doclo, Spriet, Klasen, Wouters, Moonen]
  - MMSE estimate of speech component in reference mic signal at both ears → binaural speech cues are preserved, binaural noise cues may be distorted

Extension of MWF: preservation of binaural speech and noise cues without significantly compromising noise reduction performance

3 Binaural multi-channel Wiener filter



- Configuration: microphone array at left and right hearing aid

$$Y_{0,m}(\omega) = X_{0,m}(\omega) + V_{0,m}(\omega), \quad m = 0 \dots M_0 - 1$$

$$\mathbf{Y}(\omega) = \left[Y_{0,0}(\omega) \dots Y_{0,M_0-1}(\omega) \quad Y_{1,0}(\omega) \dots Y_{1,M_1-1}(\omega) \right]^T$$

- Cooperation between hearing aids: use all available microphone signals to generate **output signal at both ears** → computation of filters $\mathbf{W}_0(\omega)$ and $\mathbf{W}_1(\omega)$

$$Z_0(\omega) = \mathbf{W}_0^H(\omega)\mathbf{Y}(\omega), \quad Z_1(\omega) = \mathbf{W}_1^H(\omega)\mathbf{Y}(\omega), \quad \mathbf{W}(\omega) = \begin{bmatrix} \mathbf{W}_0(\omega) \\ \mathbf{W}_1(\omega) \end{bmatrix}$$

- **SDW-MWF**: estimate speech component in reference microphone signal at both ears; additional trade-off between noise reduction and speech distortion

$$J_{SDW,0} = \mathcal{E}\left\{ |X_{0,r_0} - \mathbf{W}_0^H\mathbf{X}|^2 \right\} + \mu_0 \mathcal{E}\left\{ |\mathbf{W}_0^H\mathbf{V}|^2 \right\}$$

$$J_{SDW}(\mathbf{W}) = J_{SDW,0} + J_{SDW,1} = P + \mathbf{W}^H\mathbf{R}\mathbf{W} - \mathbf{W}^H\mathbf{r} - \mathbf{r}^H\mathbf{W} \Rightarrow \mathbf{W}_{SDW} = \mathbf{R}^{-1}\mathbf{r}$$

$$P = P_0 + P_1, \quad \mathbf{r} = \begin{bmatrix} \mathbf{r}_{x0} \\ \mathbf{r}_{x1} \end{bmatrix}, \quad \mathbf{R} = \begin{bmatrix} \mathbf{R}_x + \mu_0\mathbf{R}_v & \mathbf{0}_M \\ \mathbf{0}_M & \mathbf{R}_x + \mu_1\mathbf{R}_v \end{bmatrix}, \quad \mathbf{R}_x = \mathbf{R}_y - \mathbf{R}_v$$

- Binaural speech cues are generally preserved, noise cues may be distorted

4 Preservation of binaural noise cues

4.1 Partial estimation of noise component [Klasen, 2005]

- MMSE estimate of sum of speech component and scaled noise component

$$\bar{J}_{MSE,0}(\mathbf{W}_0) = \mathcal{E}\left\{ |(X_{0,r_0} + \lambda_0 V_{0,r_0}) - \mathbf{W}_0^H\mathbf{Y}|^2 \right\}$$

  - ⊖ considerable reduction of noise reduction performance

4.2 Extension of SDW-MWF with binaural cues

- Add term related to ITD and ILD cue of noise component to SDW cost function

$$J_{tot}(\mathbf{W}) = J_{SDW}(\mathbf{W}) + \beta \underbrace{|ITD_{out}(\mathbf{W}) - ITD_{des}|^2}_{J_{ITD}(\mathbf{W})} + \gamma \underbrace{|ILD_{out}(\mathbf{W}) - ILD_{des}|^2}_{J_{ILD}(\mathbf{W})}$$

  → link computation of filters $\mathbf{W}_0$ and $\mathbf{W}_1$
- **ITD: phase of cross-correlation** between two signals

  output: $\mathcal{E}\{Z_{v0}Z_{v1}^*\} = \mathbf{W}_0^H\mathbf{R}_v\mathbf{W}_1$ → *input*: $s = \mathcal{E}\{V_{0,r_0}V_{1,r_1}^*\} = \mathbf{R}_v(r_0, r_1)$

  Cost function: cosine of phase difference $\phi(\mathbf{W})$ between cross-correlations

$$J_{ITD}(\mathbf{W}) = 1 - \cos\left(\phi(\mathbf{W})\right) = 1 - \frac{s_R\,(\mathbf{W}_0^H\mathbf{R}_v\mathbf{W}_1)_R + s_I\,(\mathbf{W}_0^H\mathbf{R}_v\mathbf{W}_1)_I}{\sqrt{s_R^2 + s_I^2}\sqrt{(\mathbf{W}_0^H\mathbf{R}_v\mathbf{W}_1)_R^2 + (\mathbf{W}_0^H\mathbf{R}_v\mathbf{W}_1)_I^2}}$$

- **ILD: power ratio** of two signals

  output: $\frac{\mathcal{E}\{|Z_{v0}|^2\}}{\mathcal{E}\{|Z_{v1}|^2\}} = \frac{\mathbf{W}_0^H\mathbf{R}_v\mathbf{W}_0}{\mathbf{W}_1^H\mathbf{R}_v\mathbf{W}_1}$ → *input*: $P = \frac{\mathcal{E}\{|V_{0,r_0}|^2\}}{\mathcal{E}\{|V_{1,r_1}|^2\}} = \frac{\mathbf{R}_v(r_0,r_0)}{\mathbf{R}_v(r_1,r_1)}$

$$J_{ILD}(\mathbf{W}) = \left[\frac{\mathbf{W}_0^H\mathbf{R}_v\mathbf{W}_0}{\mathbf{W}_1^H\mathbf{R}_v\mathbf{W}_1} - P \right]^2$$

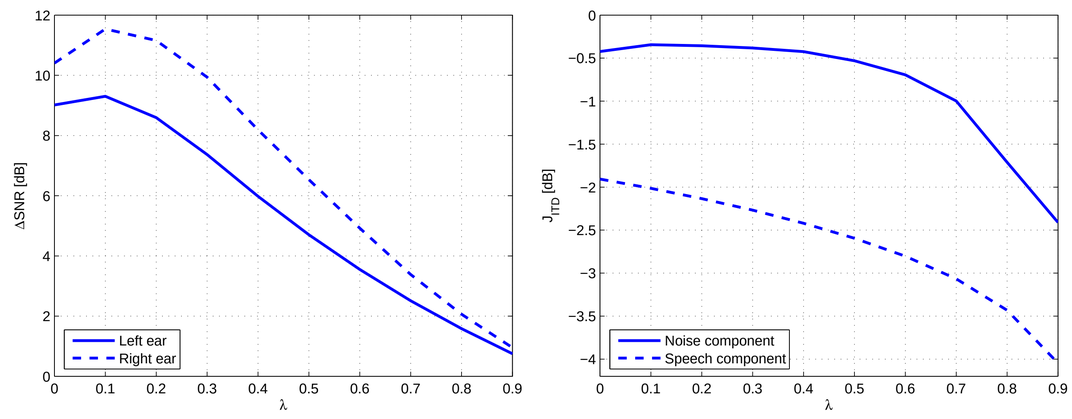
- Other possibility: specify desired angle θ_v and use HRTFS:

$$s(\omega) = \mathrm{HRTF}_0(\omega, \theta_v)\,\mathrm{HRTF}_1^*(\omega, \theta_v), \quad P = |\mathrm{HRTF}_0(\omega, \theta_v)|^2 / |\mathrm{HRTF}_1^*(\omega, \theta_v)|^2$$

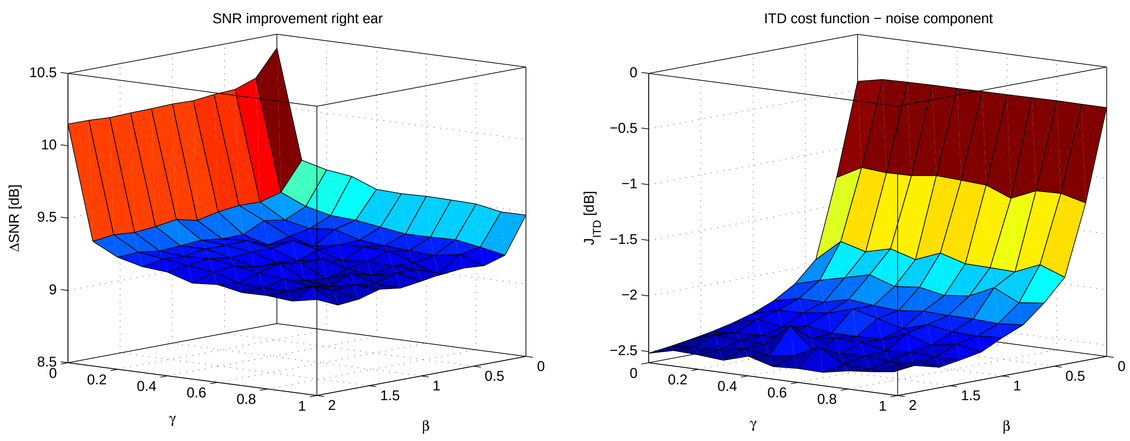
- Estimate noise cross-correlation/power during noise-dominated segments
- No closed-form expression for solution → iterative optimization techniques
- For a **single noise source**, controlling ITD and ILD corresponds to controlling Interaural Transfer Function (ITF) → interpretation for multiple noise sources ?

5 Experimental results

- Binaural recordings on KEMAR, 2 microphones at each ear ($d = 1$ cm)
- Speech source in front ($0°$), multi-talker babble noise source at $45°$
- SNR = 0 dB, $T_{60} = 125$ ms, $f_s = 16$ kHz, FFT-size $N = 256$, $\mu_0 = \mu_1 = 1$
- Performance measures: SNR improvement (left/right), mean ITD and ILD cost function (speech/noise component)
- **Partial estimation of noise component**: ITD and ILD cost function of speech and noise components decrease, SNR improvement is significantly degraded



- **Extension with ITD/ILD cost function**: ITD and ILD cost function of noise component decrease, SNR improvement is practically not compromised



- **Future work**: better perceptual performance measures, tuning of frequency-dependent weight factors β and γ, multiple noise sources