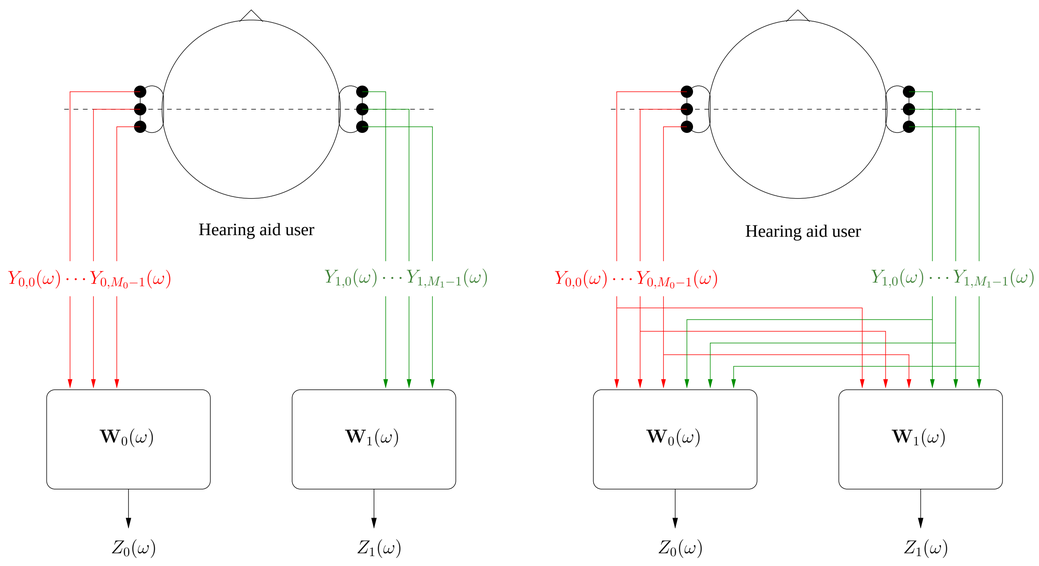
Theoretical analysis of binaural cue preservation using Multi-Channel Wiener Filtering and Interaural Transfer Functions

S. Doclo, T.J. Klasen, T. Van den Bogaert, J. Wouters, M. Moonen
Katholieke Universiteit Leuven, Belgium - Dept. Electrical Engineering (SCD), ExpORL

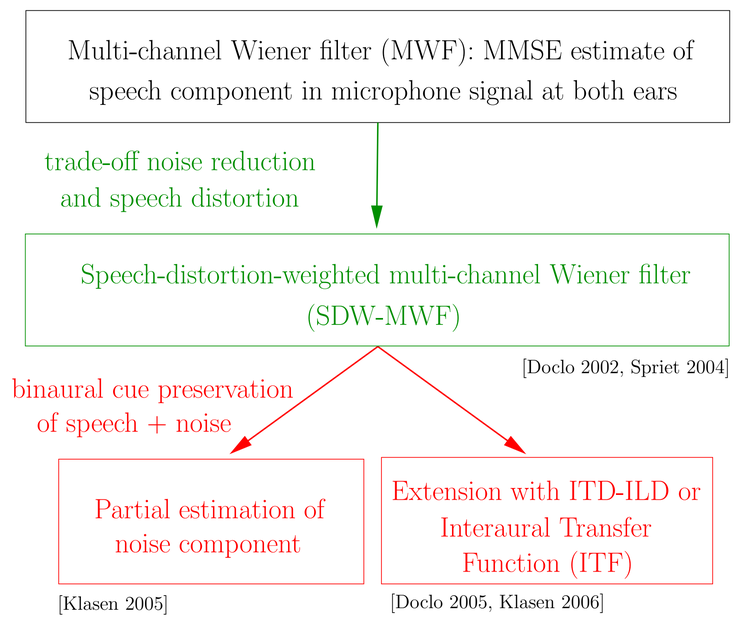
1 Binaural hearing aids

- Hearing impairment → reduction of speech intelligibility in background noise
  - signal processing to **selectively enhance/extract useful speech signal**
  - multiple microphones available: spectral + spatial processing
  - 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**



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 Binaural multi-channel Wiener filter



- Configuration: microphone array at left and right hearing aid

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

$$\mathbf{Y}(\omega) = \begin{bmatrix} Y_{0,0}(\omega) \ \ldots \ Y_{0,M_0-1}(\omega) & Y_{1,0}(\omega) \ \ldots \ Y_{1,M_1-1}(\omega) \end{bmatrix}^T = \mathbf{X}(\omega) + \mathbf{V}(\omega)$$

- 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 microphone signal at both ears; additional trade-off between noise reduction and speech distortion

$$J_{SDW}(\mathbf{W}) = \mathcal{E}\left\{ \left\| \begin{bmatrix} X_{0,r_0} - \mathbf{W}_0^H\mathbf{X} \\ X_{1,r_1} - \mathbf{W}_1^H\mathbf{X} \end{bmatrix} \right\|^2 + \mu \left\| \begin{bmatrix} \mathbf{W}_0^H\mathbf{V} \\ \mathbf{W}_1^H\mathbf{V} \end{bmatrix} \right\|^2 \right\} \Rightarrow \mathbf{W}_{SDW} = \mathbf{R}^{-1}\mathbf{r}$$

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

  - estimate $\mathbf{R}_y$ during speech-dominated segments and $\mathbf{R}_v$ during noise-dominated segments → robust VAD required
  - no assumptions about positions of microphones and sources

3 Theoretical analysis

- **Performance measures**:
  - *SNR improvement* (left/right): difference between input and output SNR
  - *ITD error* (speech/noise): phase of cross-correlation
  - *ILD error* (speech/noise): power ratio
- **Single speech source, no assumptions about noise field**:
  - $\mathbf{X} = \mathbf{A}S$ with $\mathbf{A}$ acoustic transfer function vector (head, microphones, room)

$$\mathbf{W}_{SDW,0} = \frac{\mathbf{R}_v^{-1}\mathbf{A}}{\mathbf{A}^H\mathbf{R}_v^{-1}\mathbf{A} + \frac{\mu}{P_s}} A_{0,r_0}^*, \quad \mathbf{W}_{SDW,1} = \frac{\mathbf{R}_v^{-1}\mathbf{A}}{\mathbf{A}^H\mathbf{R}_v^{-1}\mathbf{A} + \frac{\mu}{P_s}} A_{1,r_1}^*$$

- ITD/ILD of speech component is perfectly preserved
- ITD/ILD of output noise component = ITD/ILD of speech component !

4 Extension with Interaural Transfer Function

- Control binaural cues of noise (and speech) component
- **Interaural Transfer Function (ITF)**: incorporates both ITD and ILD
  - assumption: single localized noise source (constant ITF)

$$ITF_{des}^v = \frac{V_{0,r_0}}{V_{1,r_1}} = \frac{\mathcal{E}\{V_{0,r_0}V_{1,r_1}^*\}}{\mathcal{E}\{V_{1,r_1}V_{1,r_1}^*\}}, \quad ITF_{out}^v(\mathbf{W}) = \frac{\mathbf{W}_0^H\mathbf{V}}{\mathbf{W}_1^H\mathbf{V}}$$

$$J_{ITF}^v(\mathbf{W}) = \mathcal{E}\left\{ \left| \frac{\mathbf{W}_0^H\mathbf{V}}{\mathbf{W}_1^H\mathbf{V}} - ITF_{des}^v \right|^2 \right\} = \frac{\mathcal{E}\{|\mathbf{W}_0^H\mathbf{V} - ITF_{des}^v\mathbf{W}_1^H\mathbf{V}|^2\}}{\mathcal{E}\{|\mathbf{W}_1^H\mathbf{V}|^2\}} = \frac{\mathbf{W}^H\mathbf{R}_{vt}\mathbf{W}}{\mathbf{W}^H\mathbf{R}_{v1}\mathbf{W}}$$

- **Total cost function**: noise reduction, speech distortion, cue preservation

$$J_{tot}(\mathbf{W}) = J_{SDW}(\mathbf{W}) + \alpha J_{ITF}^x(\mathbf{W}) + \beta J_{ITF}^v(\mathbf{W})$$

  - subtle difference with quadratic ITF cost function in [Klasen, ICASSP 2006]
  - no-closed form expression → iterative optimization techniques

5 Simulation results

- Investigate effect of α and β on noise reduction and cue preservation
- **Data model**:
  - one speech source + one noise source, *non-reverberant environment*
  - head shadow effect → HRTFs (equal for microphones on same hearing aid)
  - sensor noise: $\mathbf{R}_v(\omega) = P_v(\omega)\left[\mathbf{g}(\omega,\theta_v)\mathbf{g}^H(\omega,\theta_v) + \delta\mathbf{I}_M\right]$
- **Simulation parameters**:
  - speech source at $-5°$ and noise source at $40°$
  - 2-microphone array ($d_0 = 2\,$cm, $d_1 = 1.5\,$cm)
  - $f = 2\,$kHz, $f_s = 16\,$kHz, SNR = $0\,$dB, $\delta = 0.01$ (sensor noise $-20\,$dB), $\mu = 1$
- **Conclusions**:
  - Increasing β substantially decreases ITD/ILD error of noise component, but also decreases SNR improvement
  - α can be used for reducing ITD/ILD error of speech component caused by increasing β

