Physical and perceptual evaluation of the Interaural Wiener Filter algorithm

Simon Doclo\textsuperscript{1}, Thomas J. Klasen\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

IHCON, Aug 19 2006

Overview

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

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

- **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

\[ \text{ITD} = \tau \]
\[ \text{ILD} = \frac{P_L}{P_R} \]
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 [Van den Bogaert, 2006]

• 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
- Audio demo
- Conclusions
Binaural noise reduction techniques

- Problem statement
- Binaural noise reduction
- Multi-channel Wiener filter
- Preservation of binaural cues
- Experimental results
- Audio demo
- 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]

  - low computational complexity
  - limited performance, known geometry, broadside array, only speech cues

---

[Desloge, 1997]
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
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
**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
  - noise cues may be distorted

---

**Extension of MWF**: preservation of binaural speech and noise cues without substantially compromising noise reduction performance
Design of hearing aid SP algorithm requires some mathematics but perceptual evaluation in a couple of minutes...
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**: \(Y(\omega) = X(\omega) + V(\omega)\)
- **Use all microphone signals to compute output signal at both ears**

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

Partial estimation of noise component

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

binaural cue preservation of speech + noise

[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

\[
J(W) = E \left\{ \left\| X_{0,i_0} - W_0^H X \right\|^2 + \mu \left\| W_0^H V \right\|^2 \right\}
\]

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

- **Problem statement**
- **Binaural noise reduction**
- **Multi-channel Wiener filter**
- **Preservation of binaural cues**
- **Experimental results**
- **Audio demo**
- **Conclusions**

- Depends on second-order statistics of speech and noise
- Estimate \( R_y \) 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

- **Binaural cues (ITD-ILD):**
  - Perfectly preserves binaural cues of speech component
  - Binaural cues of noise component $\rightarrow$ speech component !!
    (cf. physical and perceptual evaluation)

- **Extension of SDW-MWF with binaural cues**
  - Add term related to binaural cues of noise (and speech) component to SDW cost function

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

- Possible cues: ITD, ILD, Interaural Transfer Function (ITF)
- Weight factors $\alpha$ and $\beta$ can be frequency-dependent
Interaural Wiener Filter

- Problem statement
- Binaural noise reduction
- Multi-channel Wiener filter
- Preservation of binaural cues
- Experimental results
- Audio demo
- Conclusions

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

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

\[ ITF_{out}^v = \frac{Z_{v0}}{Z_{v1}} = \frac{W_H^0 V}{W_H^1 V} \]

\[ J_{tot}(W) = E\left\{\left[\begin{array}{c} X_{0,r_0} - W_H^0 X \\ X_{1,r_1} - W_H^1 X \end{array}\right]\right\}^2 + \mu \left[\begin{array}{c} W_H^0 V \\ W_H^1 V \end{array}\right]^2 \]

\[ + \alpha E\left\{\left[ W_H^0 X - ITF_{in}^v W_H^1 X \right]^2 \right\} + \beta E\left\{\left[ W_H^0 V - ITF_{in}^v W_H^1 V \right]^2 \right\} \]

ITF preservation speech

ITF preservation noise

- Closed form expression!
- Large \( \beta \) changes direction of speech component to noise component
  \( \rightarrow \) 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
- Audio demo
- Conclusions

**Problem statement**

- Binaural noise reduction
- Multi-channel Wiener filter
- Preservation of binaural cues
- Experimental results
- Audio demo
- Conclusions

**Overview of batch algorithm**

- VAD
- Off-line computation of statistics
  - \( R_v(\omega), \ R_x(\omega) \)
- Calculate binaural input cues and filter
  - \( \mu, \alpha, \beta \)
- Frequency-domain filtering
- IFFT

**Equations**

- \( Y(\omega) = X(\omega) + V(\omega) \)
- \( W(\omega) = \begin{bmatrix} W_0(\omega) \\ W_1(\omega) \end{bmatrix} \)
- \( Z_0 = Z_{x0} + Z_{v0} \)
- \( Z_1 = Z_{x1} + Z_{v1} \)
Experimental results

- Problem statement
- Binaural noise reduction
- Multi-channel Wiener filter
- Preservation of binaural cues
- Experimental results
  - Physical
  - Perceptual
- Audio demo
- Conclusions

• **Identification of HRTFs:**
  o Binaural recordings on CORTEX MK2 artificial head
  o 2 omni-directional microphones on each hearing aid (d=1cm)
  o LS = -90°:15°:90°, 90°:30°:270°, 1m from head
  o 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
- Audio demo
- Conclusions

• Speech and noise material:
  - Dutch sentences (VU list)
  - Stationary speech-weighted noise with same long-term spectrum as speech material \( \rightarrow \) spatial aspects
  - \( S_0N_{60}, SNR=0 \) dB
  - \( f_s=16 \) 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)
    \[
    \Delta SNR_L = \sum_i I(\omega_i) \Delta SNR_L(\omega_i)
    \]
    important 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
    \[
    \Delta ITD_x = \sum_i I(\omega_i) \Delta ITD_x(\omega_i)
    \]
    \[
    \Delta ITD_x(\omega_i) = \angle E\{X_{0,\omega_i} X_{1,\omega_i}^*\} - \angle E\{Z_{0} Z_{1}^*\}
    \]
    low-pass filter 1500 Hz
Physical evaluation: SNR

SNR improvement left ear

SNR improvement right ear
Physical evaluation: ILD-ITD

ILD error speech component

ILD error noise component

ITD error speech component

ITD error noise component
Physical evaluation

• Conclusions:
  o $\beta$ increases: ITD-ILD error of noise component decreases
    ... BUT... ITD-ILD error of speech component increases
  o $\alpha$ increases: ITD-ILD error of speech component decreases
    ... BUT... ITD-ILD error of noise component increases
  o Compromise between speech and noise localisation error possible
    (cf. localisation experiments)
  o 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

- Problem statement
- Binaural noise reduction
- Multi-channel Wiener filter
- Preservation of binaural cues
- Experimental results
  - Physical
  - Perceptual
- Audio demo
- Conclusions

- 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

- 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

![Graph showing SRT improvement with respect to beta]

- Problem statement
- Binaural noise reduction
- Multi-channel Wiener filter
- Preservation of binaural cues
- Experimental results
  - Physical
  - Perceptual
- Audio demo
- Conclusions
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 demonstration

- **Speech and noise material:**
  - HINT sentences, speech source in front (0°)
  - Multi-talker babble noise at 60°
  - SNR=0 dB, \( f_s = 16 \text{ kHz} \), 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>Input</td>
<td><img src="sound.png" alt="Audio" /></td>
<td><img src="sound.png" alt="Audio" /></td>
<td><img src="sound.png" alt="Audio" /></td>
</tr>
<tr>
<td>Output (( \beta = 0 ))</td>
<td><img src="sound.png" alt="Audio" /></td>
<td><img src="sound.png" alt="Audio" /></td>
<td><img src="sound.png" alt="Audio" /></td>
</tr>
<tr>
<td>Output (( \beta = 0.05 ))</td>
<td><img src="sound.png" alt="Audio" /></td>
<td><img src="sound.png" alt="Audio" /></td>
<td><img src="sound.png" alt="Audio" /></td>
</tr>
<tr>
<td>Output (( \beta = 10 ))</td>
<td><img src="sound.png" alt="Audio" /></td>
<td><img src="sound.png" alt="Audio" /></td>
<td><img src="sound.png" alt="Audio" /></td>
</tr>
</tbody>
</table>

Conclusions

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

- Suitable algorithm: multi-channel Wiener filter
  - speech cues are preserved  noise cues may be distorted

- Preservation of binaural noise cues:
  - **Interaural Wiener filter**: extension with Interaural Transfer Function of noise (and speech) component

- Perceptual evaluation:
  - **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
References


References