Hearing devices using wireless acoustic sensor networks

Prof. Dr. Simon Doclo
Signal Processing Group, University of Oldenburg

simon.doclo@uni-oldenburg.de
http://www.sigproc.uni-oldenburg.de

Hearing Aid Developers Forum, 09.06.2011

Outline

- Use of multiple microphones in hearing devices
  - Monaural → Binaural → external microphones
- Binaural signal processing
  - Objective: noise reduction and binaural cue preservation
  - Algorithms: binaural beamforming, time-frequency masking, Multi-channel Wiener filter
  - Experimental results
  - Bandwidth reduction: iterative distributed MWF
- Wireless acoustic sensor networks
  - Algorithms: extension of distributed MWF
  - Effect of bitrate on performance
- Conclusions and future work

Hearing aids

- Problems: background noise, directional hearing
  - signal processing to selectively enhance useful speech signal and improve speech intelligibility
  - signal processing to preserve directional hearing (binaural auditory cues) and spatial awareness
- Digital hearing instruments allow for advanced acoustical signal pre-processing
  - multiple microphones available → spectral + spatial processing
  - noise reduction (beamforming), computational auditory scene analysis (source localisation, environment classification, …)

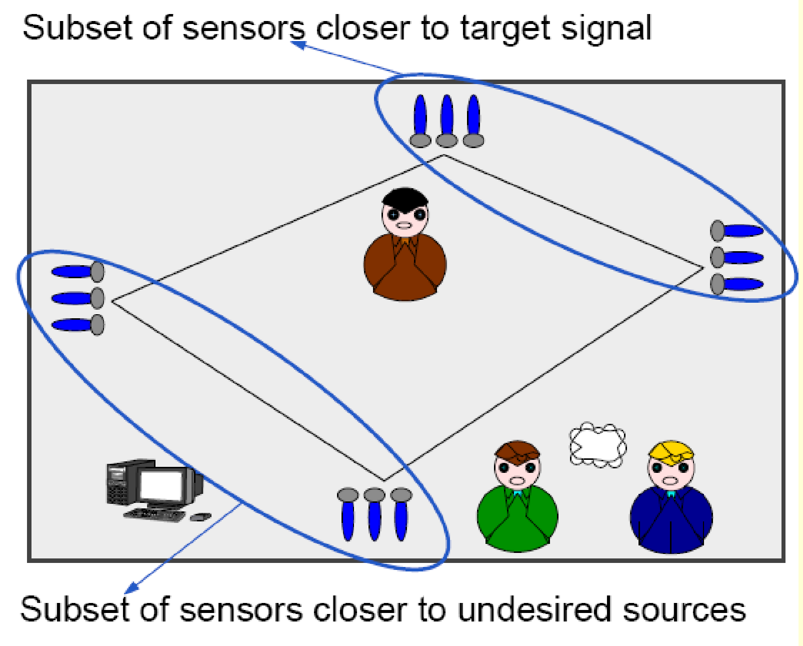
Monaural (2-3)

Binaural

External microphones

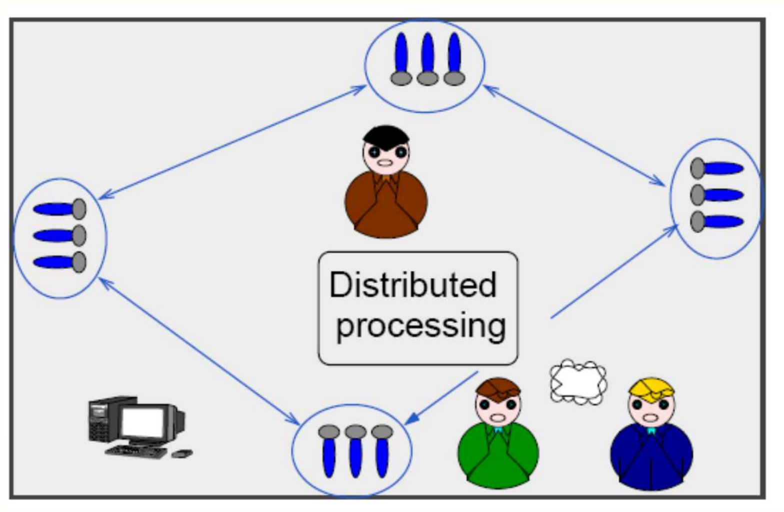
Acoustic sensor networks

- Signal acquisition in **adverse acoustic environments**:
  - Microphones at large distance from speaker → background noise and reverberation
- **Acoustic sensor networks**:
  - Network of a large number of spatially distributed nodes (each with one or multiple microphones)
  - Wireless data transmission
  - More information about spatial noise field (microphones with higher SNR, direct-to-reverberant ratio)
- **Objectives:**
  - speech enhancement
  - source localisation
  - CASA



Acoustic sensor networks

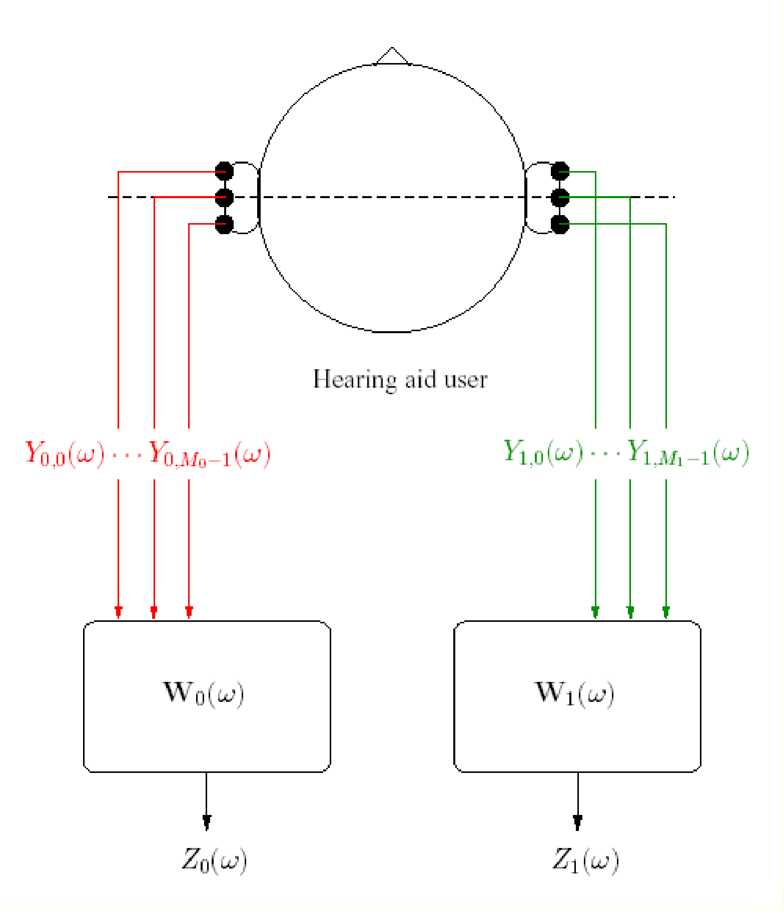
- **Challenges:**
  - *Dynamic array configuration*: large number of microphones at unknown positions, dynamic subset selection
  - *Distributed and collaborative algorithms:* power and complexity constraints, effect of limited bandwidth
  - *Calibration and synchronisation issues:* same time basis
- **Prototype applications:**
  - Hearing aids using extra microphones (room, other HA, …)
  - Video-conferencing using all microphones on laptops / room
  - Surveillance



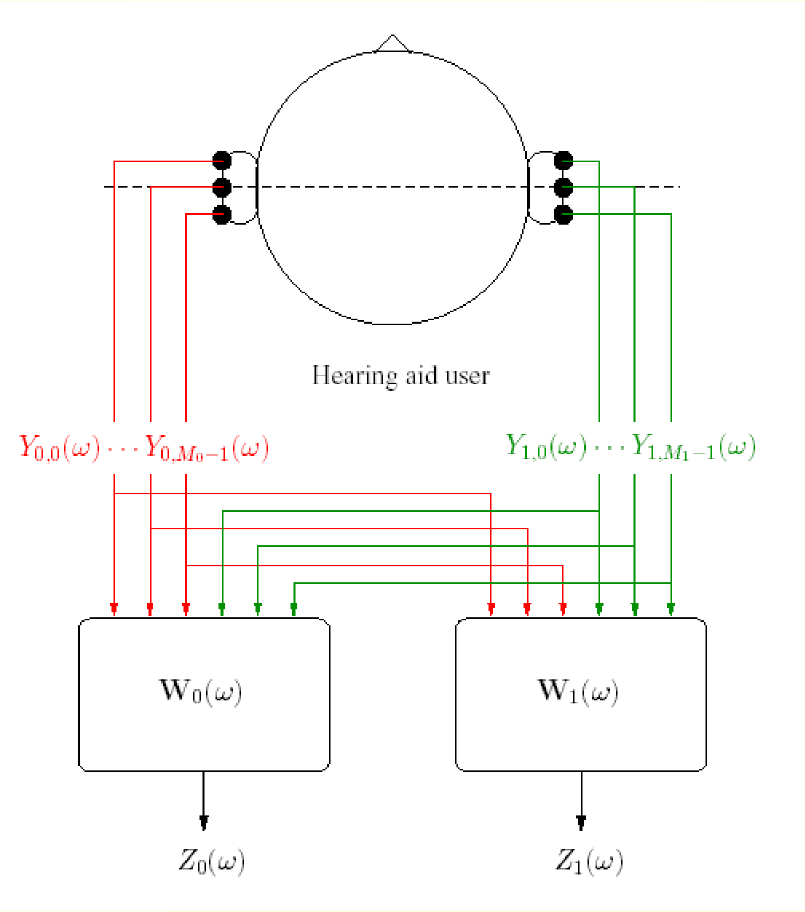
Binaural processing

Bilateral vs. Binaural

Bilateral system



Binaural system



⊖ Independent left/right processing:
Preservation of binaural cues
(ILD/ITD) for localisation ?

⊕ More microphones:
→ better performance ?
→ preservation of binaural cues ?

⊖ Need for wireless binaural link

Bilateral vs. Binaural

- Bilateral system
  - Independent processing of left and right hearing aid
  - Negative effect on localisation cues and intelligibility through binaural hearing advantage [Van den Bogaert et al., 2006]
- Binaural system
  - Cooperation between left and right hearing aid (e.g. wireless link) → **centralised** vs. **distributed** processing
  - Bandwidth constraint and latency of wireless link

Objectives/requirements for binaural algorithm:

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

- Introduction
- Binaural processing
  - -Algorithms
  - -Experiments
  - -Distributed MWF
- Acoustic sensor networks
- Conclusion

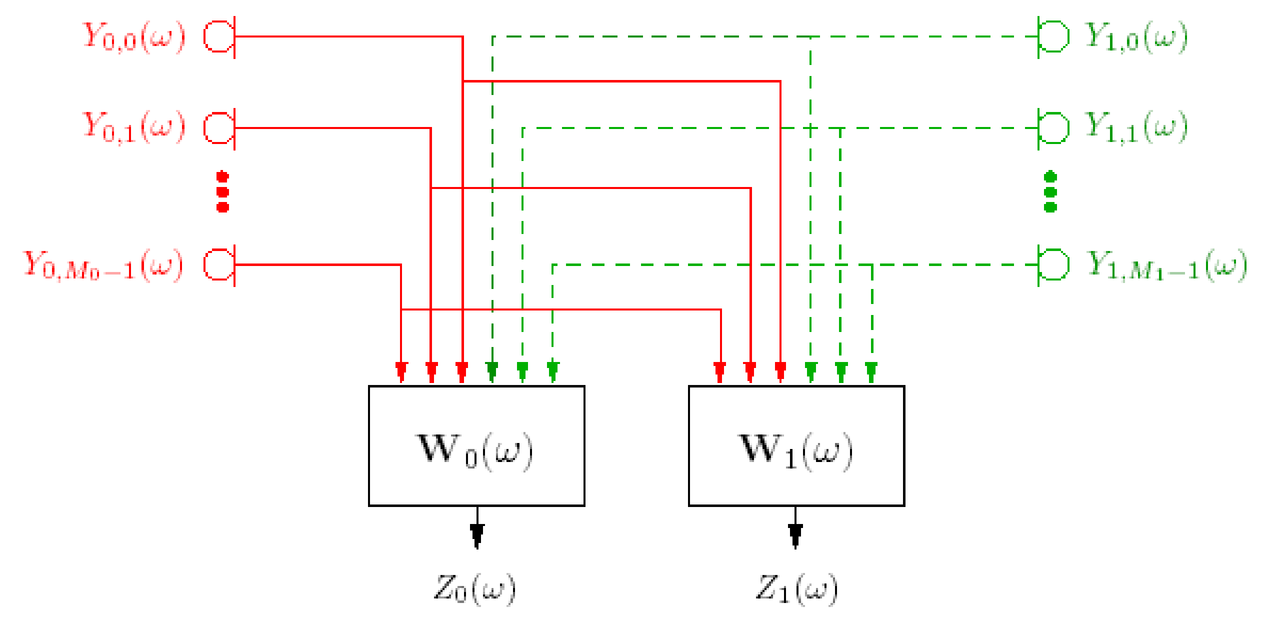
Binaural noise reduction techniques

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

$$Y_{0,m}(\omega) = \underset{\substack{\uparrow \\ \text{speech}}}{X_{0,m}(\omega)} + \underset{\substack{\uparrow \\ \text{noise}}}{V_{0,m}(\omega)}, \quad m = 0\ldots M_0 - 1$$

- Use all microphone signals to compute output signal at both ears

$$\boxed{Z_0(\omega) = \mathbf{W}_0^H(\omega)\mathbf{Y}(\omega), \quad Z_1(\omega) = \mathbf{W}_1^H(\omega)\mathbf{Y}(\omega)}$$

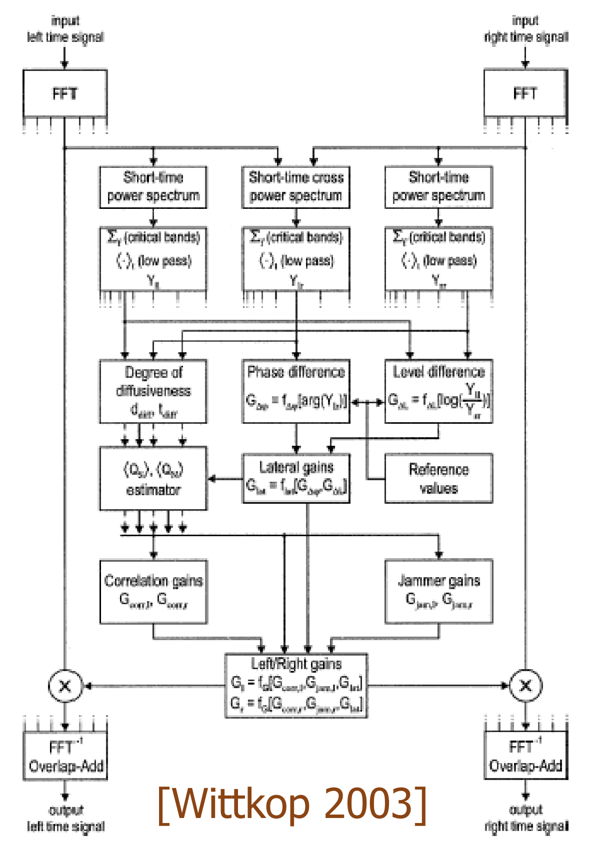


Binaural noise reduction techniques

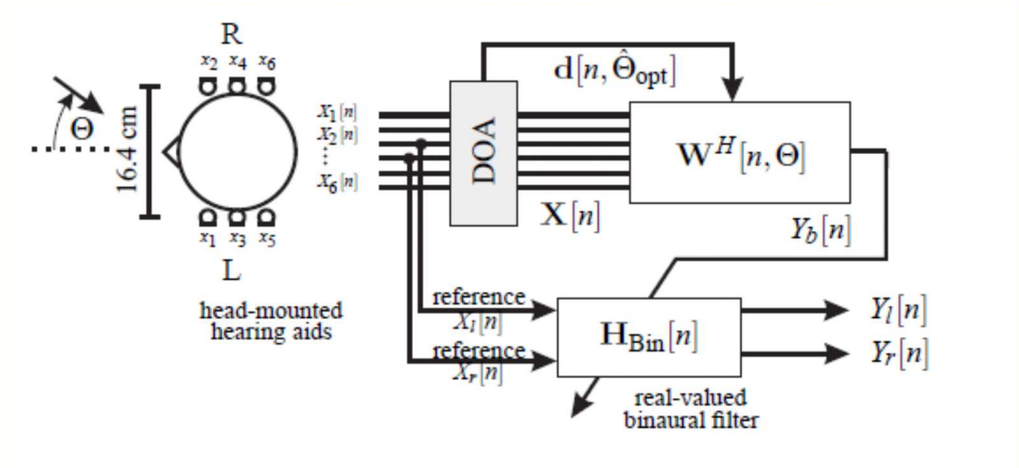
- Time-frequency post-processing/masking:
  - Computation and application of **real-valued** binaural mask based on binaural and temporal/spectral cues
  - Can be merged with MVDR-beamformer or ICA-based processing

⊕ Good preservation of binaural cues for **all** sources

⊖ “single-microphone spectral enhancement” artefacts at low SNRs



[Wittkop 2003]



Beamformer: $\mathbf{W}_b = \frac{\Gamma^{-1}\mathbf{d}}{\mathbf{d}^H\Gamma^{-1}\mathbf{d}} \quad \Rightarrow \quad Y' = \mathbf{W}_b^H\mathbf{Y}$

Post-Filter: $H_p = \frac{(|d_0|^2+|d_1|^2)|Y'|^2}{|Y_0|^2+|Y_1|^2} \quad \Rightarrow \quad Z = H_p \begin{bmatrix} Y_0 \\ Y_1 \end{bmatrix}$

[Rohdenburg 2009, Reindl 2010, Saruwatari 2010]

- Introduction
- Binaural processing
  - -Algorithms
  - -Experiments
  - -Distributed MWF
- Acoustic sensor networks
- Conclusion

Binaural noise reduction techniques

- Binaural multi-channel Wiener filter: estimate of speech component in microphone signal at both ears (usually front mic) + trade-off between noise reduction and speech distortion

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

speech component in front microphones

speech distortion

noise reduction

estimate

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

- Estimate $\mathbf{R}_y$ during speech-dominated time-frequency segments, estimate $\mathbf{R}_v$ during noise-dominated segments, requiring robust voice activity detection (VAD) mechanism
- No assumptions about positions of microphones and sources
- Different implementations:
  - Batch (off-line) vs. adaptive (update correlation matrices)
  - Using spatial prediction (SP) between speech components [Chen 2008]

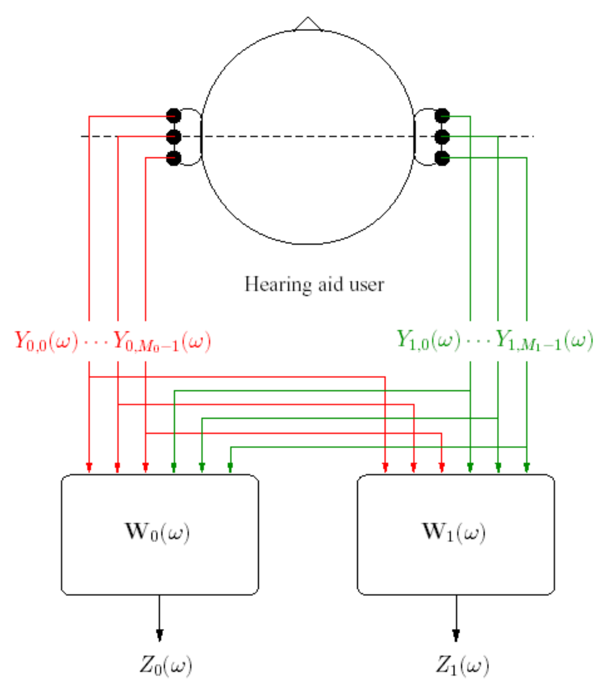
[S. Doclo, S. Gannot, M. Moonen, A. Spriet, Handbook on Array Processing and Sensor Networks, Wiley, 2010.]

[B. Cornelis, S. Doclo, T. Van den Bogaert, J. Wouters, M. Moonen, IEEE Trans. Audio, Speech and Language Processing, Feb. 2010.]

[S. Doclo, T.J. Klasen, M. Moonen, T. Van den Bogaert, J. Wouters, R.P. Derleth, S. Korl, US2010002886.]

Binaural noise reduction techniques

- Binaural multi-channel Wiener filter:
  - Preservation of binaural cues (ITD-ILD)
    - Speech cues are preserved, no a-priori assumptions
    - **Noise cues are distorted**
  - **Extensions** in order to preserve binaural cues of both speech and noise sources, without substantially compromising noise reduction
    - Partial noise estimation (MWFv)
    - Extension with Interaural Transfer Function (MWF-ITF)

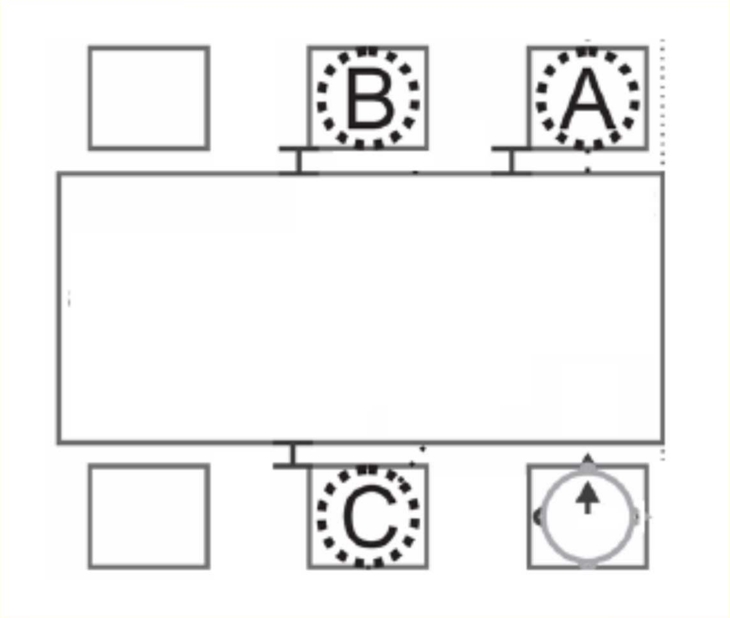


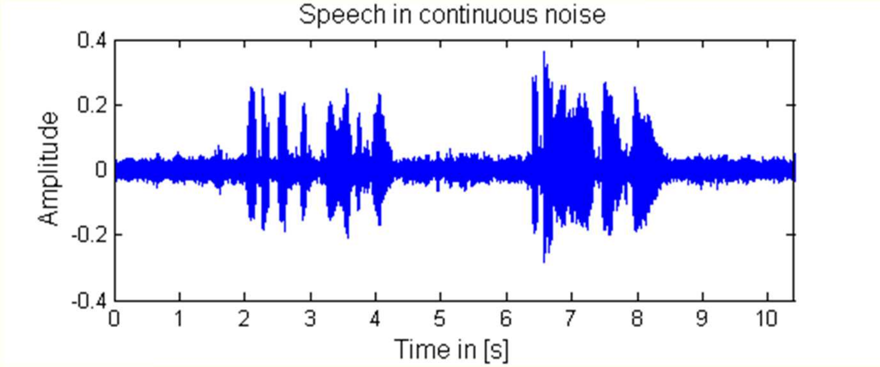
$$J_{SDW\eta}(\mathbf{W}) = E\left\{\left\|\begin{bmatrix} X_L - \mathbf{W}_L{}^H\mathbf{X} \\ X_R - \mathbf{W}_R{}^H\mathbf{X}\end{bmatrix}\right\|^2 + \mu\left\|\begin{bmatrix} \eta V_L - \mathbf{W}_L{}^H\mathbf{V} \\ \eta V_R - \mathbf{W}_R{}^H\mathbf{V}\end{bmatrix}\right\|^2\right\}, \quad 0 \leq \eta \leq 1$$

$$\begin{aligned} J_{ITF}(\mathbf{W}) &= J_{SDW}(\mathbf{W}) \\ &+ \alpha\, E\left\{\left|\mathbf{W}_L^H\mathbf{X} - ITF_{in}^{x}\ \mathbf{W}_R^H\mathbf{X}\right|^2\right\} + \beta\, E\left\{\left|\mathbf{W}_L^H\mathbf{V} - ITF_{in}^{v}\ \mathbf{W}_R^H\mathbf{V}\right|^2\right\}\end{aligned}$$

Experimental results

- **Acoustic environment**



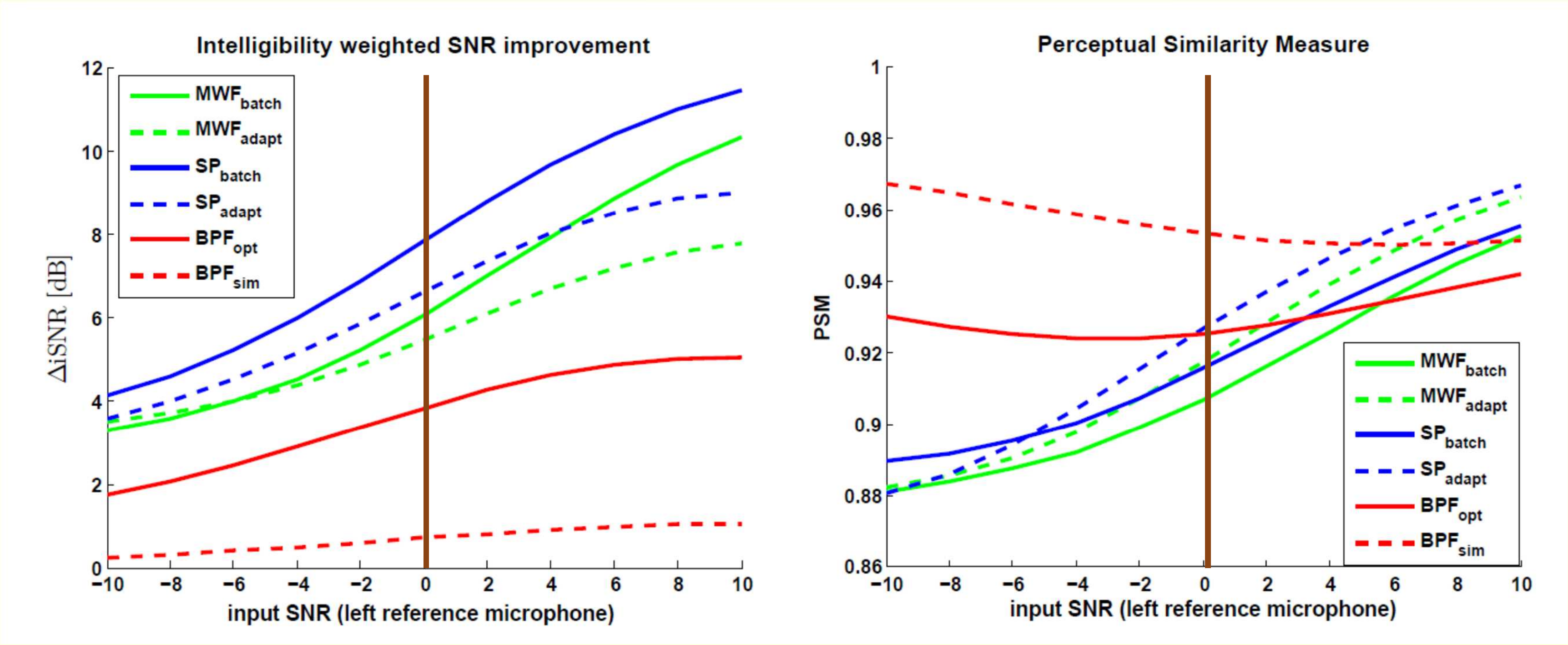


- Cafeteria with recorded babble noise and simulated speaker at position B
- Binaural hearing aid with 3 microphones
- German sentences taken from OLSA speech material
- Speech in continuous babble noise
- f_s: 16 kHz, WOLA, FFT-size: 256 samples, Overlap: 75%

http://medi.uni-oldenburg.de/hrir/ [Kayser et al. 2009]

Experimental results

- **Objective Evaluation**
  - Intelligibility weighted SNR improvement
  - Perceptual Similarity Measure (PSM)

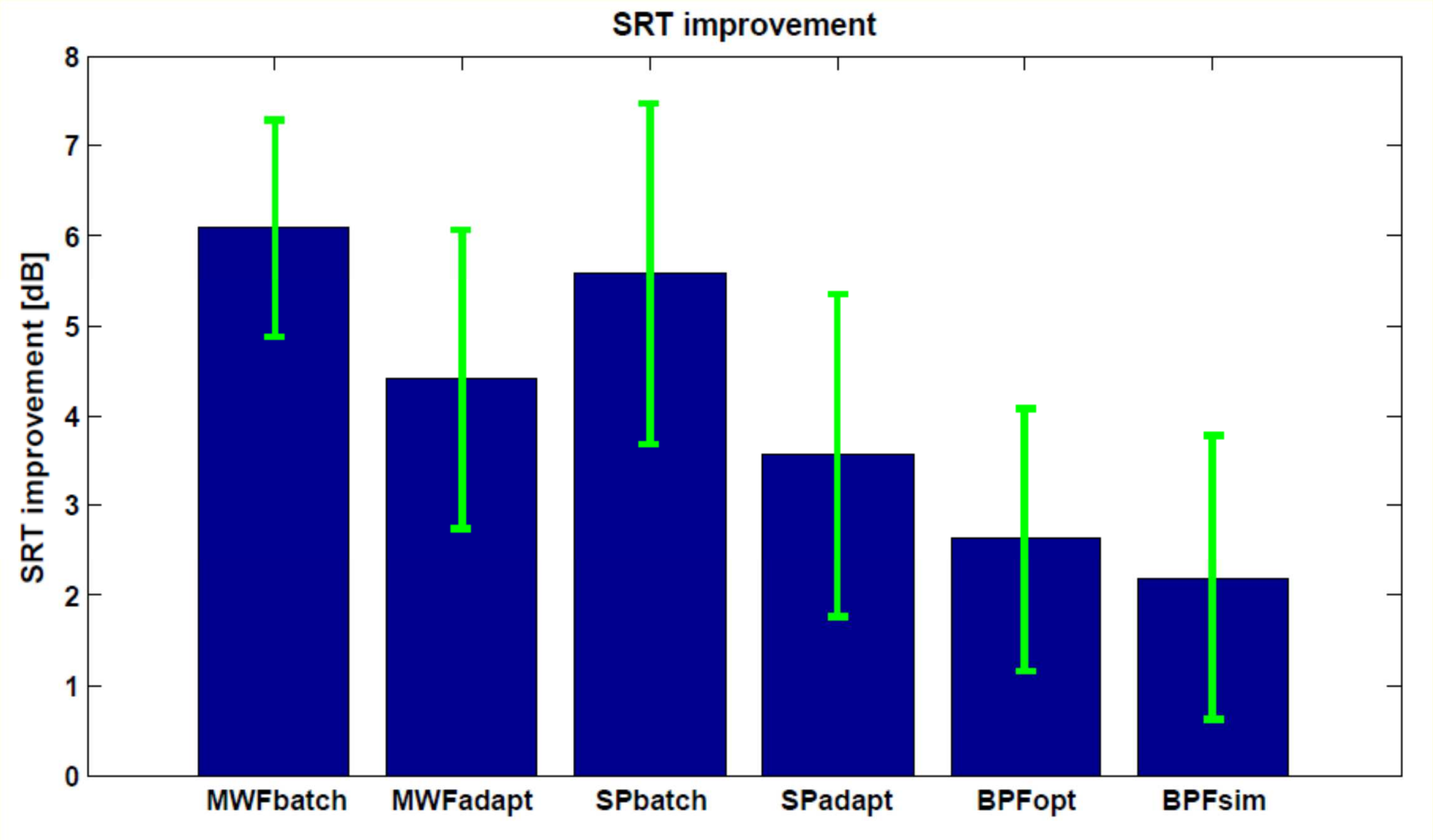


SNR	Orig.	MWF	SP	BF + Postfilt
0 dB	🔊	🔊	🔊	🔊

[D. Marquardt, V. Hohmann, S. Doclo, DAGA 2011]

Experimental results

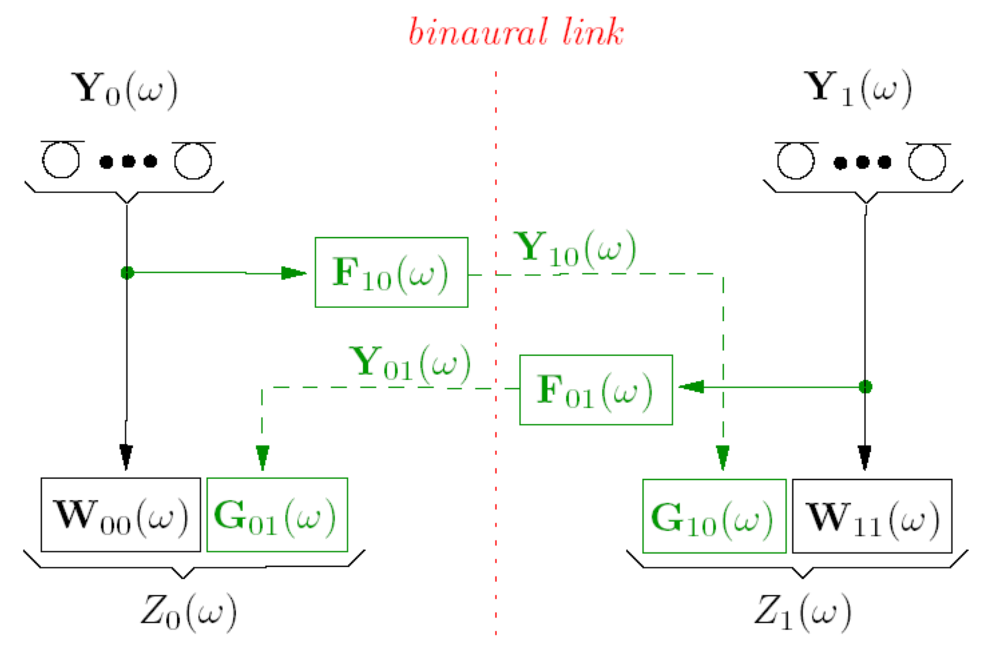
- **Subjective Evaluation**
  - Improvement of Speech Reception Threshold (SRT)
  - Oldenburg Sentence Test (10 NH subjects)
  - Binaural presentation using headphones



[D. Marquardt, V. Hohmann, S. Doclo, DAGA 2011]

Distributed MWF

- Binaural MWF
  - o **all** microphone signals are transmitted over wireless link
- Reduce bandwidth requirement of wireless link by transmitting **one** signal from contralateral ear
  - o Raw microphone signal (e.g. front)
  - o Output of fixed (e.g. superdirective) beamformer
  - o MWF-estimate using only contralateral microphone signals
  - o Iterative distributed binaural MWF scheme (DB-MWF)

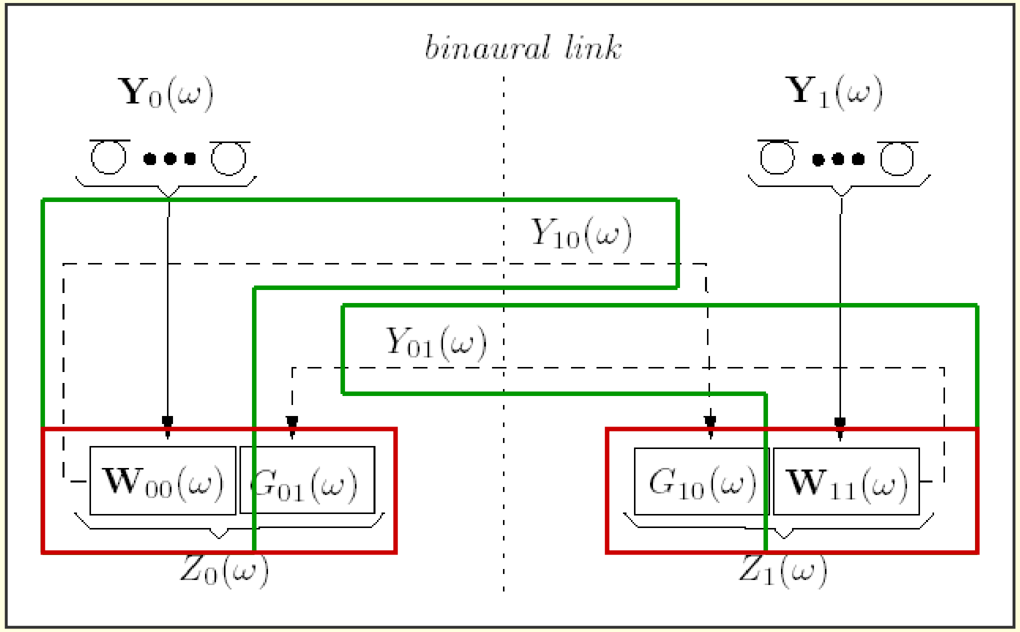


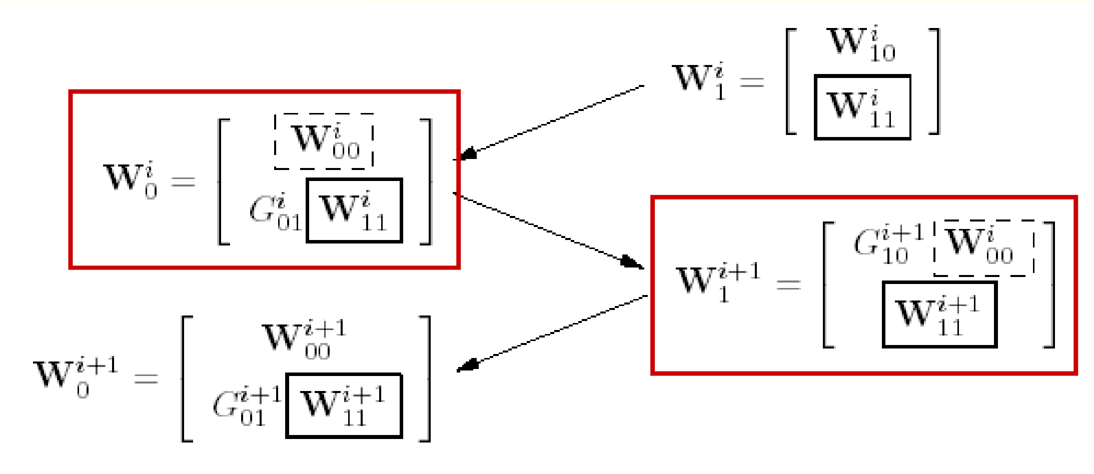
[S. Doclo, T. Van den Bogaert, M. Moonen, J. Wouters, IEEE Trans. Audio, Speech and Language Processing, Jan. 2009.]

- Introduction
- Binaural processing
  - -Algorithms
  - -Experiments
  - -Distributed MWF
- Acoustic sensor networks
- Conclusion

Distributed MWF

- Iterative procedure
  - In each iteration $\mathbf{F}_{10}$ is equal to $\mathbf{W}_{00}$ from previous iteration, and $\mathbf{F}_{01}$ is equal to $\mathbf{W}_{11}$ from previous iteration





$$\mathbf{W}_1^i = \begin{bmatrix} \mathbf{W}_{10}^i \\ \boxed{\mathbf{W}_{11}^i} \end{bmatrix}$$

$$\mathbf{W}_0^i = \begin{bmatrix} & \boxed{\mathbf{W}_{00}^i} \\ G_{01}^i & \boxed{\mathbf{W}_{11}^i} \end{bmatrix}$$

$$\mathbf{W}_1^{i+1} = \begin{bmatrix} G_{10}^{i+1} & \boxed{\mathbf{W}_{00}^i} \\ & \boxed{\mathbf{W}_{11}^{i+1}} \end{bmatrix}$$

$$\mathbf{W}_0^{i+1} = \begin{bmatrix} & \mathbf{W}_{00}^{i+1} \\ G_{01}^{i+1} & \boxed{\mathbf{W}_{11}^{i+1}} \end{bmatrix}$$

Distributed MWF

- Single speech source
  - MWF cost function decreases in each step of iteration

$$\boxed{J\left(\begin{bmatrix}\mathbf{W}_0^{i+1}\\ \mathbf{W}_1^{i+1}\end{bmatrix}\right) \leq J\left(\begin{bmatrix}\mathbf{W}_0^{i}\\ \mathbf{W}_1^{i}\end{bmatrix}\right)}$$

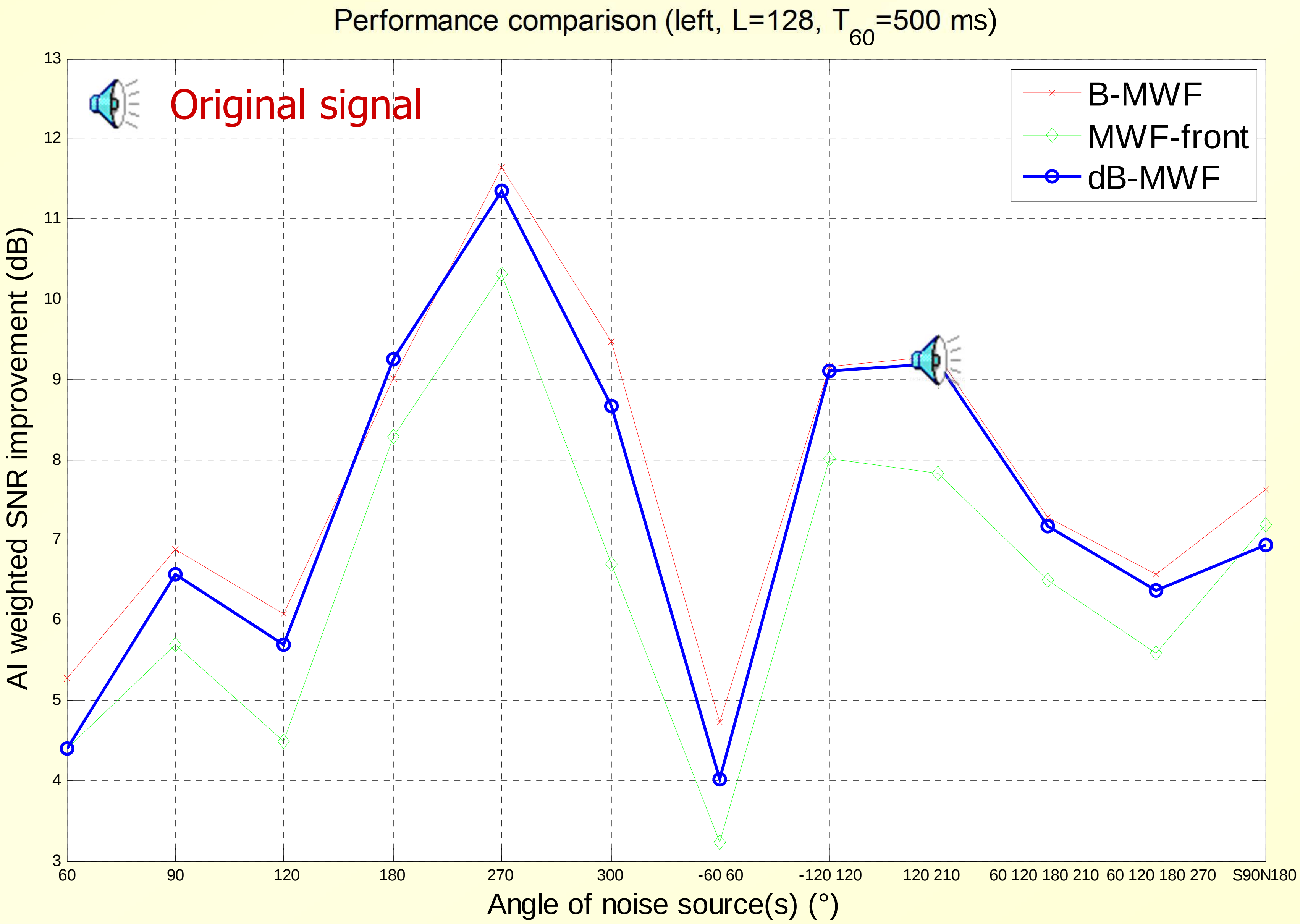
  - Remarkably: convergence to B-MWF solution (!)

$$\rightarrow \boxed{\mathbf{W}_0^{\infty} = \mathbf{W}_0^{m}, \quad \mathbf{W}_1^{\infty} = \mathbf{W}_1^{m}}$$

- General case where **R**$_x$ is not a rank-1 matrix
  - MWF cost function does not necessarily decrease in each iteration
  - usually no convergence to optimal B-MWF solution
  - Although $J_0(\mathbf{W}_0^{\infty}) \geq J_0(\mathbf{W}_0^{m}), \quad J_1(\mathbf{W}_1^{\infty}) \geq J_1(\mathbf{W}_1^{m})$, DB-MWF procedure can be used in practice and approaches binaural MWF performance

- Introduction
- Binaural processing
  - -Algorithms
  - -Experiments
  - -Distributed MWF
- Acoustic sensor networks
- Conclusion

Distributed MWF



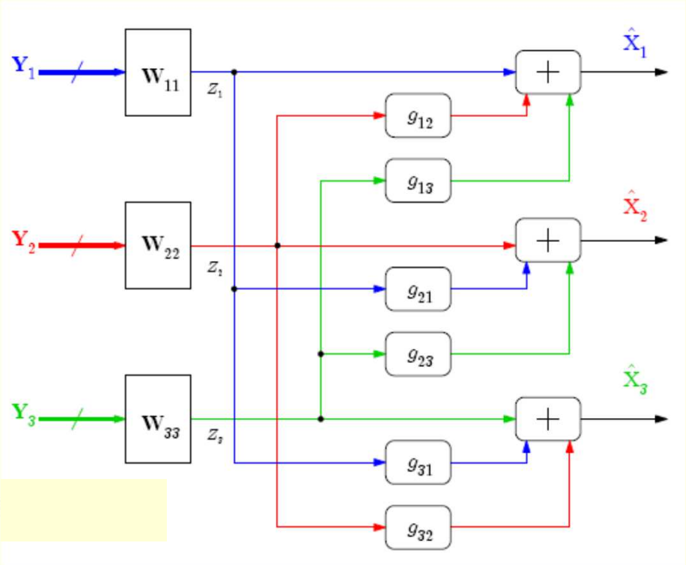
Real-world performance of db-MWF close to full binaural MWF !

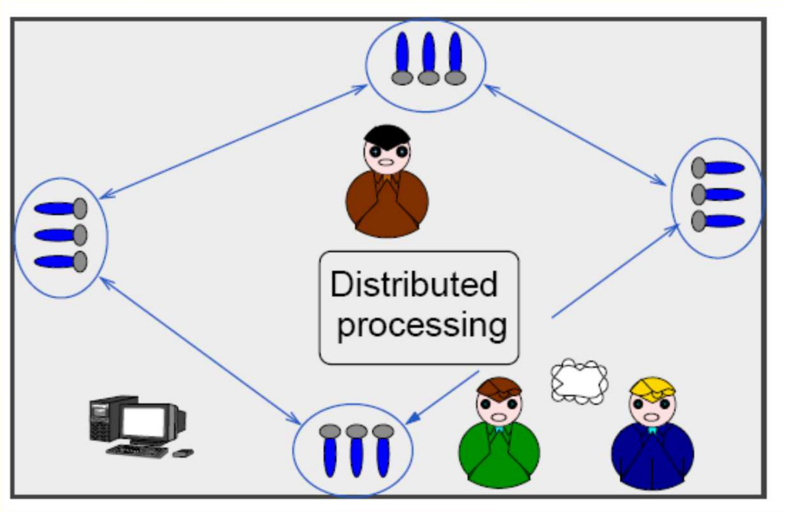
Acoustic sensor networks

Acoustic sensor networks

- Now consider more than 2 sensor nodes…
- Recently has become quite a **hot research topic**
  - Distributed MWF: extension to multiple sensor arrays and multiple desired sources (DANSE) [Bertrand 2010]
  - Distributed MVDR/LCMV-beamformer [Golan 2010, Bertrand 2011]
  - Performance analysis of a randomly spaced wireless microphone array [Golan 2011]
  - Dynamic signal combining (no synchronisation required)

    [Matheja 2011, Srinivasan 2011, Stenger 2011]

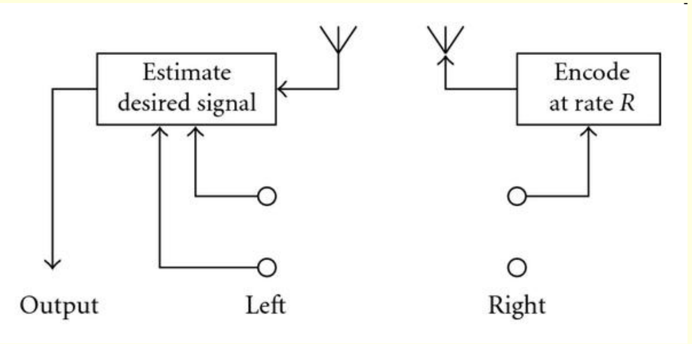


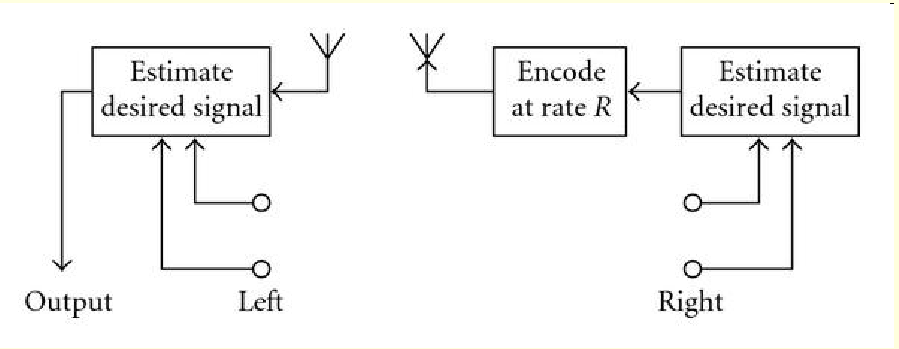


- Introduction
- Binaural processing
- Acoustic sensor networks
  - -Rate constraints
  - -Future work
- Conclusion

Rate constraints

- Investigate effect of capacity of wireless link → encode signal(s) at finite bit-rate *R* before transmission



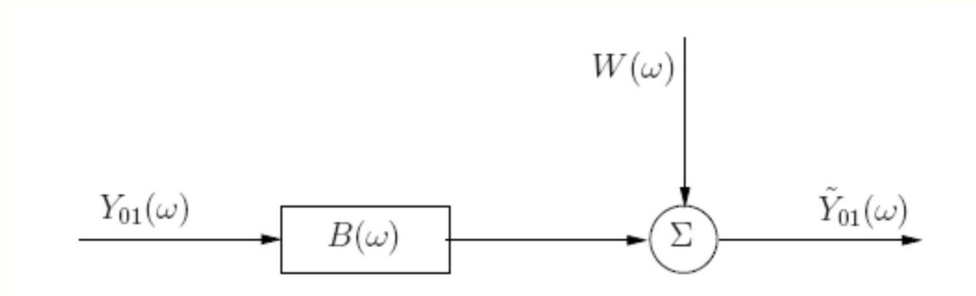


- Rate-distortion:

$$
\begin{aligned}
R(\lambda) &= \frac{1}{4\pi}\int_{-\infty}^{\infty} \max\left(0, \log_2 \frac{\Phi_Y^{01}(\omega)}{\lambda}\right) d\omega \\
D(\lambda) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \min\left(\lambda, \Phi_Y^{01}(\omega)\right) d\omega ,
\end{aligned}
$$

PSD of transmitted signal

- Upper bound on achievable performance can be calculated using forward channel representation

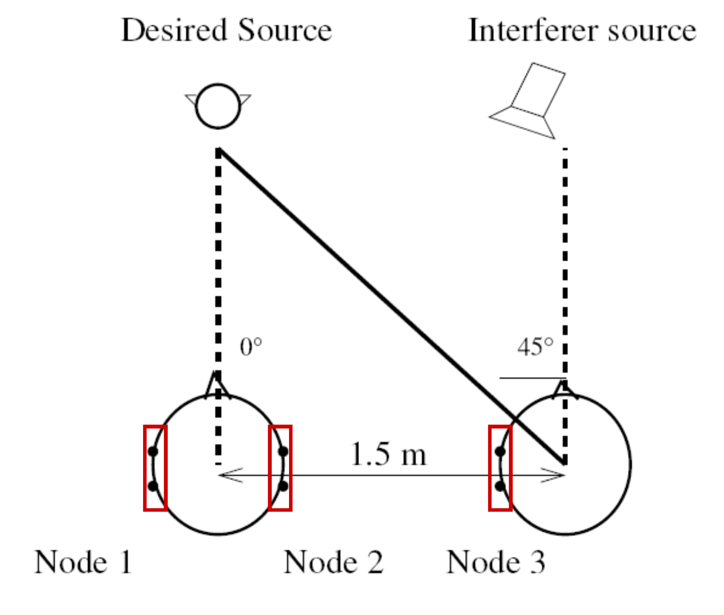


$$B = \max\left(0, \frac{\Phi_Y^{01} - \lambda}{\Phi_Y^{01}}\right)$$

$$\Phi_W = \max\left(0, \lambda \frac{\Phi_Y^{01} - \lambda}{\Phi_Y^{01}}\right)$$

Rate constraints

- Investigate effect of rate constraints on performance of centralized MWF and distributed MWF (DANSE)
- Setup and performance measures:
  - **Acoustic scenario**: 3 nodes, 2 microphones per node (d=1cm)
  - single speech source, single interference, spatially uncorrelated noise on each microphone



$$\Phi_y = \Phi_s \mathbf{A}_s \mathbf{A}_s^H + \Phi_i \mathbf{A}_i \mathbf{A}_i^H + \Phi_u \mathbf{I}_{2M}$$

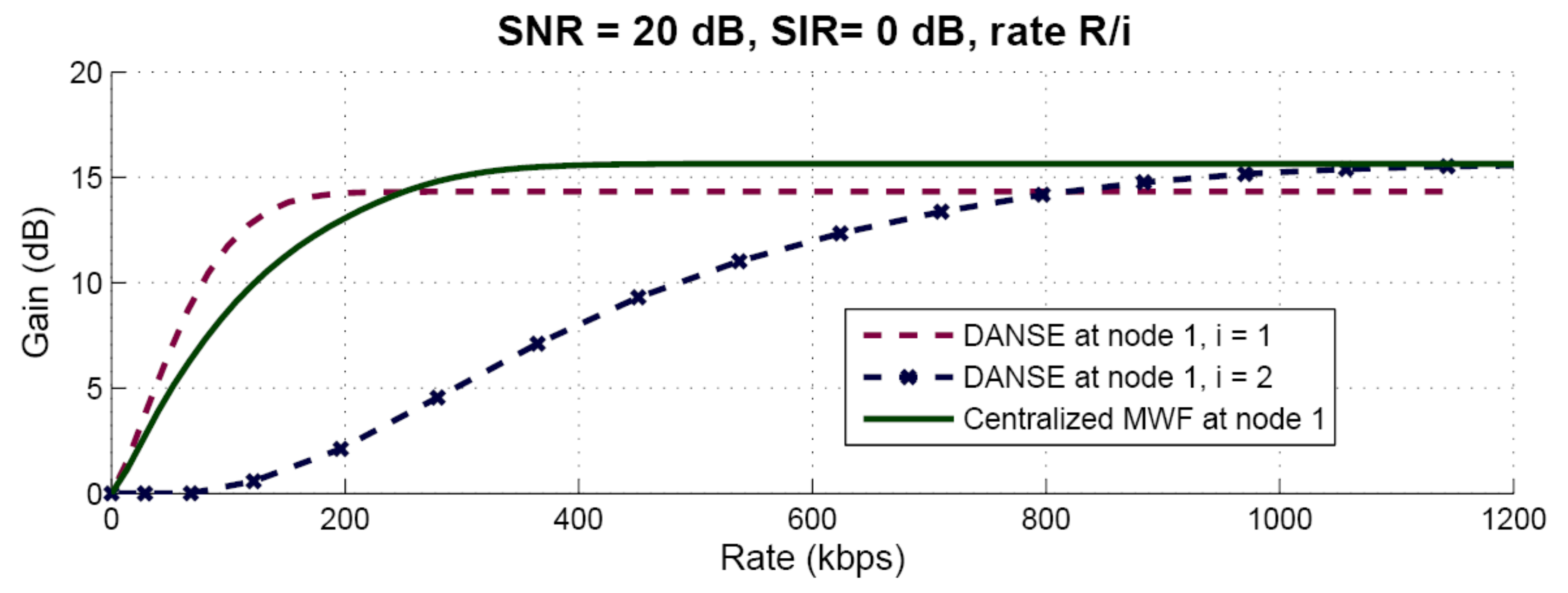
  - – Involved PSDs are assumed to be flat, SIR=0 dB, SNR=20 dB
  - – ATFs modelled using spherical head shadow model, no reverberation
- **Performance measure**: ratio between MSE at rate 0 and MSE at rate *R, i.e.* effect of availability of wireless link

$$G(R) = 10\log_{10}\frac{\xi(0)}{\xi(R)}$$

[T. C. Lawin-Ore, S. Doclo, ICASSP 2011]

Rate constraints

- Effect on performance of distributed MWF (DANSE)
  - **Case 1:** total capacity *R* evenly distributed between iterations

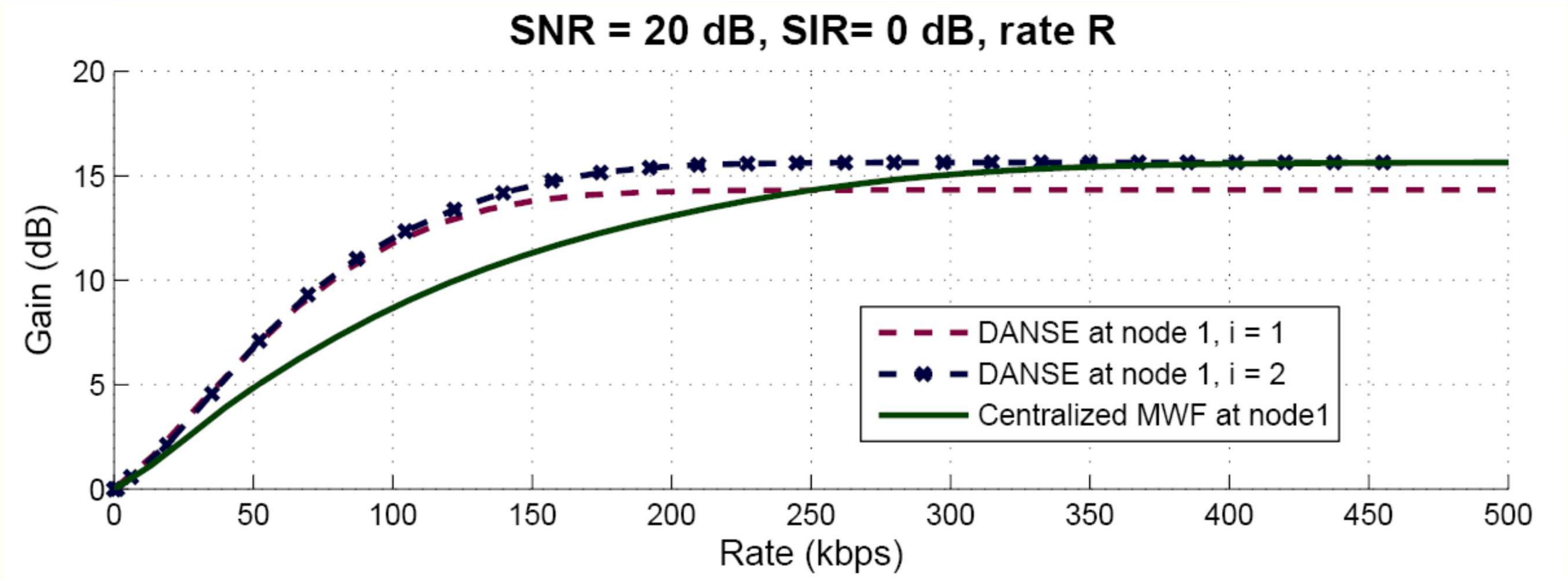


- For **infinite rate**, DANSE converges to centralized MWF
- At **low rates** highest performance gain is achieved by transmitting just a single microphone signal (i = 1).
- **More iterations** only improve performance at high rates

[T. C. Lawin-Ore, S. Doclo, ICASSP 2011]

Rate constraints

- Effect on performance of distributed MWF (DANSE)
  - o **Case 2:** spread iterations over subsequent frames (stationarity)



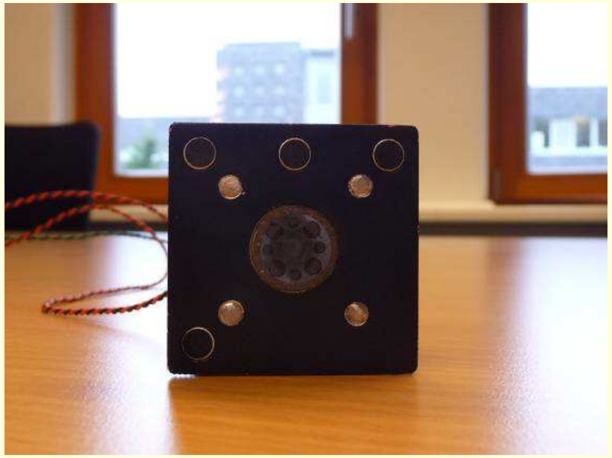
➢ **DANSE scheme** converges after i=2 iterations, moreover achieving **highest performance gain**

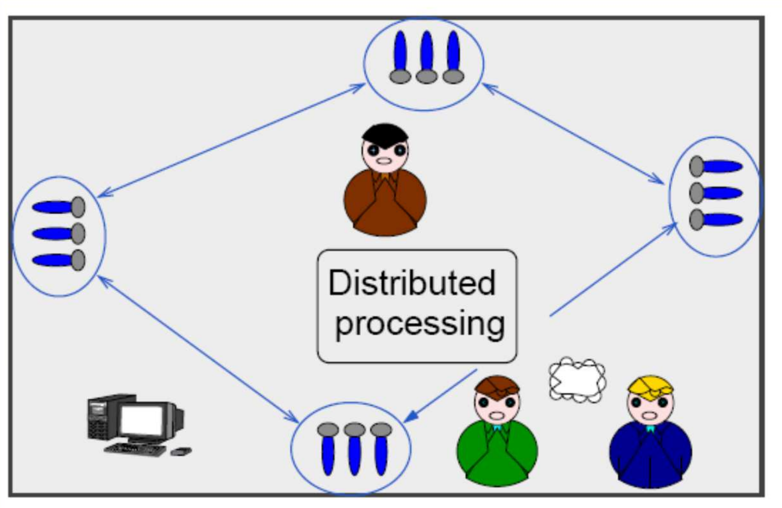
[T. C. Lawin-Ore, S. Doclo, ICASSP 2011]

- Introduction
- Binaural processing
- Acoustic sensor networks
  -Rate constraints
  -Future work
- Conclusion

Acoustic sensor networks

- **Future work/challenges:**
  - *Speech enhancement algorithms:*
    - Dynamic subset selection for time-varying situations
    - Theoretical performance analysis (statistical room acoustics) → optimal microphone configuration
  - *Computational auditory scene analysis:*
    - E.g. multi-source localisation by merging energy- and correlation-based techniques
  - *Calibration and synchronisation techniques:*
    - With and without reference signals
  - (Perceptual) *coding* of transmitted signals





Conclusions

- Speech enhancement algorithms in hearing instruments
  - More and more microphones: monaural → binaural → acoustic sensor networks
  - Algorithms: beamforming, post-processing, MWF
- Bandwidth reduction by transmitting filtered combination of microphone signals
  - D-MWF: iterative procedure, converging to centralized MWF
- Effect of bit-rate on performance using rate-distortion theory
  - D-MWF achieves highest performance gain, when iterations can be spread over subsequent frames
- Remaining challenges in acoustic sensor networks:
  - **Algorithms**: robustness, dynamic subset selection, distributed algorithms, optimal microphone configuration
  - **(Perceptual) coding** of transmitted signal
  - **Technical issues of wireless link:** latency, synchronisation

Acknowledgements

- Daniel Marquardt, Universität Oldenburg
- Toby Christian Lawin-Ore, Universität Oldenburg
- Dr. Thomas Rohdenburg, Fraunhofer HSA, Oldenburg
- Prof. Dr. Joerg Bitzer, Jade Hochschule, Oldenburg
- Christian Bartsch, Jade Hochschule, Oldenburg

- Dr. Tim Van den Bogaert, KU Leuven, Belgium
- Prof. Dr. Jan Wouters, KU Leuven, Belgium
- Prof. Dr. Marc Moonen, KU Leuven, Belgium

Questions ?

House of Hearing, Oldenburg