Binaural noise reduction for hearing aids

Simon Doclo\textsuperscript{1}, Tim Van den Bogaert\textsuperscript{2}, Marc Moonen\textsuperscript{1}, Jan Wouters\textsuperscript{2}

\textsuperscript{1}Dept. of Electrical Engineering (ESAT-SCD), KU Leuven, Belgium
\textsuperscript{2}Laboratory for Exp. ORL, KU Leuven, Belgium

ASIP-NET, Oct 26 2006
Overview

- Binaural hearing aids: noise reduction and preservation of binaural cues
- Overview of binaural noise reduction algorithms
- **Binaural multi-channel Wiener filter:**
  - Estimate of speech component at both hearing aids
  - Speech cues are preserved – noise cues may be distorted
- **Preservation of binaural noise cues:**
  - Partial estimation of noise component
  - Extension with ITD-ILD-ITF cost function
  - Postfiltering procedure
- **Experimental results:**
  - Physical evaluation (SNR, ITD, ILD)
  - Perceptual evaluation (SRT, localisation)
- **Audio demonstration**
Problem statement

• Hearing impairment → reduction of speech intelligibility in background noise
  o Signal processing to selectively enhance useful speech signal
  o Many hearing impaired are fitted with hearing aid at both ears
  o Multiple microphones available: spectral + spatial processing

• Binaural auditory cues:
  o Interaural Time Difference (ITD) – Interaural Level Difference (ILD)
  o Binaural cues, in addition to spectral and temporal cues, play an important role in binaural noise reduction and sound localization
Problem statement

- Bilateral system:
  - Independent processing of left and right hearing aid
**Problem statement**

- **Bilateral system:**
  - Independent processing of left and right hearing aid
  - Localisation cues are distorted

**RMS error per loudspeaker when accumulating all responses of the different test conditions**

- **NH** = normal hearing
- **NO** = hearing impaired without hearing aids
- **O** = omnidirectional configuration
- **A** = adaptive directional configuration

[Van den Bogaert et al., 2006]
Problem statement

- Bilateral system:
  - Independent processing of left and right hearing aid
  - Localisation cues are distorted

FIG. 5. Mean speech reception thresholds obtained in experiment I for three different noise types: FF (free field), dL (headshadow only), and dT (ITD only). The closed data points represent results of Plomp and Mimpen (1981) obtained in a free field.

[Bronkhorst and Plomp, 1988] [Beutelmann and Brand, 2006]
Problem statement

- **Bilateral system:**
  - Independent processing of left and right hearing aid
  - Localisation cues are distorted

- **Binaural system:**
  - Cooperation between left and right hearing aid (e.g. wireless link)
  - Assumption: all microphone signals are available at the same time

**Objectives/requirements for binaural algorithm:**

1. SNR improvement: noise reduction, limit speech distortion
2. Preservation of binaural cues (speech/noise) to exploit binaural hearing advantage
3. No assumption about position of speech source and microphones
Binaural noise reduction techniques

- Problem statement
- Binaural noise reduction
- Multi-channel Wiener filter
- Preservation of binaural cues
- Experimental results
- Demo (1)
- Postfiltering
- Demo (2)
- Conclusions
Binaural noise reduction techniques

- Problem statement
- Binaural noise reduction
- Multi-channel Wiener filter
- Preservation of binaural cues
- Experimental results
- Demo (1)
- Postfiltering
- Demo (2)
- Conclusions
Binaural noise reduction techniques

- **Fixed beamforming**: spatial selectivity + binaural speech cues
  - Maximize directivity index while restricting speech ITD error
    - [Desloge, 1997]
  - Superdirective beamformer using HRTFS
    - [Lotter, 2004]
  - **Pros**
    - low computational complexity
  - **Cons**
    - limited performance, known geometry, broadside array, only speech cues
Binaural noise reduction techniques

- CASA-based techniques [Kollmeier, Peissig, Wittkop, Dong, Haykin]
  - Computation and application of (real-valued) binaural mask based on binaural and temporal/spectral cues
    - Perfect preservation of binaural cues of speech/noise component
    - Mostly for 2 microphones, "spectral-subtraction"-like problems

[Diagram of binaural noise reduction process]
Binaural noise reduction techniques

- **Adaptive beamforming**: based on GSC-structure
  - Divide frequency spectrum: low-pass portion unaltered to preserve ITD cues, high-pass portion processed using GSC [Welker, 1997]

  🎧 preserves binaural cues to some extent
  🙁 substantial reduction in noise reduction performance, known geometry

---

[Welker, 1997]
Binaural noise reduction techniques

- Binaural multi-channel Wiener filter [Doclo, Klasen, Wouters, Moonen]
  - MMSE estimate of speech component in microphone signal at both ears
  - Speech cues are preserved, no assumptions about position of speech source and microphones

Extension of MWF:

- Preservation of binaural speech and noise cues without substantially compromising noise reduction performance
**Configuration and signals**

- **Configuration**: microphone array with $M$ microphones at left and right hearing aid, communication between hearing aids

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

- Vector notation: $Y(\omega) = X(\omega) + V(\omega)$

- **One desired signal source**: $X(\omega) = A(\omega)S(\omega)$

$A(\omega) =$ transfer function (TF) vector between source and mic array
Configuration and signals

- Use all available microphone signals to compute output signal at both ears → computation of filters $W_0$ and $W_1$

\[
Z_0(\omega) = W_0^H(\omega)Y(\omega), \quad Z_1(\omega) = W_1^H(\omega)Y(\omega)
\]

\[
W(\omega) = \begin{bmatrix} W_0(\omega) \\ W_1(\omega) \end{bmatrix}
\]
Overview of cost functions

Multi-channel Wiener filter (MWF): MMSE estimate of speech component in microphone signal at both ears

trade-off noise reduction and speech distortion

Speech-distortion weighted multi-channel Wiener filter (SDW-MWF)

binaural cue preservation of speech + noise

Partial estimation of noise component

Extension with ITD-ILD or Interaural Transfer Function (ITF)

[Doclo 2002, Spriet 2004]

[Klasen 2005]

[Doclo 2005, Klasen 2006]
**Binaural multi-channel Wiener filter**

- **Binaural SDW-MWF:** estimate of speech component in microphone signal at both ears (usually front microphone) + trade-off between noise reduction and speech distortion

\[
J(W) = E \left\{ \left[ X_{0,r_0} - W_0^H X \right]^2 + \mu \left[ W_0^H V \right]^2 \right\}
\]

\[
W_{SDW} = R^{-1}r
\]

- trade-off parameter

\[
R = \begin{bmatrix}
R_x + \mu R_v & 0_M \\
0_M & R_x + \mu R_v
\end{bmatrix},
\]

\[
r = \begin{bmatrix}
r_{x0} \\
r_{x1}
\end{bmatrix},
\]

\[
R_x = R_y - R_v
\]

- Depends on second-order statistics of speech and noise
- Estimate \( R_v \) during speech-dominated time-frequency segments, estimate \( R_v \) during noise-dominated segments, requiring robust voice activity detection (VAD) mechanism
- No assumptions about positions of microphones and sources
Binaural multi-channel Wiener filter

- Interpretation for single speech source:
  - Spectral and spatial filtering operation
    \[ W_{SDW,0} = \frac{\Gamma_v^{-1} A}{A^H \Gamma_v^{-1} A} + \frac{A^H \Gamma_v^{-1} A}{A^H \Gamma_v^{-1} A + \mu P_v / P_s} A^*_0, \tilde{r}_0 \]

  - Spatial separation between speech and noise sources
  - with \( \Gamma \) (spatial) coherence matrix and \( P \) (spectral) power
  - Equivalent to superdirective beamformer (diffuse noise field) or delay-and-sum beamformer (spatially white noise field)
    + single-channel WF-based postfilter (spectral subtraction)

- Binaural cues (ITD-ILD):
  - Perfectly preserves binaural cues of speech component
  - Binaural cues of noise component \( \rightarrow \) speech component !!
Binaural multi-channel Wiener filter

- Partial estimation of noise component [Klasen, 2005]
  - Estimate of sum of speech component and scaled noise component
    \[ J(W) = E\left\| \begin{pmatrix} X_{0,r_0} + \lambda V_{0,r_0} - W_0^H Y \\ X_{1,r_1} + \lambda V_{1,r_1} - W_1^H Y \end{pmatrix}^2 \right\| \]
  - Considerable reduction of noise reduction performance
  - Works for multiple noise sources

- Extension of SDW-MWF with binaural cues
  - Add term related to binaural cues of noise (and speech) component to SDW cost function
    \[ J_{tot}(W) = J_{SDW}(W) + \alpha J_{cue}^x(W) + \beta J_{cue}^v(W) \]
  - Link computation of filters \( W_0 \) and \( W_1 \)
  - Possible cues: ITD, ILD, Interaural Transfer Function (ITF)
  - Weight factors \( \alpha \) and \( \beta \) can be frequency-dependent
Interaural Wiener Filter

- Preserve binaural cues between input and output
  - ITD: phase of cross-correlation
  - ILD: power ratio
  - ITF: Interaural transfer function (incorporates ITD and ILD)

\[ \text{ITF}_{\text{in}}^v = \frac{V_{0,0}}{V_{1,1}} = \frac{E\{V_{0,0} V_{1,1}^*\}}{E\{V_{1,1} V_{1,1}^*\}} = \frac{R_v(r_0, r_1)}{R_v(r_1, r_1)} \]

\[ \text{ITF}_{\text{out}}^v = \frac{Z_v}{Z_v} = \frac{W_0^H V}{W_1^H V} \]

\[ J_{\text{tot}}(W) = E \left\{ \left\| X_{0,0} - W_0^H X \right\|^2 + \mu \left\| W_0^H V \right\|^2 \right\} + \alpha E \left\{ \left\| W_0^H X - \text{ITF}_{\text{in}}^v X \right\|^2 \right\} + \beta E \left\{ \left\| W_0^H V - \text{ITF}_{\text{in}}^v V \right\|^2 \right\} \]

- Closed form expression!
- Large \( \beta \) changes direction of speech component to noise component → increase weight \( \alpha \) (cf. physical and perceptual evaluation)
Overview of batch algorithm

- Problem statement
- Binaural noise reduction
- Multi-channel Wiener filter
- Preservation of binaural cues
- Experimental results
- Demo (1)
- Postfiltering
- Demo (2)
- Conclusions

Off-line computation of statistics

\[ \mathbf{Y}(\omega) = \mathbf{X}(\omega) + \mathbf{V}(\omega) \]

\[ \mathbf{R}_v(\omega), \mathbf{R}_x(\omega) \]

Calculate binaural input cues and filter

\[ \mathbf{W}(\omega) = \begin{bmatrix} \mathbf{W}_0(\omega) \\ \mathbf{W}_1(\omega) \end{bmatrix} \]

\[ \mu, \alpha, \beta \]

Frequency-domain filtering

\[ Z_0 = Z_{x0} + Z_{v0} \]

\[ Z_1 = Z_{x1} + Z_{v1} \]

Left input signals

Right input signals
Experimental results

- Problem statement
- Binaural noise reduction
- Multi-channel Wiener filter
- Preservation of binaural cues

- Experimental results
  - Physical
  - Perceptual
- Demo (1)
- Postfiltering
- Demo (2)
- Conclusions

• Identification of HRTFs:
  - Binaural recordings on CORTEX MK2 artificial head
  - 2 omni-directional microphones on each hearing aid (d=1cm)
  - LS = -90°:15°:90°, 90°:30°:270°, 1m from head
  - Conditions: T_60=140 ms, f_s=16 kHz, L=1366 taps
Experimental results

• Problem statement
  - Binaural noise reduction
  - Multi-channel Wiener filter
  - Preservation of binaural cues
  - Experimental results
    - Physical
    - Perceptual
  - Demo (1)
  - Postfiltering
  - Demo (2)
  - Conclusions

• Speech and noise material:
  - Dutch sentences (VU list)
  - Stationary speech-weighted noise with same long-term spectrum as speech material → spatial aspects
  - $S_0N_{60}, SNR=0 \text{ dB}$
  - $f_s=16 \text{ kHz}, FFT-size N=256, \mu=1$

• Physical evaluation:
  - Speech intelligibility: $\Delta SNR$
  - Localisation: $\Delta ITD / \Delta ILD$

• Perceptual evaluation:
  - Preliminary study
  - Speech intelligibility: SRT
  - Localisation: localise S and N
Physical evaluation

- **Performance measures:**
  - Intelligibility weighted SNR improvement (left/right)
    \[
    \Delta SNR_L = \sum_i I(\omega_i) \Delta SNR_L(\omega_i)
    \]
    importance of i-th frequency
    for speech intelligibility
  - ILD error (speech/noise component) \(\rightarrow\) power ratio
    \[
    \Delta ILD_x = \sum_i |ILD_{out}^x(\omega_i) - ILD_{in}^x(\omega_i)|
    \]
  - ITD error (speech/noise component) \(\rightarrow\) phase of cross-correlation
    \[
    \Delta ITD_x = \sum_i I(\omega_i) \Delta ITD_x(\omega_i)
    \]
    \[
    \Delta ITD_x(\omega_i) = \angle E\{X_{0,\omega_i} X_{1,\omega_i}^*\} - \angle E\{Z_{x,0} Z_{x,1}^*\}
    \]
    low-pass filter 1500 Hz
Physical evaluation: SNR

SNR improvement left ear

\[ \Delta \text{SNR}_w \text{ [dB]} \]

SNR improvement right ear

\[ \Delta \text{SNR}_w \text{ [dB]} \]
Physical evaluation: ILD-ITD

ILD error speech component

ILD error noise component

ITD error speech component

ITD error noise component
Physical evaluation

- Problem statement
  - Binaural noise reduction
  - Multi-channel Wiener filter
  - Preservation of binaural cues
- Experimental results
  - Physical
    - Perceptual
  - Demo (1)
  - Postfiltering
  - Demo (2)
  - Conclusions

- Conclusions:
  - $\beta$ increases: ITD-ILD error of noise component decreases
    ... BUT... ITD-ILD error of speech component increases
  - $\alpha$ increases: ITD-ILD error of speech component decreases
    ... BUT... ITD-ILD error of noise component increases
  - Compromise between speech and noise localisation error possible
    (cf. localisation experiments)
  - SNR improvement only slightly degraded
    (cf. SRT experiments)
Perceptual evaluation

- **Speech intelligibility: SRT**
  - How does parameter $\beta$ affect speech intelligibility?
  - Two effects: increasing $\beta$ reduces SNR improvement, but preserves binaural noise cues better, enabling binaural speech intelligibility advantage

- **Localisation performance**
  - How do parameters $\alpha$ and $\beta$ affect localisation of processed speech and noise components?
  - $\alpha$: preservation of speech cues, $\beta$: preservation of noise cues
**Perceptual evaluation: SRT**

- **Measurement procedure:**
  - SRT = SNR where 50% of speech is intelligible
  - adaptive procedure (2 dB/step)
  - headphone experiments, using HRTFs
  - $S_0N_{60}$ (Dutch VU sentences – stationary noise)
  - presentation level = 65 dB SPL
  - 5 normal-hearing subjects
  - $f_s=16$ kHz, FFT-size $N=256$, $\mu=1$, $\alpha=0$
  - Reference condition = no processing
Perceptual evaluation: SRT

- Problem statement
- Binaural noise reduction
- Multi-channel Wiener filter
- Preservation of binaural cues
- Experimental results
  - Physical
  - Perceptual
- Demo (1)
- Postfiltering
- Demo (2)
- Conclusions

**Results:**

- Average SRT without processing = -9.2 dB
- SRT improvements in the range **11-13 dB**
- Binaural speech intelligibility advantage does not seem to compensate for loss in SNR improvement
Perceptual evaluation: localisation

- Sum of localisation errors $S_x$ and $N_0$

- Parameters can be tuned to achieve better overall localization performance at the cost of some noise reduction

- Good correlation between physical and perceptual evaluation
### Audio demo 1

- **Speech and noise material:**
  - HINT sentences, speech source in front (0°)
  - Multi-talker babble noise at 60°
  - SNR=0 dB, $f_s=16$ kHz, FFT-size $N=256$, $\mu=1$, $\alpha=0$

<table>
<thead>
<tr>
<th></th>
<th>Noisy</th>
<th>Speech</th>
<th>Noise</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Input</strong></td>
<td><img src="images/speaker.png" alt="Speaker" /></td>
<td><img src="images/speaker.png" alt="Speaker" /></td>
<td><img src="images/speaker.png" alt="Speaker" /></td>
</tr>
<tr>
<td><strong>Output ($\beta=0$)</strong></td>
<td><img src="images/speaker.png" alt="Speaker" /></td>
<td><img src="images/speaker.png" alt="Speaker" /></td>
<td><img src="images/speaker.png" alt="Speaker" /></td>
</tr>
<tr>
<td><strong>Output ($\beta=0.05$)</strong></td>
<td><img src="images/speaker.png" alt="Speaker" /></td>
<td><img src="images/speaker.png" alt="Speaker" /></td>
<td><img src="images/speaker.png" alt="Speaker" /></td>
</tr>
<tr>
<td><strong>Output ($\beta=10$)</strong></td>
<td><img src="images/speaker.png" alt="Speaker" /></td>
<td><img src="images/speaker.png" alt="Speaker" /></td>
<td><img src="images/speaker.png" alt="Speaker" /></td>
</tr>
</tbody>
</table>


**Postfiltering**

- **CASA-based techniques:** [Kollmeier, Peissig, Wittkop, Dong, Haykin]
  - Mostly restricted to 2 microphones, “spectral-subtraction”-like
  - Computation and application of real-valued gain $G(\omega)$

\[
Z_0(\omega) = G(\omega) Y_{0,r_0}(\omega) = G(\omega) |Y_{0,r_0}(\omega)| e^{j\phi}
\]

- "Optimal gain": $G(\omega) = \frac{S_{0,r_0}(\omega)}{|Y_{0,r_0}(\omega)|}$ (SNR-based)
- Same gain $\rightarrow$ preservation of binaural cues for all sources
- Gain function: based on coherence, binaural cues (ITD/ILD) and temporal/spectral cues (pitch, onset, modulation frequencies)
- Disadvantages: noisy phase $\rightarrow$ performance limit, musical noise
Postfiltering

- Combination of advantages of different techniques: [Lotter, 2004]
  - Multi-microphone noise reduction (SDW-MWF, TF-LCMV, fixed BF)
  - Real-valued gain → preservation of binaural cues for all sources

\[
G(\omega) = \frac{2|Z(\omega)|^\kappa}{|Y_{0,0}(\omega)|^\kappa + |Y_{1,0}(\omega)|^\kappa} \quad \Rightarrow \quad D(\omega, \theta) = \frac{2|W^H(\omega)d(\omega, \theta)|^\kappa}{|d_{0,0}(\omega, \theta)|^\kappa + |d_{1,0}(\omega, \theta)|^\kappa}
\]

- Disadvantage: spectral amplitude filtering → noisy phase, (musical noise)
Experimental results

- Binaural recordings on KEMAR, $T_{60} = 290$ ms
- 3 microphones on each hearing aid ($d=1$cm)
- Speech source in front, 2-4-6 babble noise sources ($45^\circ$, $65^\circ$, $135^\circ$)
- SNR=0 dB, $f_s=16$ kHz, FFT-size $N=256$
- SDW-MWF: $\mu=1-5$, postfilter: $\kappa=1$
Experimental results

SDW-MWF, $T_{60} = 290$ ms

- Problem statement
- Binaural noise reduction
- Multi-channel Wiener filter
- Preservation of binaural cues
- Experimental results
- Demo (1)
- Postfiltering
- Demo (2)
- Conclusions

![Graph showing noise reduction vs. number of noise sources for different channel numbers and mu values.](image)
Audio demo 2 ($T_{60} = 290$ ms)

<table>
<thead>
<tr>
<th>Noise sources</th>
<th>Input</th>
<th>“Optimal” postfilter</th>
<th>Nr Mics</th>
<th>SDW-MWF ($\mu=5$)</th>
<th>+postfilter ($k=1$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td>2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Problem statement
- Binaural noise reduction
- Multi-channel Wiener filter
- Preservation of binaural cues
- Experimental results
- Demo (1)
- Postfiltering
- Demo (2)
- Conclusions
Conclusions

- **Speech enhancement for binaural hearing aids:**
  - Improve **speech intelligibility**
  - **Localization:** preserve binaural speech and noise cues
  - No assumptions about position speech source and microphones

- **Suitable algorithm: multi-channel Wiener filter, (TF-LCMV)**
  - speech cues are preserved  noise cues may be distorted

- **Preservation of binaural noise cues:**
  - Partial estimation of noise component: **reduction SNR improvement**
  - **Interaural Wiener filter:** extension with Interaural Transfer Function of noise (and speech) component, **good results for single noise source, less effect for multiple noise sources**
  - Postfiltering: preservation of cues for all sources, **performance limit**

- **Perceptual evaluation of Interaural Wiener filter:**
  - **S₀N₆₀:** SRT improvements in the range **11-13 dB**
  - Binaural speech intelligibility advantage does not seem to compensate for (small) loss in SNR improvement
  - Parameters can be tuned to achieve better overall localization performance at the cost of some noise reduction
Future work

- **Algorithmic extensions:**
  - Better perceptual cost functions/performance measures
  - Multiple noise sources
  - Adaptive version with low computational complexity

- **Perceptual evaluation:**
  - Various speech-noise scenarios
  - Different fixed/adaptive beamforming algorithms

... perfect procedure providing noise reduction and cue preservation
References


