Binaural multi-channel Wiener filtering for hearing aids: Preserving interaural time and level differences

[1,2]**T.J. Klasen,** [1]**Simon Doclo,** [1,2]**Tim Van den Bogaert,** [1]**Marc Moonen,** [2]**Jan Wouters**

[1]KU Leuven ESAT, Kasteelpark Arenberg 10, Leuven [2]KU Leuven Lab. ORL, Kapucijnenvoer 33, Leuven

tklasen@esat.kuleuven.be

Introduction

- Hearing impaired persons localize sounds better without their bilateral hearing aids than with them.
- Current hearing aids are not designed to preserve localization cues
- Advantages of preserving localization cues
  - Visual cues ⇒ Improvement in intelligibility
  - Spatial separation ⇒ Improvement in intelligibility

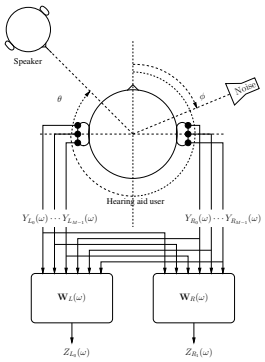
Interaural localization cues

- Interaural time difference (ITD)
  - ITD is difference in arrival of signal between ears
  - ITD cues reside in low frequencies < 1500Hz
- interaural level difference (ILD)
  - ILD is intensity difference between ears
  - ILD cues reside in high frequencies > 3000Hz

State of the art

- Binaural Wiener filter ⇒ Preserves speech ITD cues
- Controlled binaural Wiener filter ⇒ Preserves noise ITD cues at cost of noise reduction
- Extended cost function includes ITD and ILD terms ⇒ Iterative optimization techniques

System model



- Signals received at the mth microphone pair

$$Y_{L_m}(\omega) = \underbrace{X_{L_0}(\omega)}_{Speech} + \underbrace{V_{L_m}(\omega)}_{Noise} \quad Y_{R_m}(\omega) = \underbrace{X_{R_m}(\omega)}_{Speech} + \underbrace{V_{R_m}(\omega)}_{Noise}$$

- $2M$-dimensional signal vector

$$\mathbf{Y}(\omega) = \left[Y_{L_0}(\omega)\dots Y_{L_{M-1}}(\omega) Y_{R_0}(\omega)\dots Y_{R_{M-1}}(\omega)\right]^T$$
$$\mathbf{Y}(\omega) = \mathbf{X}(\omega) + \mathbf{V}(\omega)$$

- Left and Right$2M$-dimensional filters

$$\mathbf{W}(\omega) = \begin{bmatrix} \mathbf{W}_L(\omega) \\ \mathbf{W}_R(\omega) \end{bmatrix} = \begin{bmatrix} \left[W_{L_0}(\omega)\dots W_{L_{2M-1}}(\omega)\right]^T \\ \left[W_{R_0}(\omega)\dots W_{R_{2M-1}}(\omega)\right]^T \end{bmatrix}$$

Interaural transfer function (ITF)

- Input and Output ITFs (speech and noise)

$$ITF_{X_{des}} = \frac{X_{L_0}}{X_{R_0}} \qquad ITF_{V_{out}}(\mathbf{W}) = \frac{\mathbf{W}_L^H \mathbf{V}}{\mathbf{W}_R^H \mathbf{V}}$$

- Desired ITFs of the speech and noise components
  - In function of the desired angles θ_X and θ_V, and frequency, ω

$$ITF_{X_{des}} = \frac{HRTF_{X_L}(\omega, \theta_X)}{HRTF_{X_L}(\omega, \theta_X)}$$

  - As original ITFs

$$ITF_{X_{des}} = \frac{\mathcal{E}\left\{X_{L_0} X_{R_0}^*\right\}}{\mathcal{E}\left\{X_{R_0} X_{R_0}^*\right\}} \qquad ITF_{V_{des}} = \frac{\mathcal{E}\left\{V_{L_0} V_{R_0}^*\right\}}{\mathcal{E}\left\{V_{R_0} V_{R_0}^*\right\}}$$

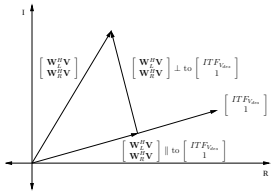
- Preserve binaural cues ⇒ original ITFs as desired ITFs

Binaural Wiener filtering

- Original cost function

$$J(\mathbf{W}) = \mathcal{E}\left\{ \underbrace{\left\| \begin{bmatrix} X_{L_0} - \mathbf{W}_L^H \mathbf{X} \\ X_{R_0} - \mathbf{W}_R^H \mathbf{X} \end{bmatrix} \right\|^2}_{Speech\ Distortion} + \mu \underbrace{\left\| \begin{bmatrix} \mathbf{W}_L^H \mathbf{V} \\ \mathbf{W}_R^H \mathbf{V} \end{bmatrix} \right\|^2}_{Residual\ Noise} \right\}$$

- Goal: Output speech and noise parallel to desired ITFs



- Add ITF terms to cost function minimize perpendicular part

$$J(\mathbf{W}) = \mathcal{E}\Bigg\{ \underbrace{\left\| \begin{bmatrix} X_{L_0} - \mathbf{W}_L^H \mathbf{X} \\ X_{R_0} - \mathbf{W}_R^H \mathbf{X} \end{bmatrix} \right\|^2 + \mu \left\| \begin{bmatrix} \mathbf{W}_L^H \mathbf{V} \\ \mathbf{W}_R^H \mathbf{V} \end{bmatrix} \right\|^2}_{Original\ SDW\ Cost\ Function} + \underbrace{\alpha \left\| \begin{bmatrix} \mathbf{W}_L^H \mathbf{X} \\ \mathbf{W}_R^H \mathbf{X} \end{bmatrix}_{\perp} \right\|^2 + \beta \left\| \begin{bmatrix} \mathbf{W}_L^H \mathbf{V} \\ \mathbf{W}_R^H \mathbf{V} \end{bmatrix}_{\perp} \right\|^2}_{Additional\ ITF\ Terms} \Bigg\}$$

- Rewrite using definition of the cross product

$$\mathbf{J}(\mathbf{W}) = \mathcal{E}\Bigg\{ \left\| \begin{bmatrix} X_{L_0} - \mathbf{W}_L^H \mathbf{X} \\ X_{R_0} - \mathbf{W}_R^H \mathbf{X} \end{bmatrix} \right\|^2 + \mu \left\| \begin{bmatrix} \mathbf{W}_L^H \mathbf{V} \\ \mathbf{W}_R^H \mathbf{V} \end{bmatrix} \right\|^2 + \alpha \frac{\left| \mathbf{W}_L^H \mathbf{X} - ITF_{X_{des}} \mathbf{W}_R^H \mathbf{X} \right|^2}{\left\| \begin{bmatrix} ITF_{X_{des}} \\ 1 \end{bmatrix} \right\|^2} + \beta \frac{\left| \mathbf{W}_L^H \mathbf{V} - ITF_{V_{des}} \mathbf{W}_R^H \mathbf{V} \right|^2}{\left\| \begin{bmatrix} ITF_{V_{des}} \\ 1 \end{bmatrix} \right\|^2} \Bigg\}.$$

- Take derivative of $\mathbf{J}(\mathbf{W})$, set to zero, and solve for $\mathbf{W}$

$$\mathbf{W} = \left(\mathcal{E}\left\{ \mathbf{R}_{\mathbf{R}_X} + \mu \mathbf{R}_{\mathbf{R}_V} + \alpha \mathbf{R}_{\mathbf{R}_{XC}} + \beta \mathbf{R}_{\mathbf{R}_{VC}} \right\} \right)^{-1} \mathcal{E}\left\{ \mathbf{r}_X \right\}$$

where,

$$\mathbf{r}_X = \begin{bmatrix} X_{L_0}^* \mathbf{X} \\ X_{R_0}^* \mathbf{X} \end{bmatrix} \quad \mathbf{R}_X = \mathbf{X}\mathbf{X}^H \quad \mathbf{R}_V = \mathbf{V}\mathbf{V}^H$$

$$\mathbf{R}_{\mathbf{R}_X} = \begin{bmatrix} \mathbf{R}_X & \mathbf{0}_{2M} \\ \mathbf{0}_{2M} & \mathbf{R}_X \end{bmatrix} \quad \mathbf{R}_{\mathbf{R}_V} = \begin{bmatrix} \mathbf{R}_V & \mathbf{0}_{2M} \\ \mathbf{0}_{2M} & \mathbf{R}_V \end{bmatrix}$$

$$\mathbf{R}_{\mathbf{R}_{XC}} = \begin{bmatrix} \mathbf{R}_X & -ITF_{X_{des}}^* \mathbf{R}_X \\ -ITF_{X_{des}} \mathbf{R}_X & |ITF_{X_{des}}|^2 \mathbf{R}_X \end{bmatrix}$$

$$\mathbf{R}_{\mathbf{R}_{VC}} = \begin{bmatrix} \mathbf{R}_V & -ITF_{V_{des}}^* \mathbf{R}_V \\ -ITF_{V_{des}} \mathbf{R}_V & |ITF_{V_{des}}|^2 \mathbf{R}_V \end{bmatrix}$$

Simulations

Setup

- $T_{60} = 0.76\ sec$, $f_s = 16\ kHz$, and FFT size = 256
- HINT speech at 345 degrees and HINT noise at 60 degrees
- Input SNR Left 2.8dB Right -6.8dB
- GN ReSound Canta behind the ear hearing aids on CORTEX MK2 artificial head
- Varied α and β from 0 to 100 with $\mu = 1$

Performance measures

- ITD Error (N bins < 1500Hz)

$$\frac{1}{N} \sum_{i=1}^{N} \left(1 - \cos\left(\angle\mathcal{E}\left\{ X_{L_0} X_{R_0}^* \right\} - \angle\mathcal{E}\left\{ \mathbf{W}_L^H \mathbf{X} \left(\mathbf{W}_R^H \mathbf{X} \right)^* \right\} \right) \right)$$

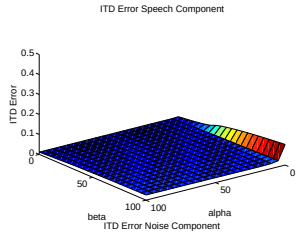
- ILD Error (All N bins)

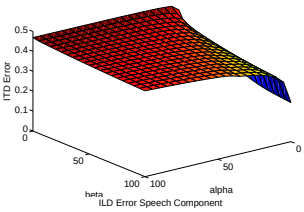
$$\frac{1}{N} \sum_{i=1}^{N} \left| 10 \log_{10} \frac{P_{L_{in}}(\omega_i)}{P_{R_{in}}(\omega_i)} - 10 \log_{10} \frac{P_{L_{out}}(\omega_i)}{P_{R_{out}}(\omega_i)} \right|$$

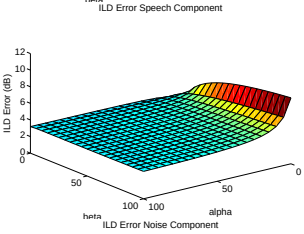
- Improvement in speech intelligibility weighted signal-to-noise-ratio ($\mathrm{SNR_{INT}}$)

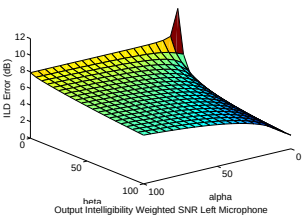
$$\mathrm{SNR_{INT}} = \sum_{j=1}^{J} w_j \mathrm{SNR}_j$$

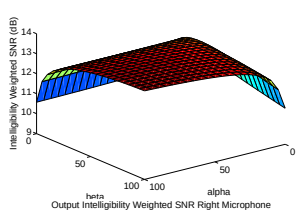
Results

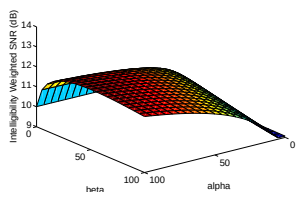


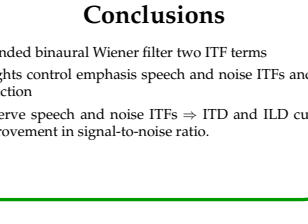












Conclusions

- Extended binaural Wiener filter two ITF terms
- Weights control emphasis speech and noise ITFs and noise reduction
- Preserve speech and noise ITFs ⇒ ITD and ILD cues and improvement in signal-to-noise ratio.