Design and low-cost implementation of a robust multichannel noise reduction scheme for cochlear implants

Simon Doclo\textsuperscript{1}, Ann Spriet\textsuperscript{1,2}, Jean-Baptiste Maj\textsuperscript{1,2}, Marc Moonen\textsuperscript{1}, Jan Wouters\textsuperscript{2}, Bas Van Dijk\textsuperscript{3}, Jan Janssen\textsuperscript{3}

\textsuperscript{1}Dept. of Electrical Engineering (ESAT-SCD), KU Leuven, Belgium
\textsuperscript{2}Laboratory for Exp. ORL, KU Leuven, Belgium
\textsuperscript{3}Cochlear Technology Centre Europe, Belgium

DARTS, 22 October 2003
Overview

• Problem statement: hearing in background noise
• Adaptive beamforming: GSC
  • not robust against model errors
• Design of robust noise reduction algorithm
  • robust fixed spatial preprocessor
  • robust adaptive stage
• Experimental results
• Low-cost implementation of adaptive stage
  • stochastic gradient algorithms
  • computational complexity + memory requirements
• Conclusions
Problem statement

- Hearing problems effect more than 12% of population

- Digital hearing instruments allow for advanced signal processing, resulting in improved speech understanding

- Major problem: (directional) hearing in background noise
  - reduction of noise wrt useful speech signal
  - multiple microphones + DSP in BTE
  - current systems: simple fixed and adaptive beamforming
  - robustness important due to small inter-microphone distance

Design of robust multi-microphone noise reduction scheme
Cochlear implants

- **Working principle:** sound is converted to electrical stimuli in speech processor, allowing deaf people to hear again.

![Cochlear implants diagram]
Algorithmic requirements

• ‘Blind’ techniques: unknown noise sources and acoustic environment

• Adaptive: time-variant signals and acoustic environment

• Robustness:
  o microphone characteristics (gain, phase, position)
  o other deviations from assumed signal model (e.g. VAD)

• Implementation issues:
  o number of microphones
  o low computational complexity
  o memory
State-of-the art noise reduction

- **Single-microphone techniques:**
  - spectral subtraction, Kalman filter, subspace-based
  - only temporal and spectral information → **limited performance**

- **Multi-microphone techniques:**
  - exploit spatial information
  - *Fixed beamforming*: fixed directivity pattern
  - *Adaptive beamforming* (e.g. GSC): adapt to different acoustic environments → improved performance

  - Sensitive to a-priori assumptions
  - *Multi-channel Wiener filtering* (MWF): MMSE estimate of speech component in microphones → improved robustness

---

Robust scheme, encompassing both GSC and MWF
Adaptive beamforming: GSC

- **Fixed spatial preprocessor:**
  - Fixed beamformer creates speech reference \( y_0[k] \)
  - Blocking matrix creates noise references \( y_i[k] = x_i[k] + v_i[k] \)
- **Adaptive noise canceller:**
  - Standard GSC minimises output noise power

\[
\begin{align*}
\min_{\mathbf{w}[k]} & \left\{ \mathbf{v}_0[k] - \mathbf{w}^T[k] \mathbf{v}[k] \right\}^2 \\
\mathbf{w}[k] & = \begin{bmatrix} \mathbf{w}_1^T[k] & \mathbf{w}_2^T[k] & \cdots & \mathbf{w}_{N-1}^T[k] \end{bmatrix}^T \\
\mathbf{v}[k] & = \begin{bmatrix} \mathbf{v}_1^T[k] & \mathbf{v}_2^T[k] & \cdots & \mathbf{v}_{N-1}^T[k] \end{bmatrix}^T
\end{align*}
\]
Robustness against model errors

- Spatial preprocessor and adaptive stage rely on assumptions (e.g. no microphone mismatch, no reverberation, ...)
- In practice, these assumptions are often not satisfied
  - Distortion of speech component in speech reference $x_0[k]
  - Leakage of speech into noise references, i.e. $x[k] \neq 0$

Speech component in output signal gets distorted

$$z_x[k] = x_0[k - \Delta] - w^T[k]x[k]$$

- Design of robust noise reduction algorithm:
  1. Design of robust spatial preprocessor (fixed beamformer) using statistical knowledge about microphone characteristics
  2. Design of robust adaptive stage by taking speech distortion into account in optimisation criterion $\rightarrow$ speech distortion weighted multichannel Wiener filter (SDW MWF)

Limit distortion both in $x_0[k]$ and $w^T[k]x[k]$
Design of fixed beamformer

- **FIR filter-and-sum structure**: arbitrary spatial directivity pattern for arbitrary microphone configuration
- **Objective**: calculate fixed FIR filters $w_n[k]$ such that beamformer performs desired spatial and spectral filtering

Spatial directivity pattern: $H(\omega, \theta) = \frac{Z(\omega, \theta)}{S(\omega)} = w^T g(\omega, \theta)$

Desired spatial directivity pattern: $D(\omega, \theta)$
Design procedures

- Design filter \( \mathbf{w} \) such that spatial directivity pattern \( H(\omega, \theta) \) optimally fits \( D(\omega, \theta) \) → minimisation of cost function
  - Broadband problem: no design for separate frequencies \( \omega_i \)
    → design over complete frequency-angle region

- Cost functions:
  - Least-squares → quadratic function
    \[
    J_{LS}(\mathbf{w}) = \int_{\Theta} \int_{\Omega} F(\omega, \theta) \left[ H(\omega, \theta) - D(\omega, \theta) \right]^2 d\omega d\theta
    \]
    amplitude and phase
  - Non-linear cost function → iterative optimisation = complex!
    \[
    J_{NL}(\mathbf{w}) = \int_{\Theta} \int_{\Omega} F(\omega, \theta) \left[ \left| H(\omega, \theta) \right|^2 - \left| D(\omega, \theta) \right|^2 \right]^2 d\omega d\theta
    \]
    only amplitude
  - Eigenfilter based on TLS-criterion → GEVD
    \[
    J_{TLS}(\mathbf{w}) = \int_{\Theta} \int_{\Omega} F(\omega, \theta) \frac{\left[ H(\omega, \theta) - D(\omega, \theta) \right]^2}{\mathbf{w}^T \mathbf{Q}_{e}^{tot} \mathbf{w} + 1} d\omega d\theta
    \]
Simulations

Parameters:
- $N=5$, $d=4\text{cm}$
- $L=20$, $f_s=8\text{kHz}$
- Pass: $40^\circ-80^\circ$
- Stop: $0^\circ-30^\circ + 90^\circ-180^\circ$

Delay-and-sum

Non-linear procedure

TLS-Eigenfilter
Robust broadband beamforming

- Small deviations from assumed microphone characteristics (gain, phase, position) → large deviations from desired directivity pattern, especially for small-size microphone arrays
- In practice microphone characteristics are never exactly known
  - Measurement or calibration procedure
  - Incorporate specific (random) deviations in design

\[ A_n(\omega, \theta) = a_n(\omega, \theta) \cdot e^{-j\psi_n(\omega, \theta)} \cdot e^{-j\omega \delta_n \cos \theta f_s / c} \]

- Consider all feasible microphone characteristics and optimise
  - average performance using probability as weight

\[ J_{\text{mean}} = \int_{A_0} \ldots \int_{A_{N-1}} J(A_0, \ldots, A_{N-1}) f_A(A_0) \ldots f_A(A_{N-1}) dA_0 \ldots dA_{N-1} \]

  - requires statistical knowledge about probability density functions
  - worst-case performance → minimax optimisation problem
    - finite grid of microphone characteristics → high complexity
Simulations

- Non-linear design procedure
- N=3, positions: [-0.01 0 0.015] m, L=20, f_s=8 kHz
- Passband = 0°-60°, 300-4000 Hz (endfire)
  Stopband = 80°-180°, 300-4000 Hz
- Robust design - average performance:
  Uniform pdf = gain (0.85-1.15) and phase (-5°-10°)
- Deviation = [0.9 1.1 1.05] and [5° -2° 5°]

<table>
<thead>
<tr>
<th>Design</th>
<th>$J$</th>
<th>$J_{dev}$</th>
<th>$J_{mean}$</th>
<th>$J_{max}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-robust</td>
<td>0.1585</td>
<td>87.131</td>
<td>275.40</td>
<td>3623.6</td>
</tr>
<tr>
<td>Average cost</td>
<td>0.2196</td>
<td>0.2219</td>
<td>0.3371</td>
<td>0.4990</td>
</tr>
<tr>
<td>Maximum cost</td>
<td>0.1707</td>
<td>0.1990</td>
<td>0.4114</td>
<td>0.4167</td>
</tr>
</tbody>
</table>
### Introduction
- Fixed beamforming
  - Broadband design
  - Robustness
- Adaptive stage
- Implementation
- Conclusions

### Simulations

<table>
<thead>
<tr>
<th>Non-robust design</th>
<th>Robust design</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Graph" /></td>
<td><img src="image2" alt="Graph" /></td>
</tr>
<tr>
<td><img src="image3" alt="Graph" /></td>
<td><img src="image4" alt="Graph" /></td>
</tr>
</tbody>
</table>

- **Angle (deg)**
- **Frequency (Hz)**
- **dB**
<table>
<thead>
<tr>
<th>Non-robust design</th>
<th>Robust design</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="robust1.avi" alt="Non-robust design" /></td>
<td><img src="robust2.avi" alt="Robust design" /></td>
</tr>
</tbody>
</table>
Design of robust adaptive stage

- Distorted speech in output signal: \( z_x[k] = x_0[k - \Delta] - w^T[k] x[k] \)

- Robustness: limit \( w^T[k] x[k] \) by controlling adaptive filter \( w[k] \)
  - Quadratic inequality constraint (QIC): \( \|w[k]\| \leq \beta \)
    - conservative approach, constraint \( \neq f(\text{amount of leakage}) \)
  - Take speech distortion into account in optimisation criterion
    \[
    \min_{w[k]} E\left\{ (v_0[k - \Delta] - w^T[k] v[k])^2 \right\} + \frac{1}{\mu} E\left\{ (w^T[k] x[k])^2 \right\}
    \]
    noise reduction \quad speech distortion
  - \( 1/\mu \) trades off noise reduction and speech distortion
    - \( 1/\mu = 0 \) or no speech leakage \( \rightarrow \) GSC
    - \( 1/\mu = 1 \) \( \rightarrow \) MMSE estimate of speech component in speech reference signal
  - Regularisation term \( \sim \) amount of speech leakage

\[\rightarrow\] Limit speech distortion, while not affecting noise reduction performance in case of no model errors \( \leftrightarrow \) QIC
**Wiener solution**

- **Optimisation criterion:**

\[
\begin{align*}
\min_{w[k]} E\left\{ (v_0[k - \Delta] - w^T[k]v[k])^2 \right\} + \frac{1}{\mu} E\left\{ (w^T[k]x[k])^2 \right\}
\end{align*}
\]

\[
w[k] = \left[ \frac{1}{\mu} E\{x[k]x^T[k]\} + E\{v[k]v^T[k]\} \right]^{-1} E\{v[k]v_0[k - \Delta]\}
\]

- **Problem:** clean speech \(x[k]\) and hence speech correlation matrix \(E\{x[k]x^T[k]\}\) are unknown!

**Approximation:** \(E\{x[k]x^T[k]\} = E\{y[k]y^T[k]\} - E\{v[k]v^T[k]\}\)

\[
w[k] = \left[ \frac{1}{\mu} E\{x[k]x^T[k]\} + \left( 1 - \frac{1}{\mu} \right) E\{v[k]v^T[k]\} \right]^{-1} E\{v[k]v_0[k - \Delta]\}
\]

- **VAD (voice activity detection) mechanism required!**

- **Introduction**
- **Fixed beamforming**
- **Adaptive stage**
  - SP SDW MWF
  - Experimental validation
- **Implementation**
- **Conclusions**
Spatially-preprocessed SDW-MWF (1)

- In new optimisation criterion additional filter $w_0[k]$ on speech reference signal may be added

$$\min_{w[k]} E\left\{ (v_0[k - \Delta] - w^T[k] v[k])^2 \right\} + \frac{1}{\mu} E\left\{ (w^T[k] x[k])^2 \right\}$$

$$w[k] = [w_0^T[k] \quad w_1^T[k] \quad \cdots \quad w_{N-1}^T[k]]^T$$

⇒ Speech Distortion Weighted Multichannel Wiener Filter (SDW-MWF)
Spatially-preprocessed SDW-MWF (2)

- SP-SDW-MWF encompasses both GSC and SDW-MWF as special cases:
  - No filter $w_0[k]$ on speech reference
    - speech distortion regularised GSC (SDR-GSC)
      - regularisation term added to GSC: the larger the speech leakage, the larger the regularisation
      - special case: $1/\mu = 0$ corresponds to traditional GSC
      - SDR-GSC outperforms GSC with quadratic inequality constraint
  - Filter $w_0[k]$ on speech reference
    - SDW-MWF on pre-processed microphone signals
      - in absence of model errors = cascade of GSC + single-channel postfilter (SDW Wiener filter)
      - Model errors do not effect its performance!

Outperforms QIC-GSC and SDR-GSC
Experimental validation (1)

- Set-up:
  - 3-mic BTE mounted on dummy head in office room (d = 1cm, 1.5cm)
  - Speech source in front of dummy head (90°)
  - 5 stationary speech-like noise sources: 75°, 120°, 180°, 240°, 285°
  - Microphone gain mismatch Ψ₂ at 2nd microphone

- Performance measures:
  - Intelligibility-weighted signal-to-noise ratio
    \[ \text{SNR}_{\text{intellig}} = \sum_{i=1}^{I} I_i \text{SNR}_i \]
    - \( I_i \) = band importance of \( i \)th one-third octave band
    - \( \text{SNR}_i \) = signal-to-noise ratio in \( i \)th one-third octave band
  - Intelligibility-weighted spectral distortion
    \[ \text{SD}_{\text{intellig}} = \sum_{i=1}^{I} I_i \text{SD}_i \]
    - \( \text{SD}_i \) = average spectral distortion in \( i \)th one-third octave band

\[ \text{SD}_i = \frac{\int_{2^{-1/6}}^{2^{1/6}} f_{c,i} \left| 10 \log_{10} G_x(f) \right| df}{\left(2^{1/6} - 2^{-1/6}\right)f_{c,i}} \]

\[ G_x(f) = \frac{E\left\{Z_x^2(f)\right\}}{E\left\{X^2(f)\right\}} \]

(Power Transfer Function for speech component)
Experimental validation (2)

- **SDR-GSC: \( w_0 = 0 \)**
  - GSC (1/\( \mu = 0 \)): degraded performance if significant leakage
  - 1/\( \mu > 0 \) increases robustness (speech distortion \( \leftrightarrow \) noise reduction)

- **SP-SDW-MWF: \( w_0 \neq 0 \)**
  - No mismatch: same \( \Delta \text{SNR}_{\text{ intellig}} \) as SDR-GSC, larger \( \Delta \text{SD}_{\text{ intellig}} \) due to SDW-WF post-filter
  - Performance is not degraded by mismatch
Experimental validation (3)

- GSC with QIC (\( \| w[k] \| \leq \beta \)) : QIC increases robustness GSC
  - QIC \( \neq f( \text{amount of speech leakage} ) \) \( \rightarrow \) less noise reduction than SDR-GSC for small mismatch
- For large mismatch: less noise reduction than SP-SDW-MWF

SP-SDW-MWF achieves better noise reduction than QIC-GSC, for a given maximum speech distortion level
**Low-cost implementation (1)**

- **Algorithms** (in decreasing order of complexity):
  - GSVD-based – *chic et très cher*
  - QRD-based, fast QRD-based – *chic et moins cher*
  - Stochastic gradient algorithms – *chic et pas cher*

- **Stochastic gradient algorithm (time-domain):**

  - Cost function
    
    \[ J(w) = E\left\{ \left( v_0[k - \Delta] - w^T[k]v[k]\right)^2 \right\} + \frac{1}{\mu} E\left\{ w^T[k]x[k] \right\} \]

    results in LMS-based updating formula

    \[ w[k+1] = w[k] + \rho[v[k]v_0[k - \Delta] - v^T[k]w[k]] - \frac{1}{\mu}x[k]x^T[k]w[k] \]

    - Classical GSC
    - regularisation term

  - Allows transition to classical LMS-based GSC by tuning some parameters (1/\(\mu\), \(w_0\))
Low-cost implementation (2)

- **Stochastic gradient algorithm (time-domain):**
  - Regularisation term $-\frac{1}{\mu}x[k]x^T[k]w[k]$ is unknown
    - Store samples in memory buffer during speech-and-noise periods and approximate regularisation term
      - Large buffer required
    - Better estimate of regularisation term can be obtained by smoothing (low-pass filtering)

- **Stochastic gradient algorithm (frequency-domain):**
  - Block-based implementation: improve gradient estimate by averaging over $K$ samples
  - Frequency-domain: fast convolution and fast correlation
    - Complexity reduction
    - Tuning of $\rho$ and $1/\mu$ per frequency
    - Still large memory requirement due to data buffers
  - Approximations allow to replace data buffers by correlation matrices in frequency-domain $\rightarrow$ memory reduction
Complexity + memory

- Parameters: \( N = M = 2 \) (\#mics, \#adaptive filters), \( L = 64, f_s = 16\text{kHz}, L_{buf} = 10000 \)
- Computational complexity:

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Complexity</th>
<th>MIPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>QIC-GSC</td>
<td>((3N-1)\text{FFT} + 14N - 12)</td>
<td>1.38</td>
</tr>
<tr>
<td>SDW-MWF (no approximation)</td>
<td>((3M+5)\text{FFT} + 28M + 6)</td>
<td>3.46</td>
</tr>
<tr>
<td>SDW-MWF (approximations)</td>
<td>((3M+2)\text{FFT} + 8M^2 + 14M + 3)</td>
<td>2.8</td>
</tr>
</tbody>
</table>

Complexity comparable to FD implementation of QIC-GSC

- Memory requirement:

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Memory</th>
<th>kWords</th>
</tr>
</thead>
<tbody>
<tr>
<td>QIC-GSC</td>
<td>(4(N-1)L + 6L)</td>
<td>0.64</td>
</tr>
<tr>
<td>SDW-MWF (no approximation)</td>
<td>(2ML_{buf} + 6LM + 7L)</td>
<td>41.22</td>
</tr>
<tr>
<td>SDW-MWF (approximations)</td>
<td>(4LM^2 + 6LM + 7L)</td>
<td>2.24</td>
</tr>
</tbody>
</table>

Substantial memory reduction through approximations
Conclusions

- **Design of robust multimicrophone noise reduction algorithm:**
  - Design of robust fixed spatial preprocessor
    - need for statistical information about microphones
  - Design of robust adaptive stage
    - take speech distortion into account in cost function
  - **Spatially pre-processed SDW Multichannel Wiener Filter**
- SP-SDW-MWF encompasses GSC and MWF as special cases
- **Experimental results:**
  - SP-SDW-MWF achieves better noise reduction than QIC-GSC, for a given maximum speech distortion level
  - Filter $w_0$ improves performance in presence of model errors
- **Implementations:** Stochastic gradient algorithms available at affordable complexity and memory
- **Further research:** robustness against VAD-errors
  - e.g. parameters dependent on input SNR
Relevant publications


Available at SISTA publication engine: http://www.esat.kuleuven.ac.be/~sistawww/cgi-bin/pub.pl