Signal processing algorithms for wireless connected hearing devices

Prof. Dr. Simon Doclo

Signal Processing Group, University of Oldenburg, Germany

simon.doclo@uni-oldenburg.de

http://www.sigproc.uni-oldenburg.de

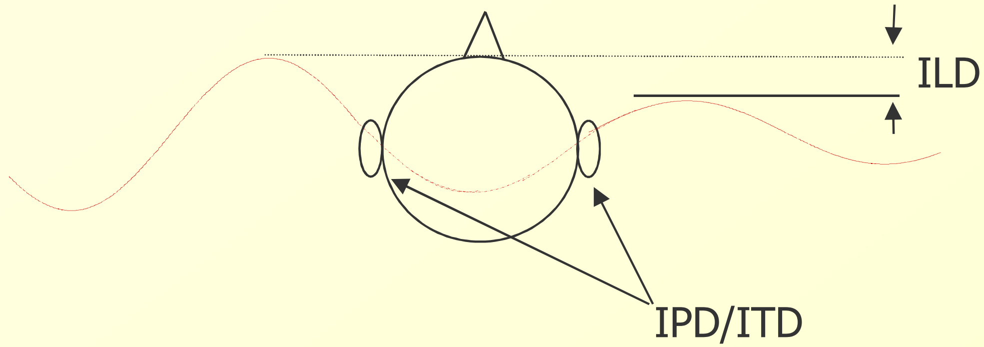
Nordic Audiology College, 21.09.2012

Outline

- Use of multiple microphones in hearing aids
  - Monaural $\rightarrow$ Binaural $\rightarrow$ 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
  - o signal processing to selectively enhance useful speech signal and improve speech intelligibility
  - o signal processing to preserve directional hearing (binaural auditory cues) and spatial awareness
  - o robustness important due to small inter-microphone distance
- Binaural auditory cues
  - o Interaural Time Difference (ITD) – Interaural Level Difference (ILD) – Interaural Coherence (IC)
  - o Binaural cues, in addition to spectral and temporal cues, play an important role in binaural noise reduction and sound localisation
  - o ITD: f < 1500Hz, ILD: f > 2000Hz



Hearing aids - Algorithms

- Introduction
- Binaural processing
- Acoustic sensor networks
- Conclusion

- Cochlear loss:
  - Frequency-specific amplification
  - Dynamic range compression
- Binaural and central loss:
  - Noise reduction
  - Binaural Algorithms
- "Technical" requirements
  - Feedback control
  - Occlusion effect / 'own voice' detection
  - Classification of acoustic environment
  - (fully digital, 1V supply from very small battery, 5-6d battery time, wireless binaural link)

Hearing aids - Algorithms

- Introduction
- Binaural processing
- Acoustic sensor networks
- Conclusion

- Digital hearing instruments allow for advanced acoustical signal pre-processing
  - multiple microphones available $\rightarrow$ spectral + spatial processing
  - noise reduction (beamforming), computational auditory scene analysis (source localisation, environment classification, …)

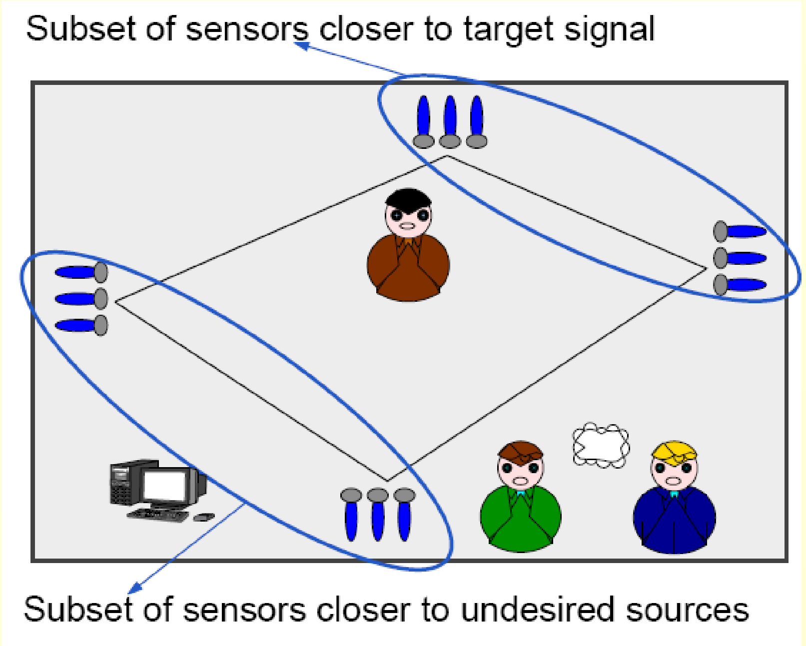
Monaural (2-3)

Binaural

External microphones

Acoustic sensor networks

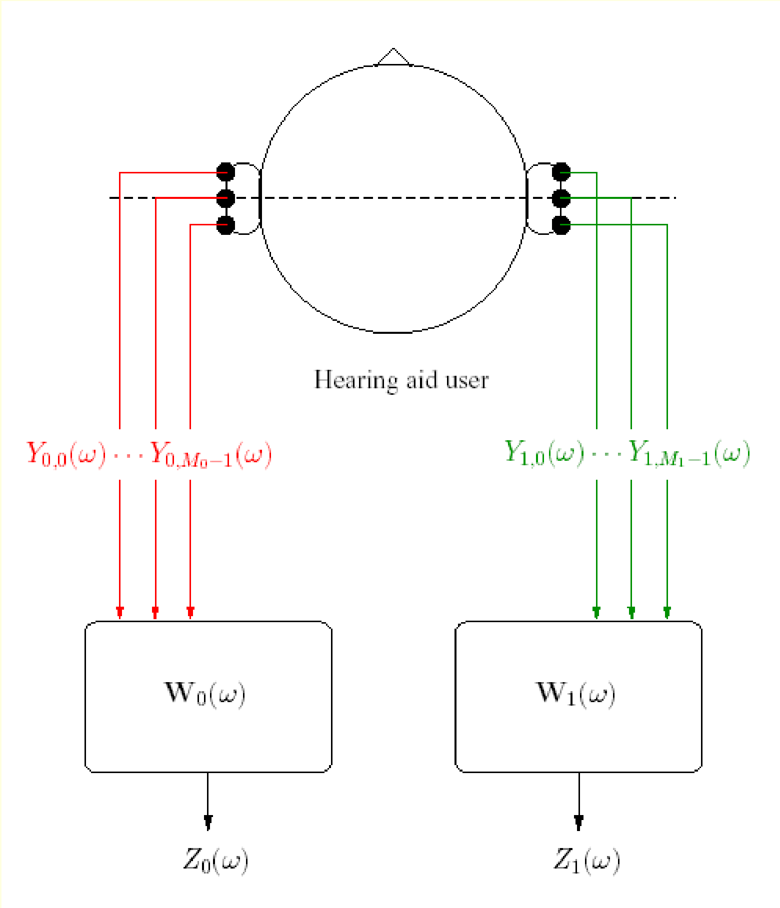
- Signal acquisition in **adverse acoustic environments**:
  - Microphones at large distance from speaker $\rightarrow$ background noise and reverberation
- **Acoustic sensor networks**:
  - Network of a large number of spatially distributed nodes (each with one or multiple microphones)
  - Wireless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



Binaural processing

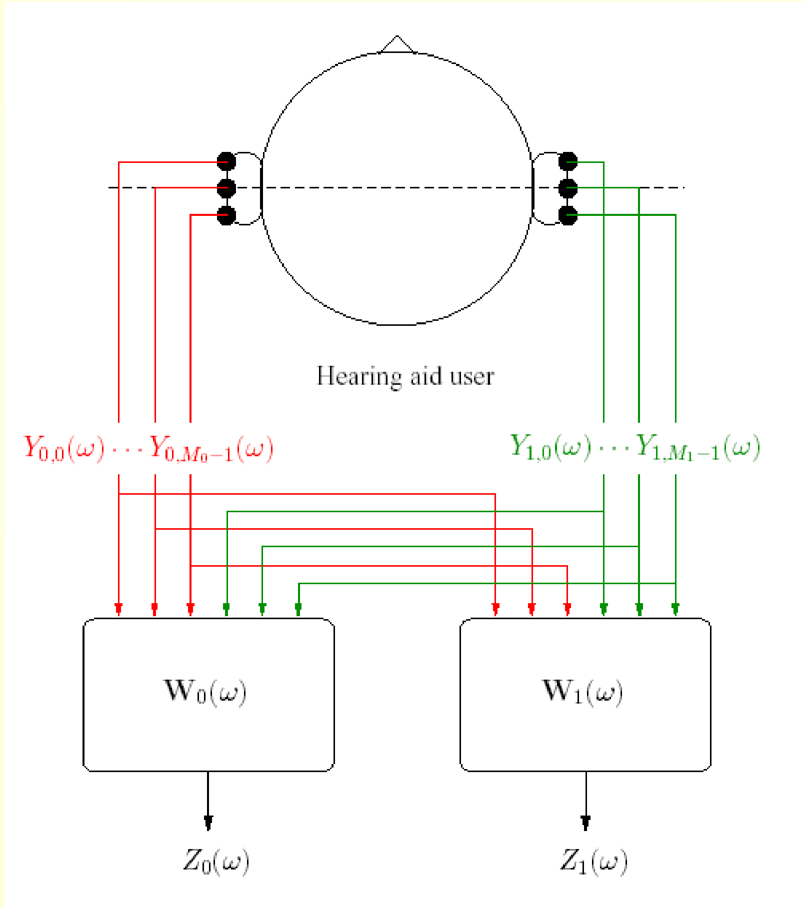
Bilateral vs. Binaural

Bilateral system



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

Binaural system



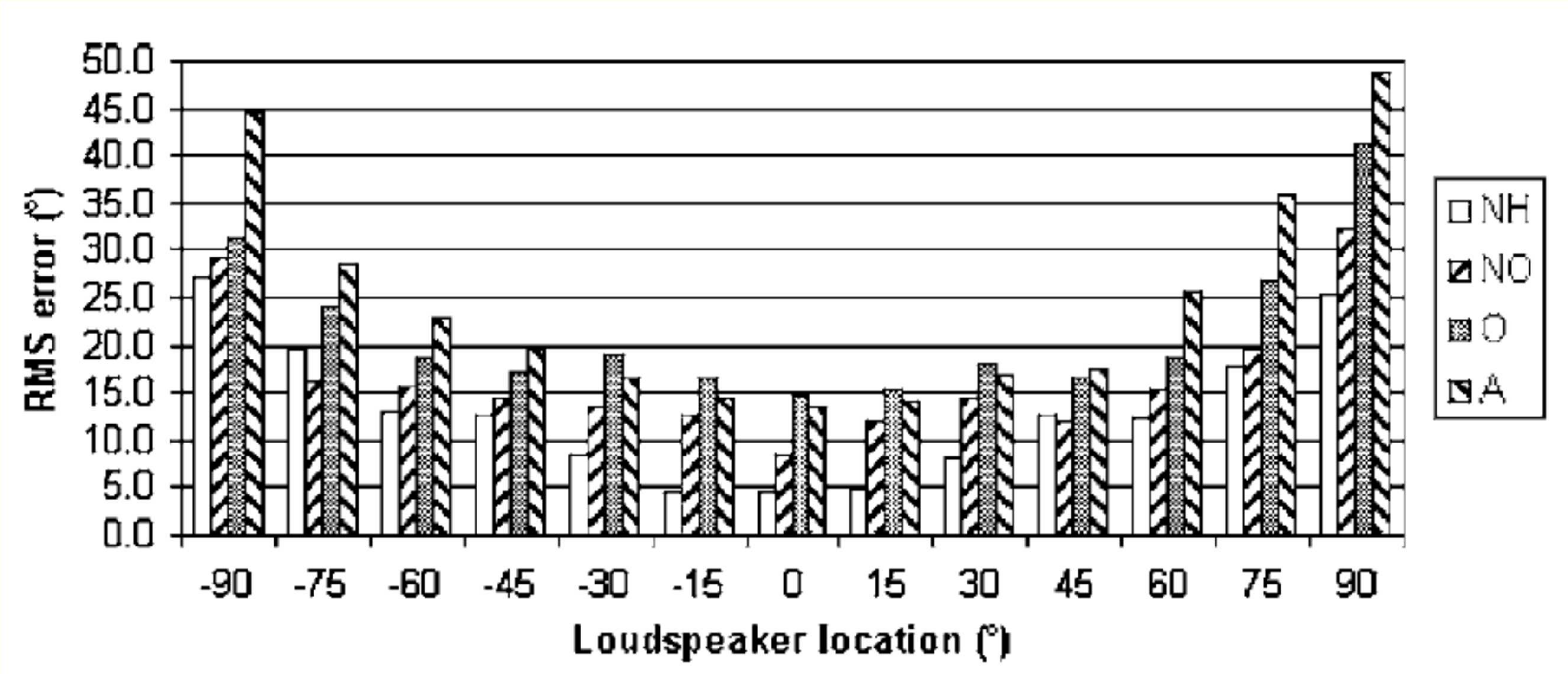
⊕ More microphones:
$\rightarrow$ better performance ?
$\rightarrow$ preservation of binaural cues ?

⊖ Need for wireless binaural link

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

Bilateral vs. Binaural

- Bilateral system:
  - o Independent processing of left and right hearing aid
  - o Negative effect on localisation cues and intelligibility through binaural hearing advantage [Van den Bogaert, 2006; Keidser, 2009]



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 microphone)

Bilateral vs. Binaural

- Bilateral system:
  - o Independent processing of left and right hearing aid
  - o Negative effect on localisation cues and intelligibility through binaural hearing advantage [Van den Bogaert, 2006; Keidser, 2009]

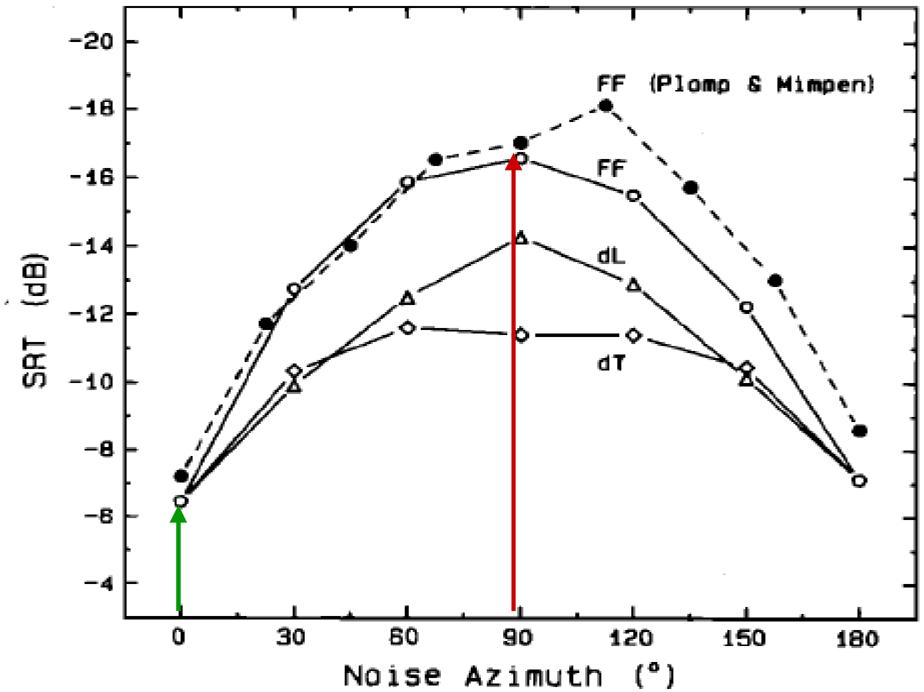
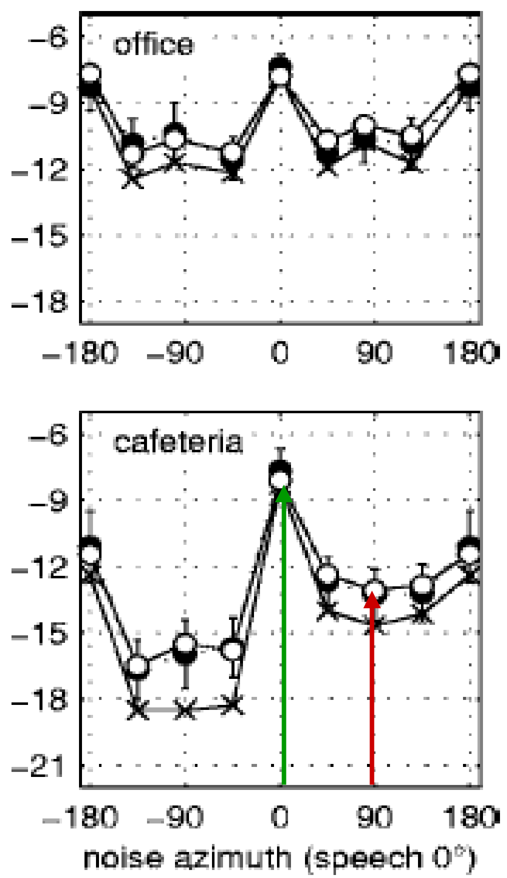


FIG. 5. Mean speech reception thresholds obtained in experiment I for three different noise types : FF (free field), dL (headshadow only), and dT (ITD only). The closed data points represent results of Plomp and Mimpen (1981) obtained in a free field.



[Bronkhorst and Plomp, 1988]

[Beutelmann and Brand, 2006]

Hearing 4all

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, 2006; Keidser, 2009]
- 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

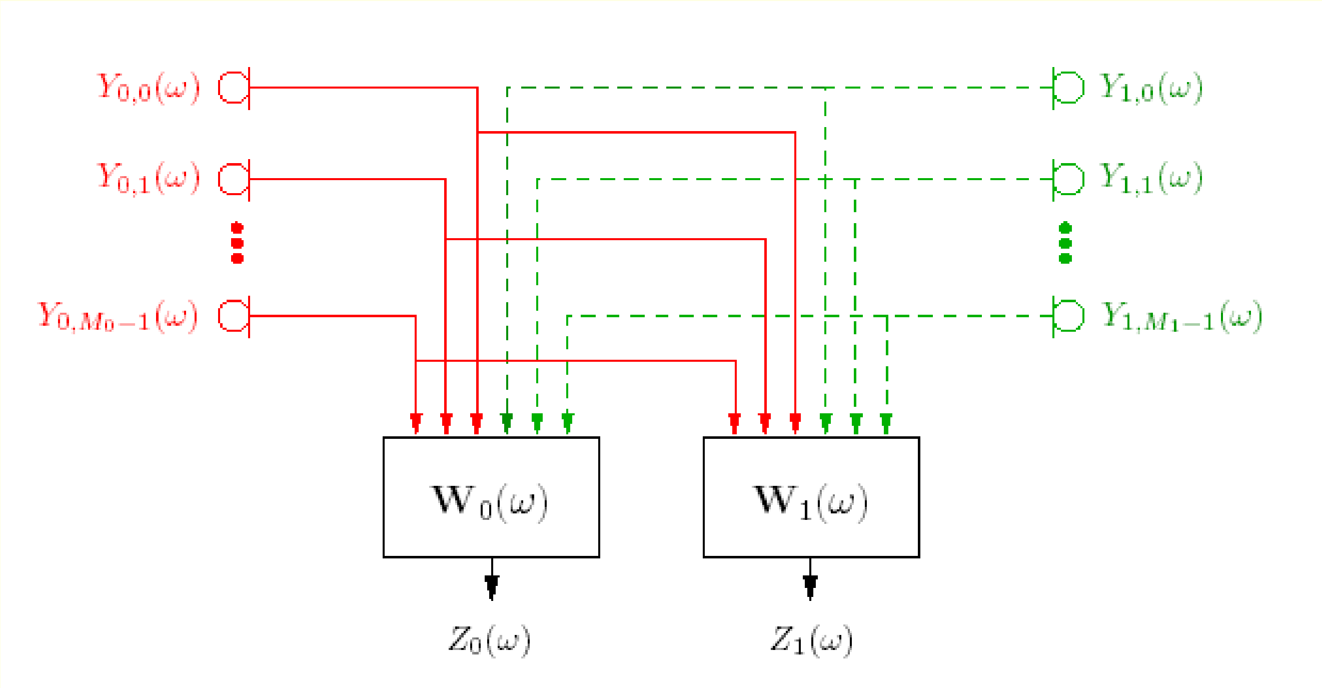
Binaural noise reduction techniques

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

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

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

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

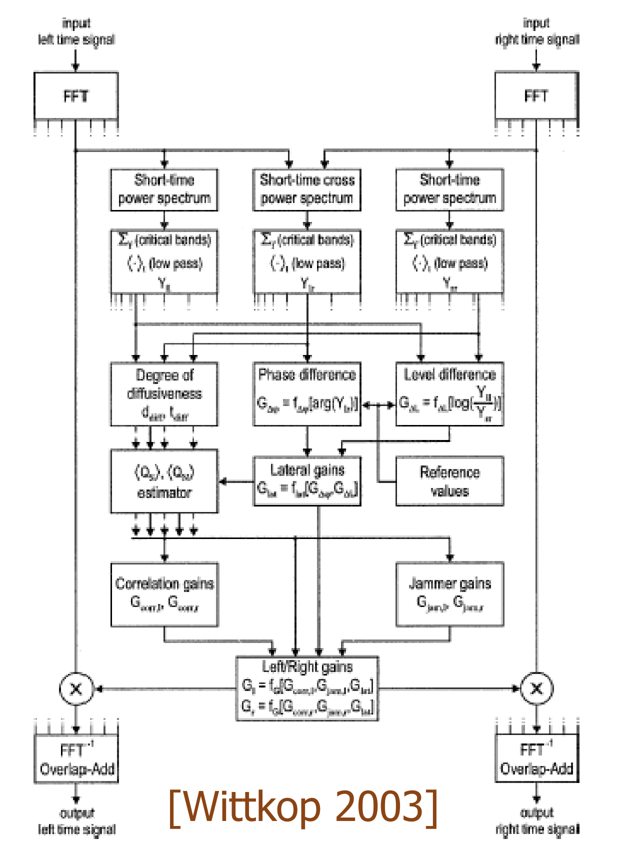


Binaural noise reduction techniques

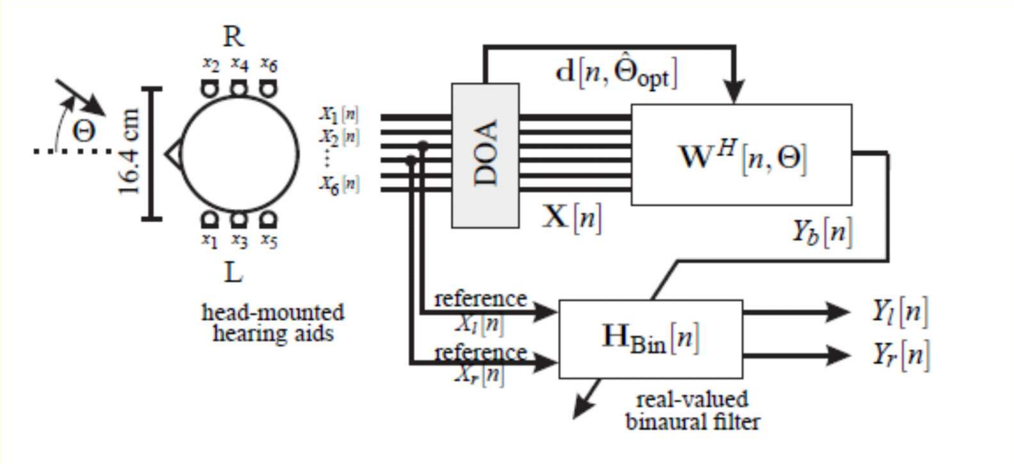
- Time-frequency post-processing/masking:
  - o Computation and application of **real-valued** binaural mask/weight based on binaural and temporal/spectral cues
  - o Can be merged with MVDR-beamformer or ICA-based processing

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

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



[Wittkop 2003]



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

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

[Rohdenburg 2009, Reindl 2010, Saruwatari 2010]

Hearing 4all

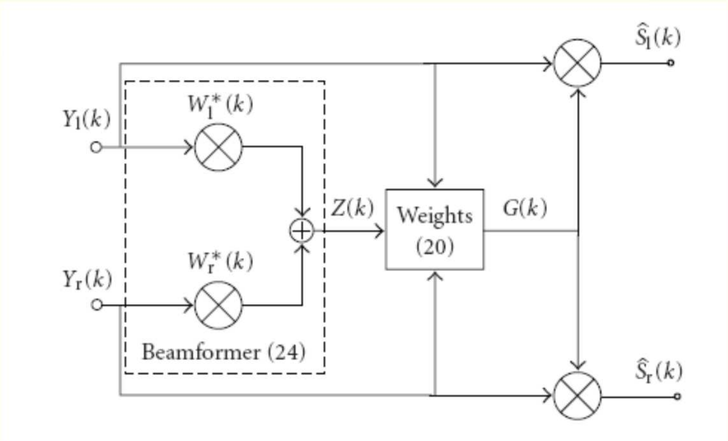
- Introduction
- Binaural processing
  - -Algorithms
  - -Experiments
  - -Distributed MWF
- Acoustic sensor networks
- Conclusion

Binaural noise reduction techniques

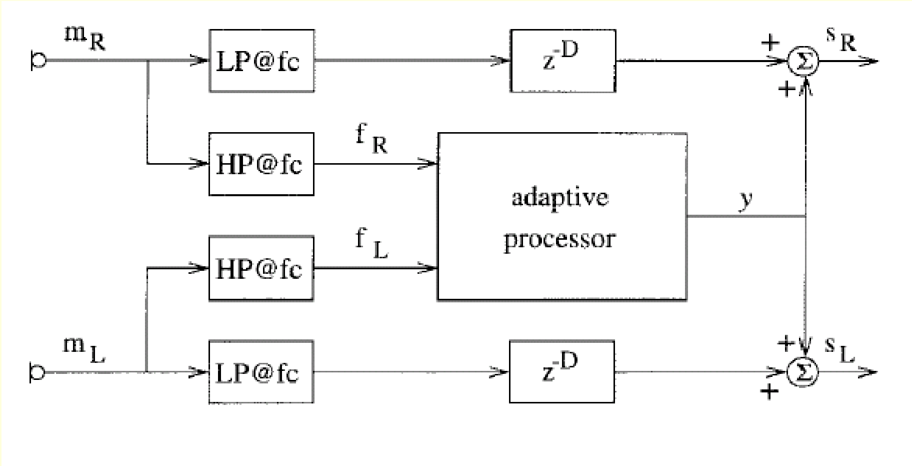
- Beamforming: spatial selectivity + binaural speech cues
  - o Maximize directivity index while restricting speech ITD error [Desloge, 1997]
  - o Superdirective beamformer using HRTFs [Lotter, 2004]
  - o Adaptive beamforming based on GSC [Welker, 1997]

⊕ low computational complexity

⊖ limited performance, known geometry, only speech cues may be preserved (in ideal situations)



[Lotter, 2006]



[Welker, 1997]

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

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

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

speech component in front microphones

speech distortion

noise reduction

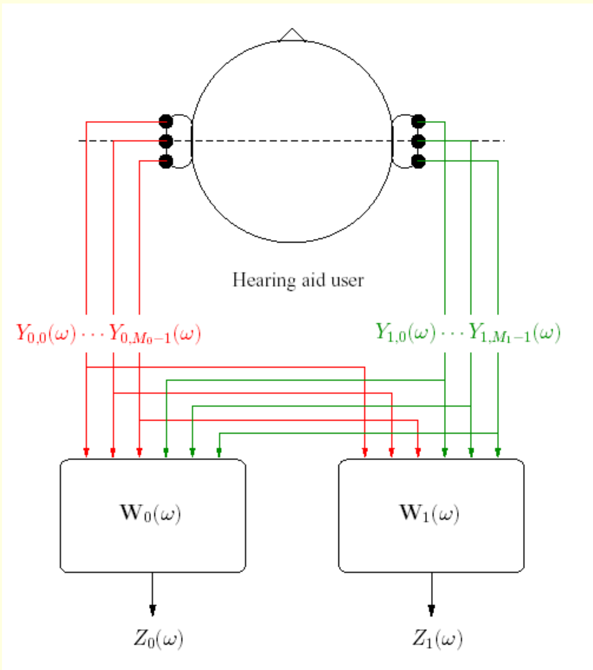
estimate

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

- o 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
- o No assumptions about positions of microphones and sources
- o Different implementations:
  - – Batch (off-line) vs. adaptive (update correlation matrices)
  - – Using spatial prediction (SP) between speech components

Binaural noise reduction techniques

- Binaural multi-channel Wiener filter:
  - o Preservation of binaural cues (ITD-ILD)
    - – Speech cues are preserved, no a-priori assumptions
    - – **Noise cues are distorted**
  - o **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) or Interaural Coherence (MWF-IC) of noise source



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

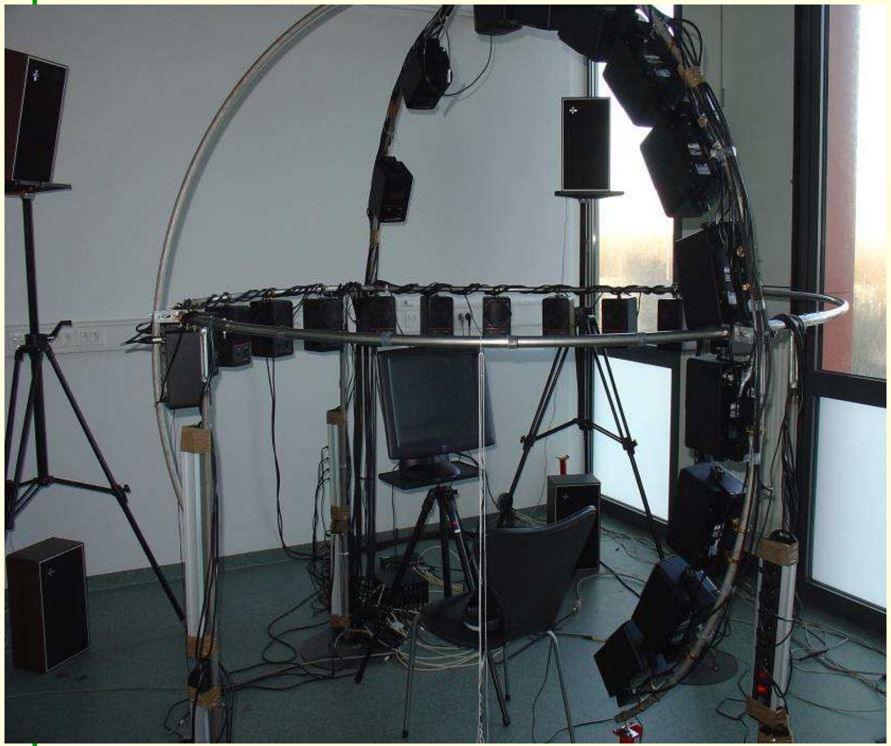
$$J_{MWF-ITF} = J_{MWF} + \delta \mathcal{E}\left\{ \left| \mathbf{W}_0^H \mathbf{V} - ITF_v^{des} \mathbf{W}_1^H \mathbf{V} \right|^2 \right\}$$

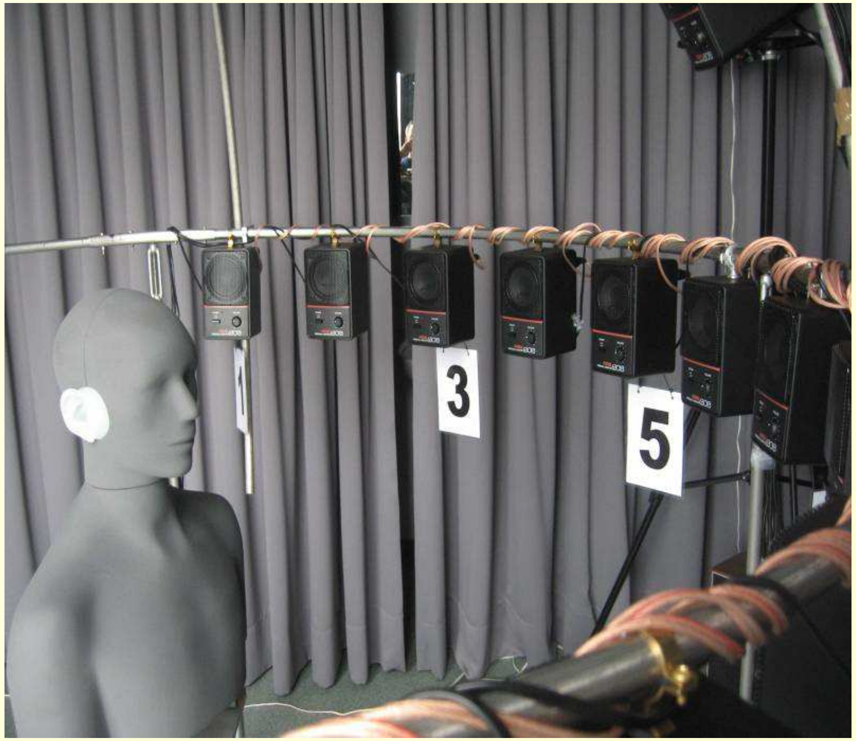
$$J_{MWF-IC} = J_{MWF} + \delta \left| \frac{\mathcal{E}\left\{ \mathbf{W}_0^H \mathbf{V} \mathbf{V}^H \mathbf{W}_1^H \right\}}{\sqrt{\mathcal{E}\left\{ \left| \mathbf{W}_0^H \mathbf{V} \right|^2 \right\} \mathcal{E}\left\{ \left| \mathbf{W}_1^H \mathbf{V} \right|^2 \right\}}} - IC_v^{des} \right|^2$$

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

Experimental results (1)

- 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}=510 ms, f_s=16 kHz





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

Audio demo

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

	Noisy	Speech	Noise
Input			
Output (β=0)			
Output (β=0.05)			

Hearing 4all

Perceptual evaluation: SRT

- SRT measurements (headphone presentation, 10 NH)
- Speech sentences + babble noise
- Algorithms: state-of-the art bilateral (ADM), MWF, MWFv (η=0.2)
- Conditions: S_0N_{60}, $S_{90}N_{270}$ and $S_0N_{90/180/270}$

	S_0N_{60}			$S_{90}N_{270}$			$S_0N_{90/180/270}$		
Bilat/bin ΔSRT (dB)	Perceptual	Left	Right	Perceptual	Left	Right	Perceptual	Left	Right
ADM	$\boxed{2.1 \pm 1.9}$	2.7	2.8	$\boxed{-4.3 \pm 1.3^*}$	4.3	−3.2	$\boxed{1.3 \pm 1.4}$	6.0	5.9
MWF_{2+2}	$\boxed{4.3 \pm 1.5^*}$	4.9	9.6	$\boxed{0.7 \pm 1.4}$	10.0	2.5	$\boxed{4.6 \pm 0.8^*}$	7.1	7.2
MWF_{2+1}	$3.8 \pm 1.6^*$	4.0	6.2	0.3 ± 2.0	9.6	2.1	$4.0 \pm 1.5^*$	6.6	6.0
MWF_{2+0}	$1.0 \pm 0.7^*$	1.9	3.3	-1.2 ± 1.6	3.8	1.0	$2.8 \pm 1.3^*$	5.1	4.9
MWF_{2+2}-$N_{0.2}$	$\boxed{3.6 \pm 1.4^*}$	3.3	5.4	$\boxed{2.0 \pm 1.4^*}$	4.3	1.9	$\boxed{3.2 \pm 0.8^*}$	4.1	4.2
MWF_{2+1}-$N_{0.2}$	$2.7 \pm 1.3^*$	2.6	3.0	1.5 ± 1.6	3.9	1.6	$3.4 \pm 0.8^*$	3.7	3.3
MWF_{2+0}-$N_{0.2}$	1.0 ± 2.1	1.1	0.9	0.0 ± 1.5	1.0	0.7	$2.3 \pm 1.4^*$	2.8	2.6

[T. Van den Bogaert, S. Doclo, J. Wouters, M. Moonen, Journal of the Acoustical Society of America, vol. 124, no. 1, Jul. 2008.]

Perceptual evaluation: SRT

- SRT measurements (10 NH, 8 HI)

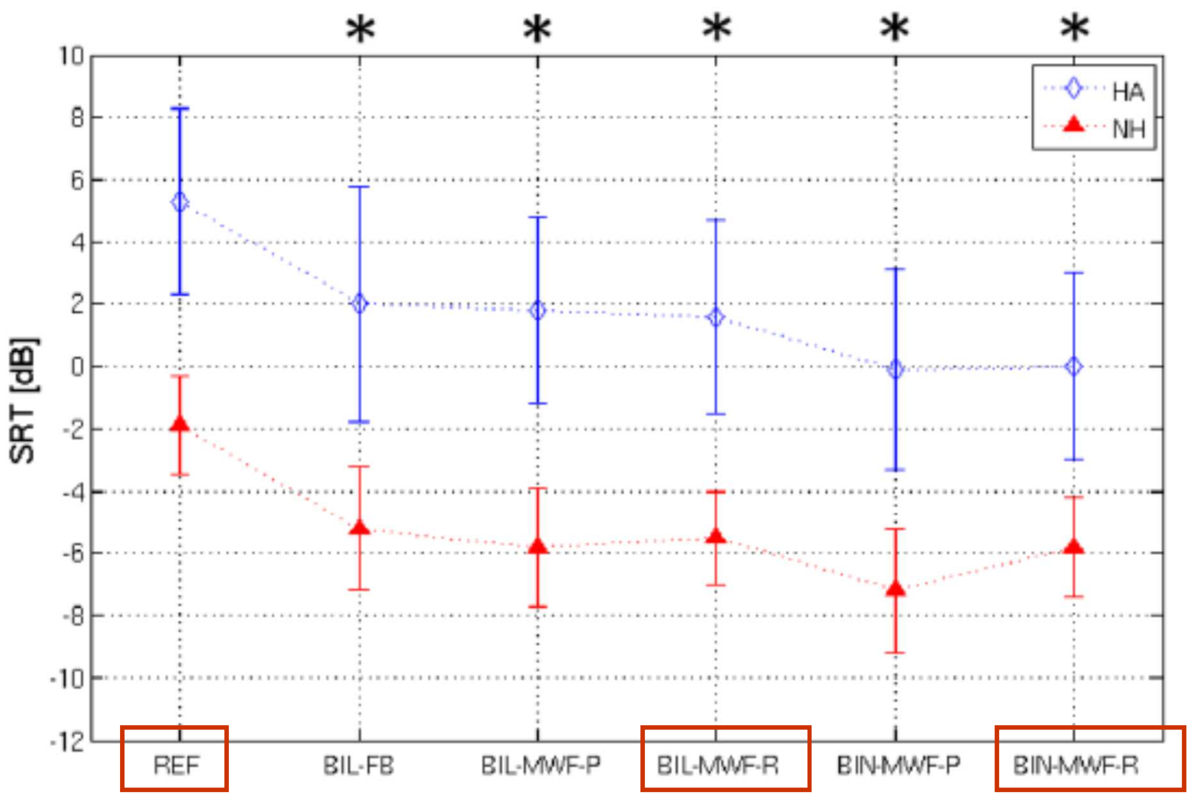


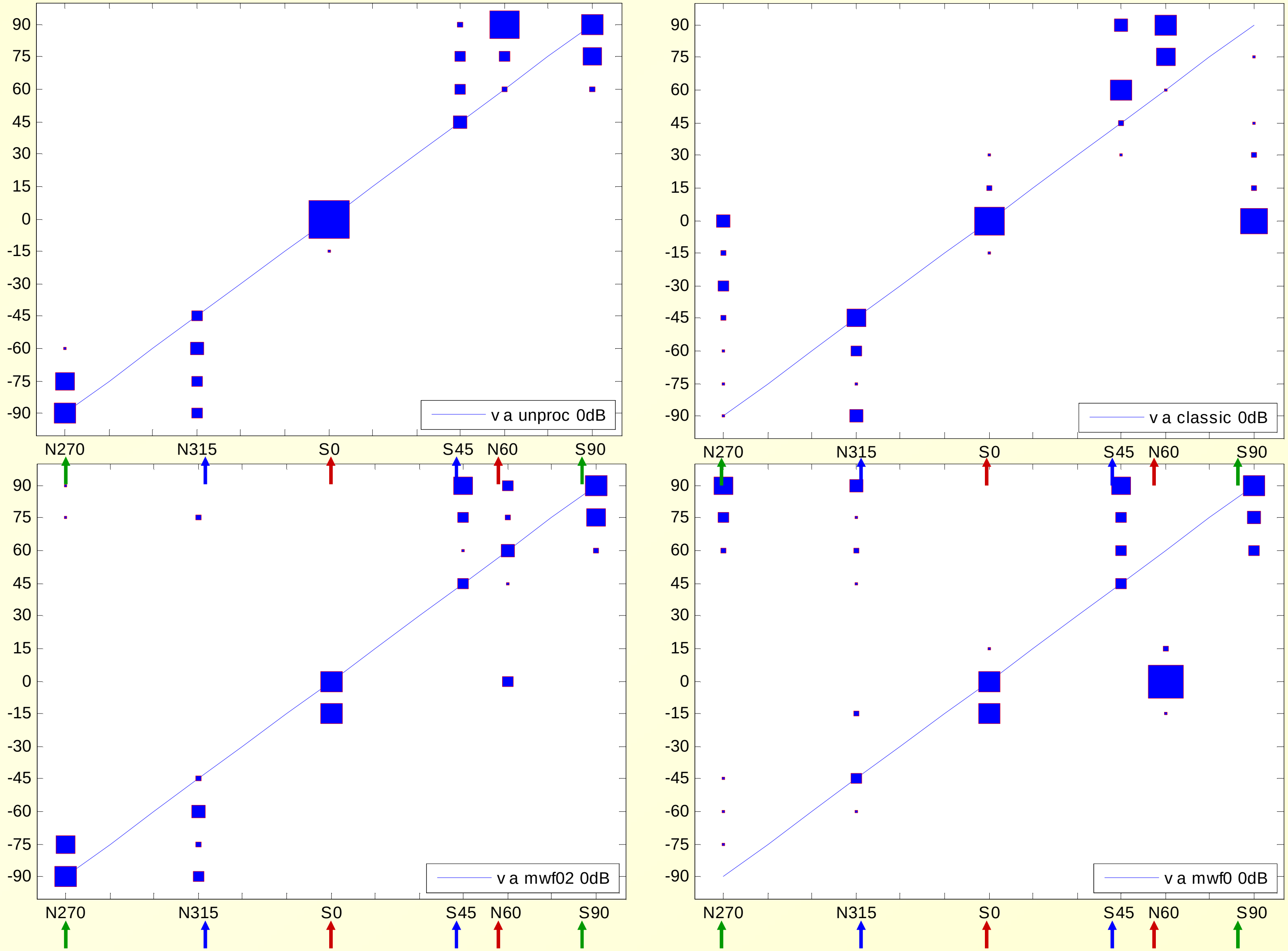
FIG. 7. (Color online) Average SRT results of ten normal hearing subjects (NH) and eight hearing aid users (HA) for the $S_0N_{90/180/270}$ scenario. Standard deviations are indicated by error bars. Algorithms which significantly improve the SRT for both groups, compared to the unprocessed (REF) condition, are marked by an "*" above the graph.

[Cornelis, 2012]

Perceptual evaluation: localisation

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

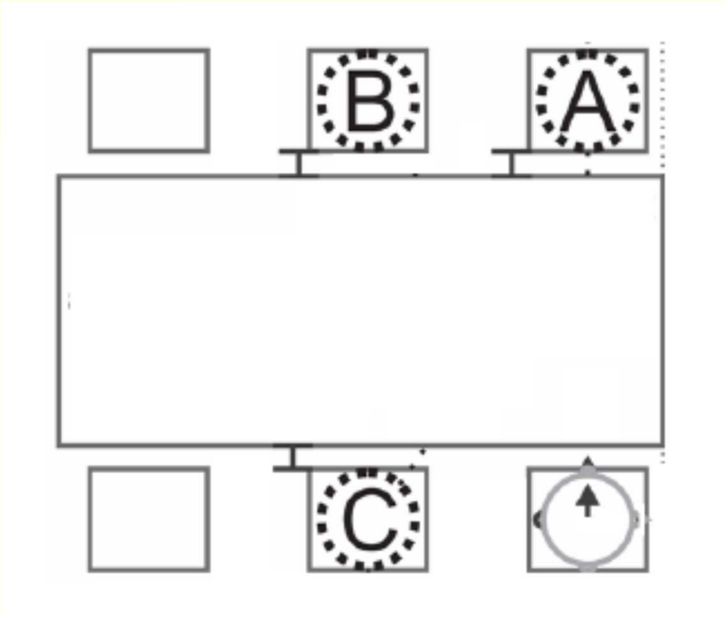
- **Algorithms**: unprocessed, state-of-the art bilateral, MWF, MWFv (η=0.2)
- **Conditions**: S_0N_{60}, $S_{45}N_{315}$ and $S_{90}N_{270}$

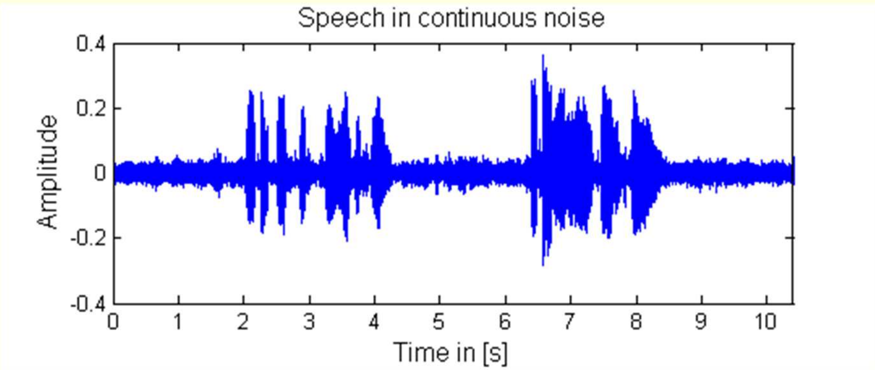


[T. Van den Bogaert, S. Doclo, J. Wouters, M. Moonen, Journal of the Acoustical Society of America, Jan. 2009.]

Experimental results (2)

- **Acoustic environment**





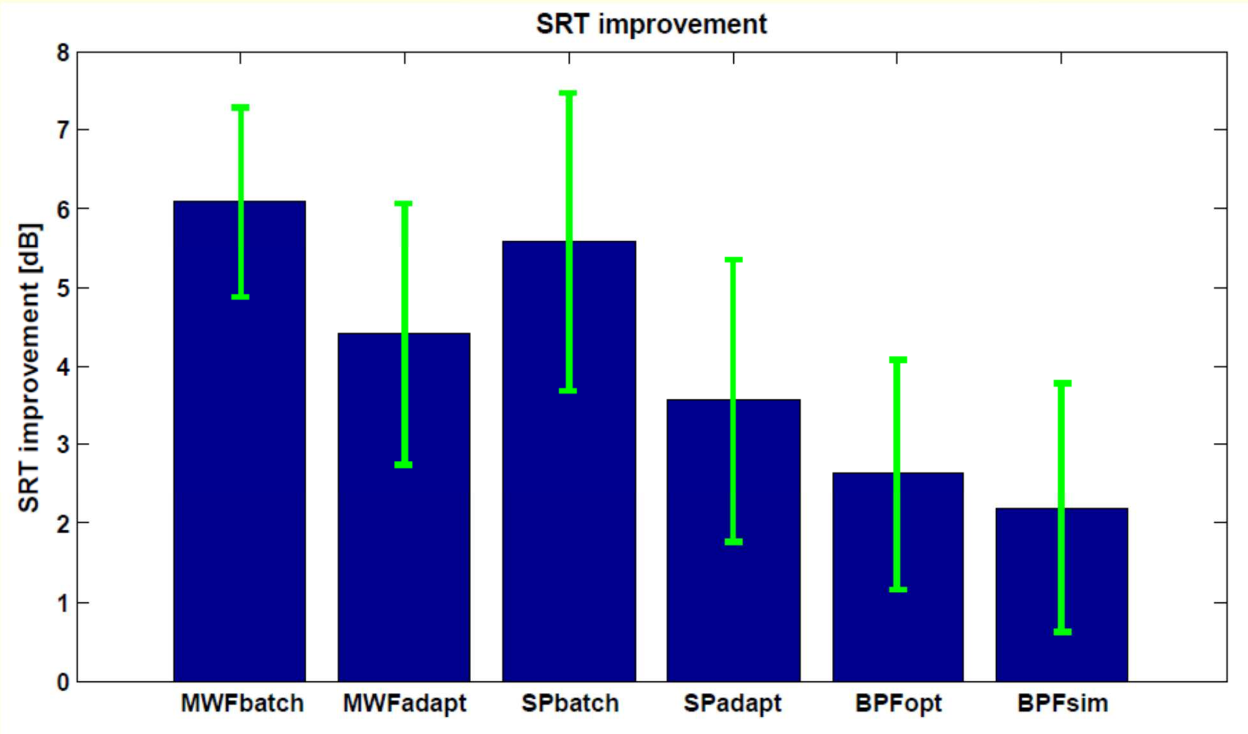
- Cafeteria with recorded babble noise and simulated speaker at position B
- Binaural hearing aid with 3 microphones
- German sentences taken from OLSA speech material
- Speech in continuous babble noise
- f_s: 16 kHz, WOLA, FFT-size: 256 samples, Overlap: 75%

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

Experimental results

- Intro
- **Binaural processing**
  - -Algorithms
  - -Experiments
  - -Distributed MWF
- Acoustic sensor networks
- Conclusion

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



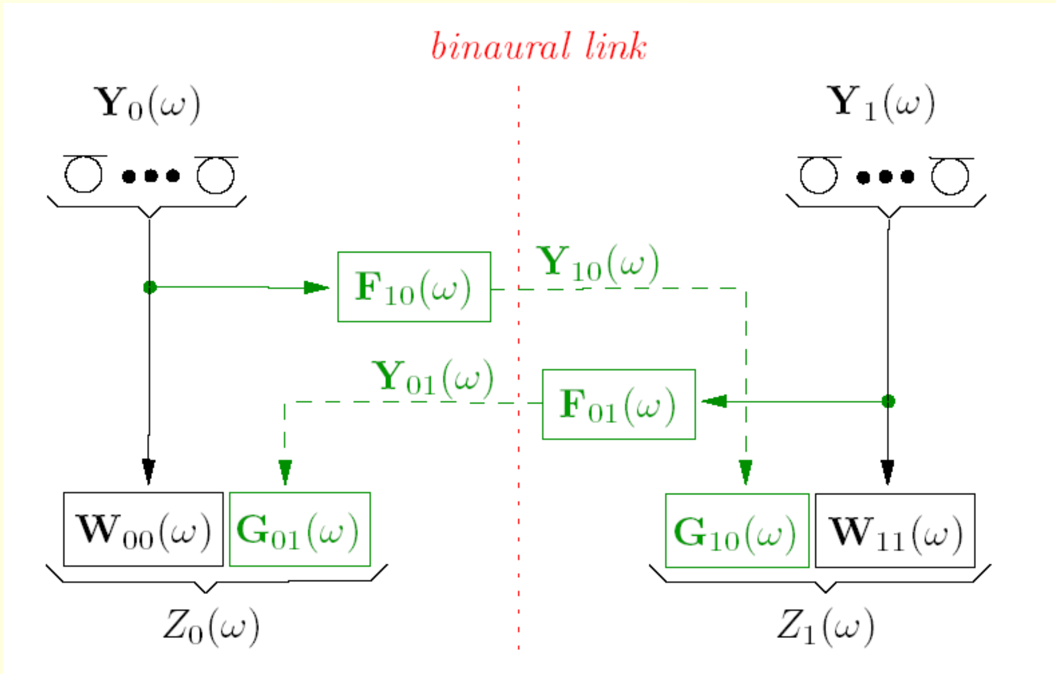
SNR	Orig.	MWF	SP	BF + Postfilt
0 dB				

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

Distributed MWF

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

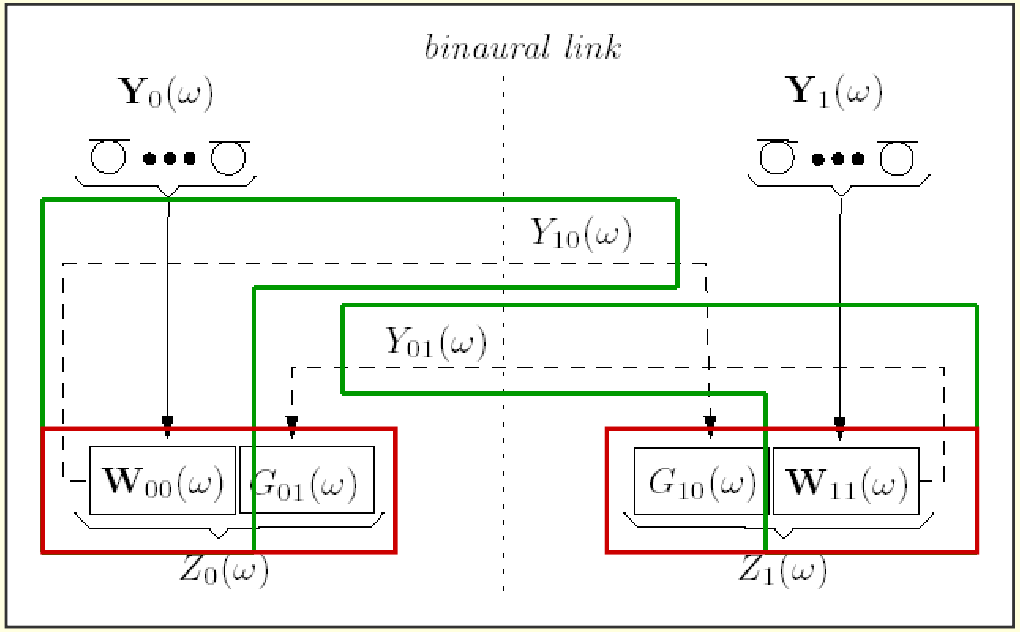
- Binaural MWF
  - o **all** microphone signals are transmitted over wireless link
- Reduce bandwidth requirement of wireless link by transmitting **one** signal from contralateral ear
  - o Raw microphone signal (e.g. front)
  - o Output of fixed (e.g. superdirective) beamformer
  - o MWF-estimate using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

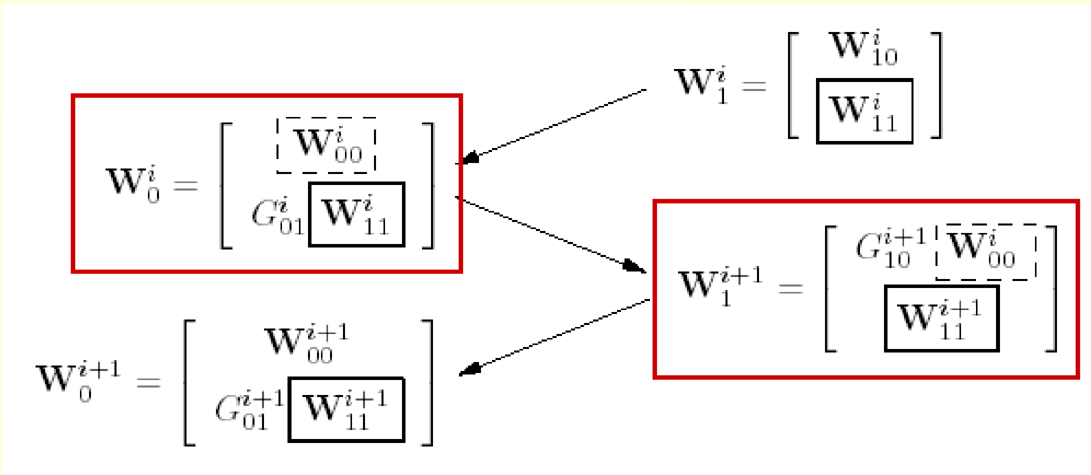


[S. Doclo, T. Van den Bogaert, M. Moonen, J. Wouters, IEEE Trans. Audio, Speech and Language Processing, Jan. 2009.]

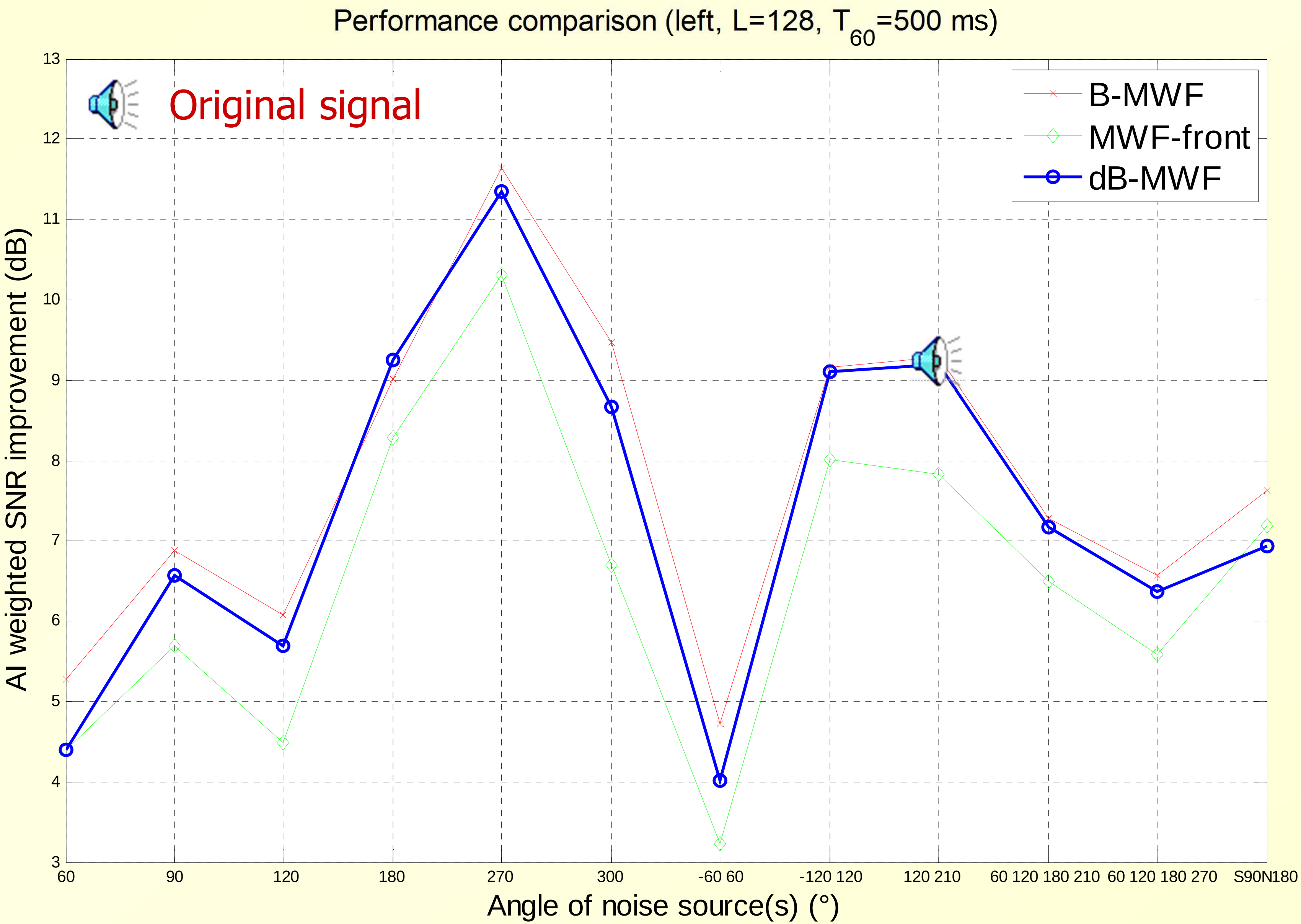
Distributed MWF

- Iterative procedure
  - o 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





Distributed MWF

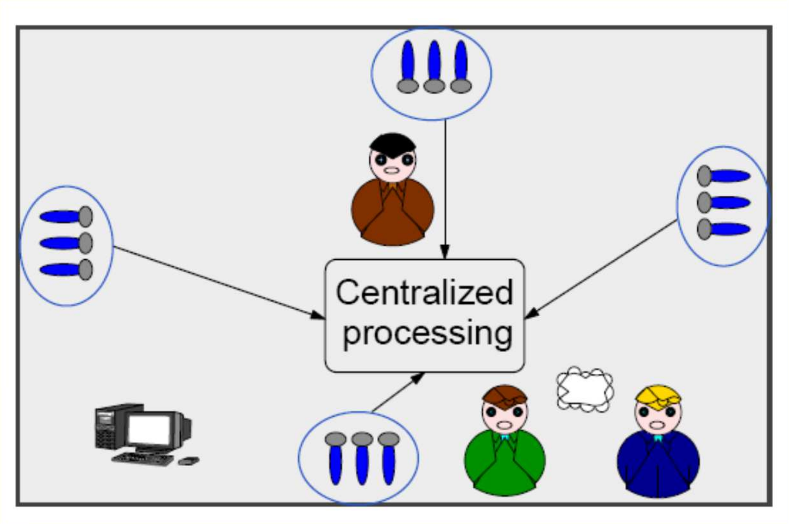


Real-world performance of db-MWF close to full binaural MWF !

Acoustic sensor networks

Acoustic sensor networks

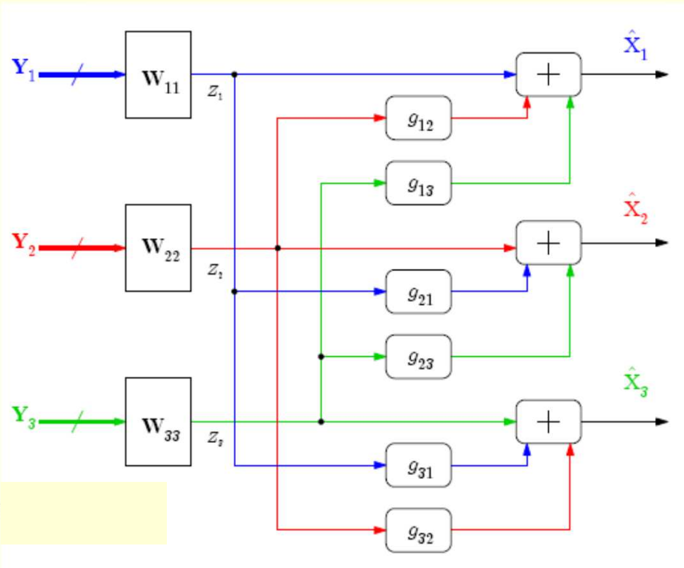
- Now consider more than 2 sensor nodes…
- **Challenges:**
  - o *Dynamic array configuration*: large number of microphones at unknown positions, dynamic subset selection
  - o *Distributed and collaborative algorithms:* power and complexity constraints, effect of limited bandwidth
  - o *Calibration and synchronisation issues:* same time basis
- **Prototype applications:**
  - o Hearing aids using extra microphones (room, other HA, …)
  - o Video-conferencing using all microphones on laptops / room
  - o Surveillance

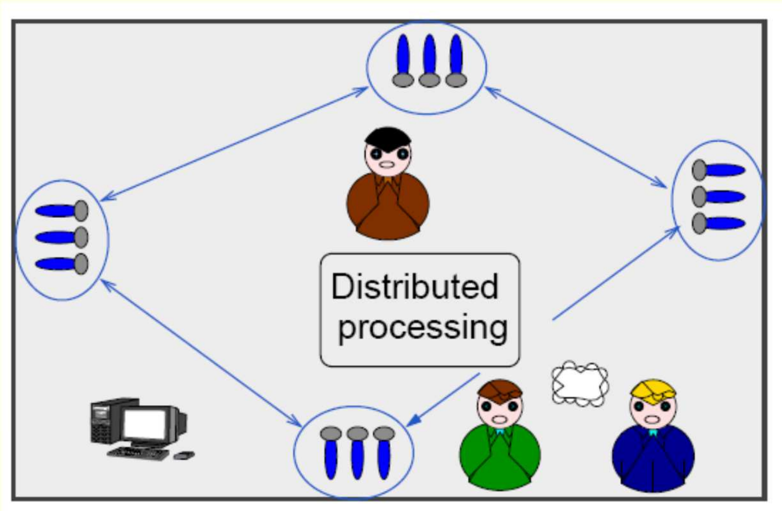


Acoustic sensor networks

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

- Recently has become quite a **hot research topic**
  - o Distributed MWF: extension to multiple sensor arrays and multiple desired sources (DANSE) [Bertrand 2010]
  - o Distributed MVDR/LCMV-beamformer [Golan 2010, Bertrand 2011]
  - o Performance analysis of a randomly spaced wireless microphone array, statistical performance of MWF [Golan 2011, Lawin-Ore 2012]
  - o Dynamic signal combining (no synchronisation required) [Matheja 2011, Srinivasan 2011, Stenger 2011]

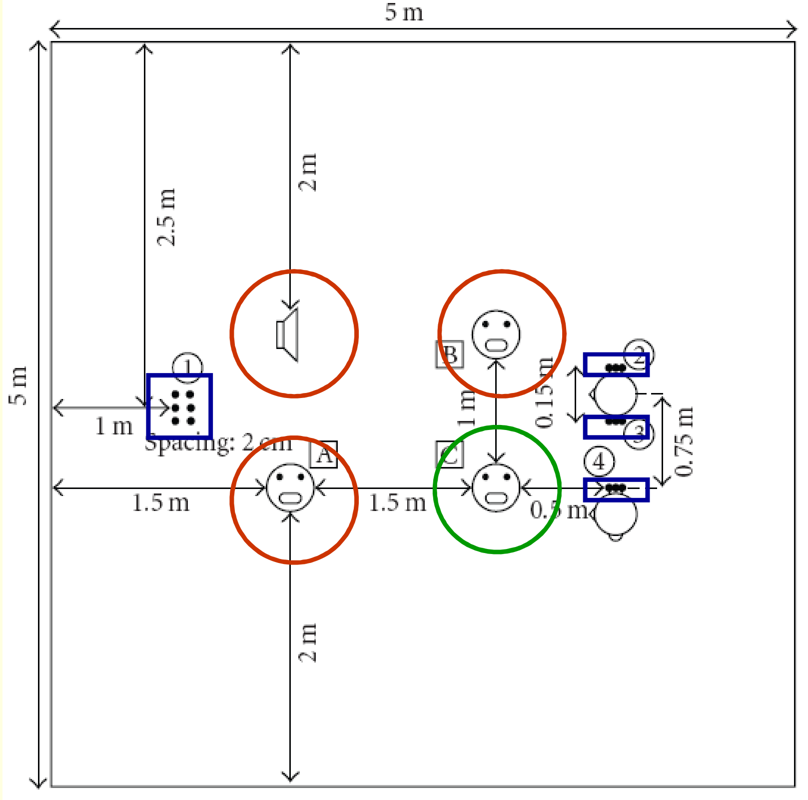


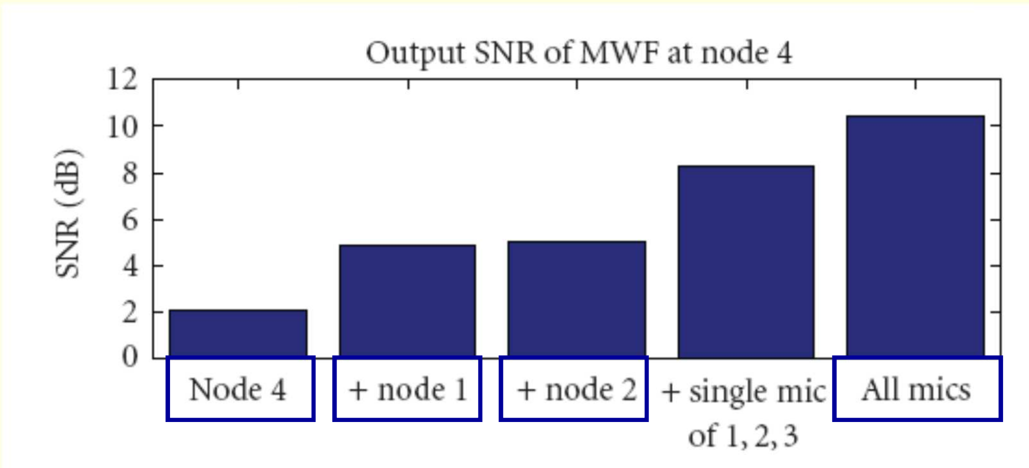


Experimental results

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

T_{60} = 0,22 s

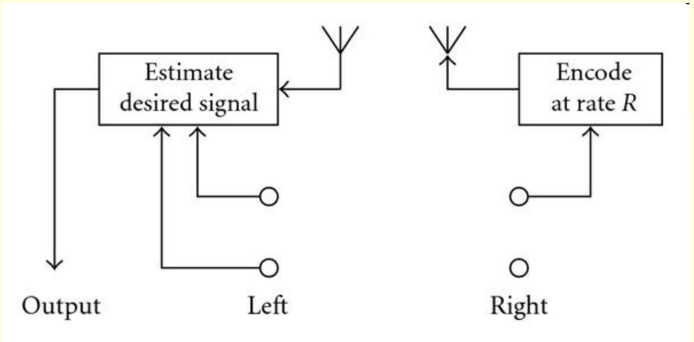


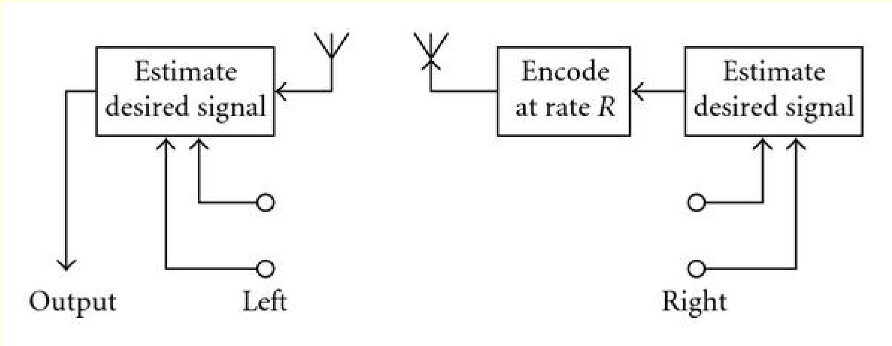


[Bertrand 2009]

Rate constraints

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





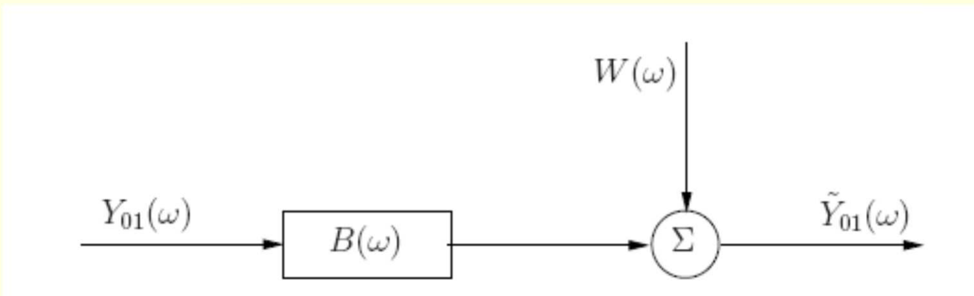
- Rate-distortion:

$$R(\lambda) = \frac{1}{4\pi}\int_{-\infty}^{\infty} \max\left(0, \log_2 \frac{\Phi_Y^{01}(\omega)}{\lambda}\right) d\omega$$

$$D(\lambda) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \min\left(\lambda, \Phi_Y^{01}(\omega)\right) d\omega\,,$$

PSD of transmitted signal

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



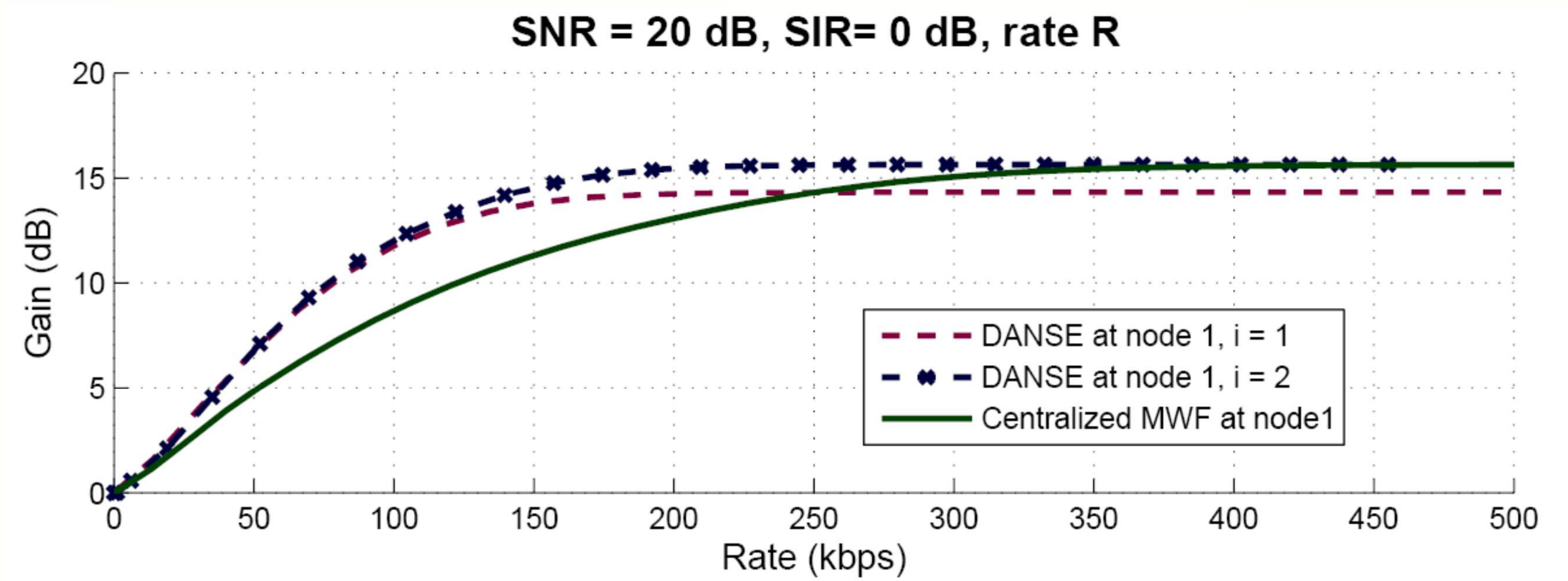
$$B = \max\left(0, \frac{\Phi_Y^{01} - \lambda}{\Phi_Y^{01}}\right)$$

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

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

Rate constraints

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



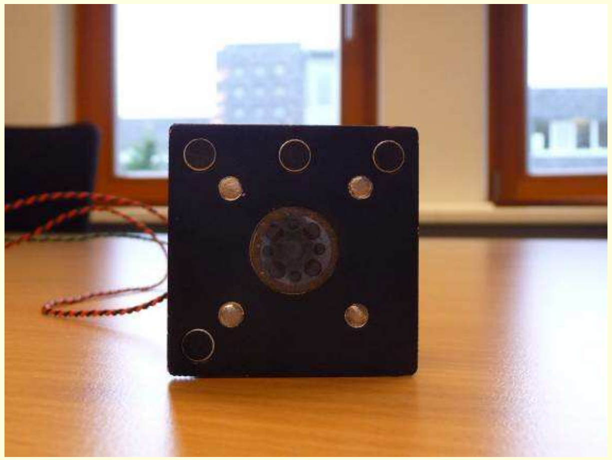
➢ **DANSE scheme** converges after i=2 iterations, moreover achieving **highest performance gain**

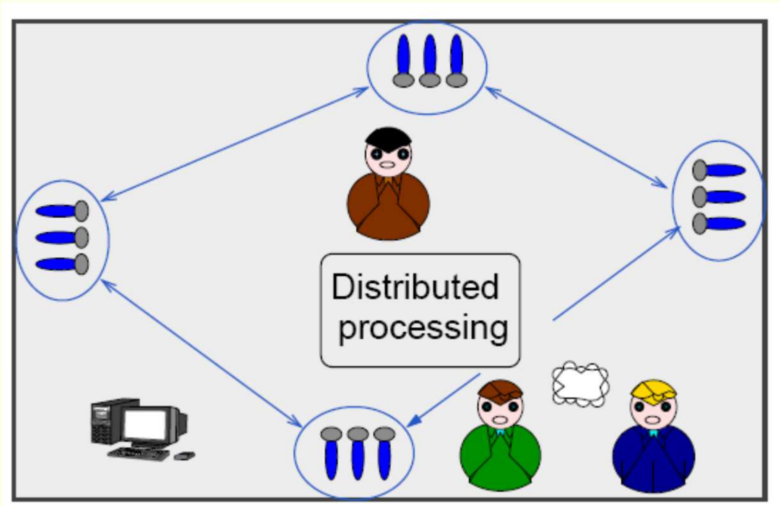
[T. C. Lawin-Ore, S. Doclo, ICASSP 2011]

Acoustic sensor networks

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

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





Conclusions

- Speech enhancement algorithms in hearing instruments
  - More and more microphones: monaural $\rightarrow$ binaural $\rightarrow$ 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