3rd joint meeting of the Acoustical Society of America and the Acoustical Society of Japan,
Honolulu, Hawaii, 2–6/12/96

Adaptive Microphone Arrays for Speech Enhancement in Coherent and Incoherent Noise Fields

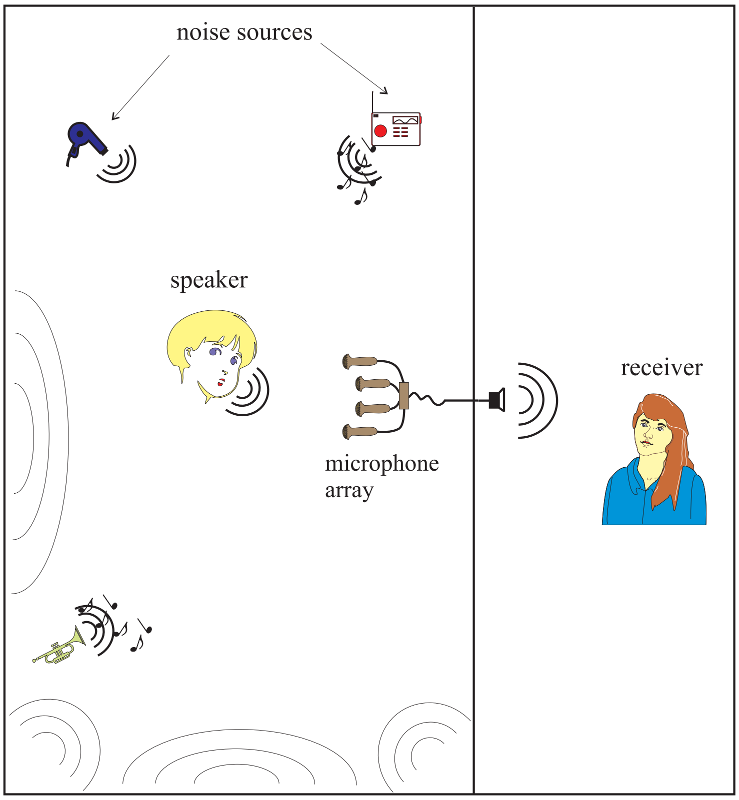
Sven Fischer[a], Karl-Dirk Kammeyer[a] & Klaus Uwe Simmer[b]

[a] University of Bremen
Dept. of Telecommunications
P.O. Box 330 440
D-28334 Bremen
fischer@comm.uni-bremen.de

[b]
Theresenstr. II
D-28203 Bremen
Germany

1. Introduction / Problem Statement
2. Correlation Properties of Noise Fields
3. Microphone Arrays for Noise Reduction
4. Linearly Constrained Beamforming with Adaptive Constraint Values
5. Experimental Results and Conclusion

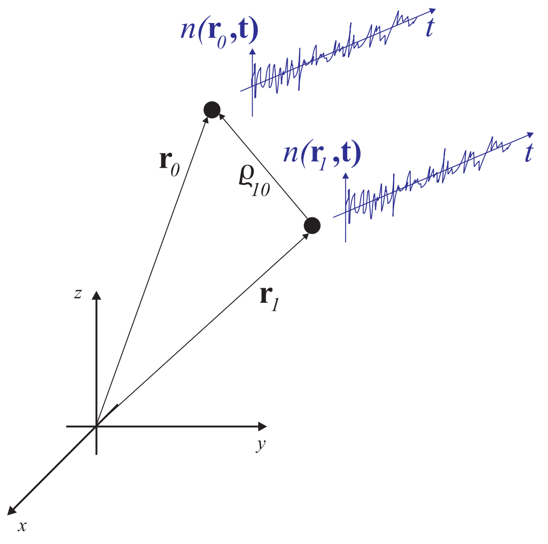
The noise reduction problem in hands–free voice communication systems:



Complexity of the noise field:

- point noise sources / spatialy extended noise sources
- number of noise sources
- movement of noise sources
- kind of noise spectrum (narrowband / broadband)
- noise level
- multipath propagation / reverberation

Spatial coherence of the noise field:



space–time correlation:

$$R(\boldsymbol{\varrho},\tau) = \mathfrak{E}\{n(\mathbf{r},t)n(\mathbf{r}+\boldsymbol{\varrho},t+\tau)\}$$

spatial coherence function:

$$\gamma(\boldsymbol{\varrho}_{ij},\omega) = \frac{\Phi(\boldsymbol{\varrho}_{ij},\omega)}{\sqrt{\Phi(\mathbf{r}_i,\omega)\Phi(\mathbf{r}_j,\omega)}}$$

magnitude squared coherence (MSC):

$$C(\boldsymbol{\varrho}_{ij},\omega) = |\gamma(\boldsymbol{\varrho}_{ij},\omega)|^2 = \frac{|\Phi(\boldsymbol{\varrho}_{ij},\omega)|^2}{\Phi(\mathbf{r}_i,\omega)\Phi(\mathbf{r}_j,\omega)}$$

$\mathbf{r}$: position vector

$\boldsymbol{\varrho}$: spatial difference vector

$$R(\boldsymbol{\varrho},\tau) \;\underset{\tau}{\circ\!\!-\!\!\bullet}\; \Phi(\boldsymbol{\varrho},\omega)$$

free sound field:

- unhindered wave propagation

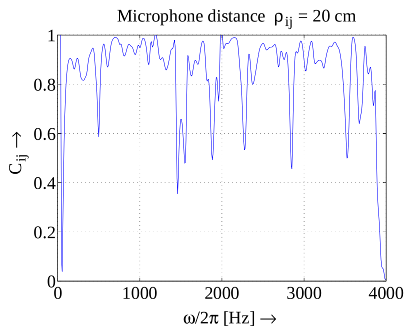
$$C = 1$$

diffuse sound field:

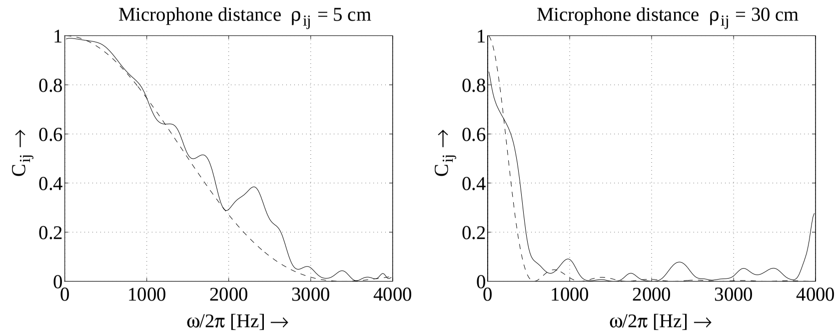
- superposition of an infinite number of plane waves
  - ○ one sound source and reverberation time $\rightarrow \infty$
  - ○ number of sound sources $\rightarrow \infty$ and any reverberation time

$$C(\varrho_{ij},\omega) = \mathbf{si}^2\left(\omega\frac{\varrho_{ij}}{c}\right)$$

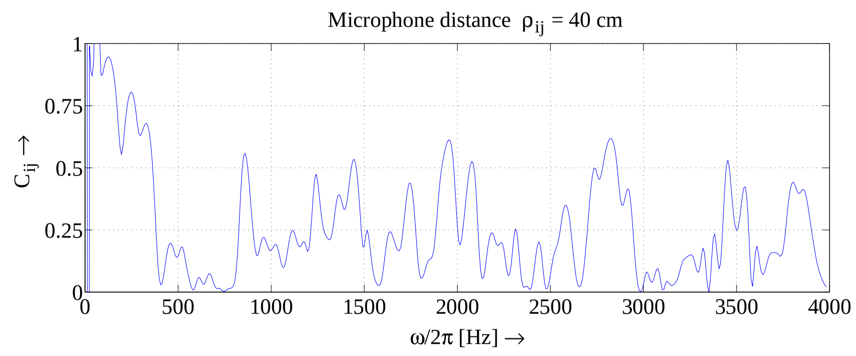
anechoic chamber:



office room:



Combined noise field:



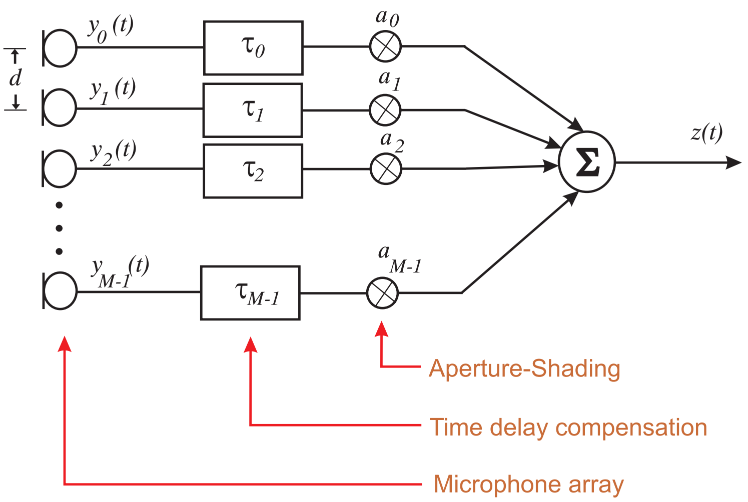
- The distance of the noise sources and the microphones as well as the reflection coefficients of the walls determine the ratio of **coherent direct path noise** and **incoherent diffuse noise**.
- A practical system for noise reduction has to operate independently of the correlation properties of the noise field.

Microphone Arrays for Noise Reduction:

The different approaches can be classified into three main categories:

- **conventional (delay&sum) beamforming**
  (Flanagan et al., 1985; Flanagan et al., 1991; Kellermann, 1991; Khalil et al., 1994)
- **adaptive beamforming**
  (Frost, 1972; Griffiths & Jim, 1982; Kaneda & Ohga, 1986; Sondhi & Elko, 1986)
- **microphone arrays with adaptive postfiltering**
  (Zelinski, 1988; Zelinski, 1990; Simmer & Wasiljeff, 1992)

(a) conventional (delay&sum) beamforming:



The sensor signals are delayed so that the desired signal becomes in phase, weighted (shading) and summed up:

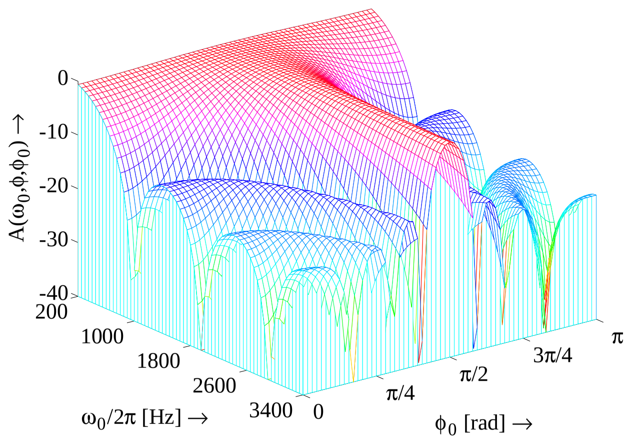
$$z(t) = \sum_{i=0}^{M-1} a_i y_i(t - \tau_i)$$

Achievable noise reduction (NR):

coherent noise field:

- NR is given by the inverse of the beampattern

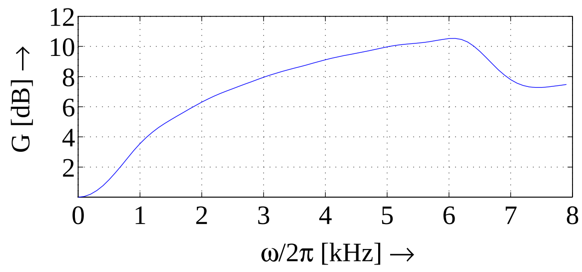
Example:
(M=7, d=0.05 m, uniform shading)



diffuse noise field:

- NR is given by the directivity index (ratio of sound energy received from look–direction to the average energy received from all directions)

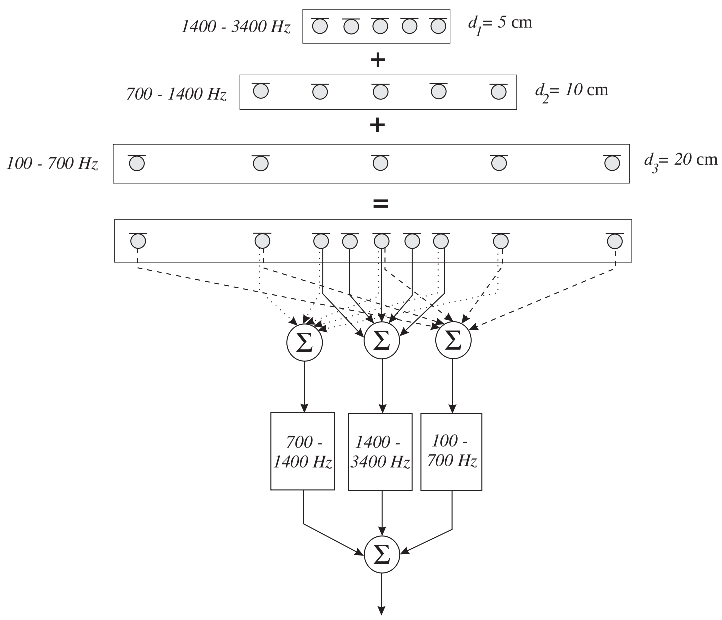
Example:
(M=7, d=0.05 m, uniform shading)

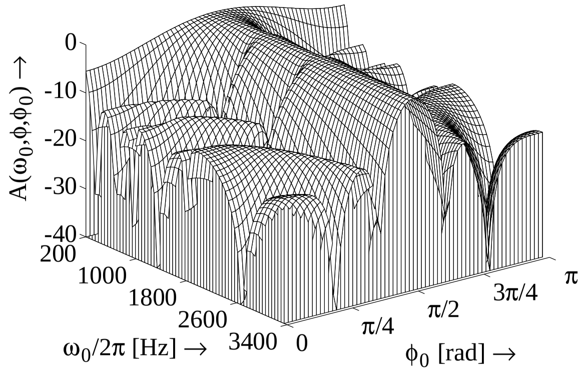


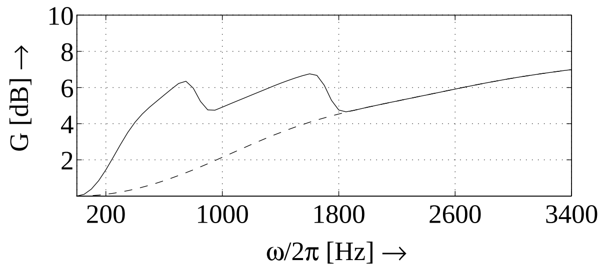
⇒ poor performance at low frequencies

⇒ "frequency invariant" beamforming is desired

Example: harmonically nested subarrays (Flanagan et al., 1985; Kellermann, 1991)

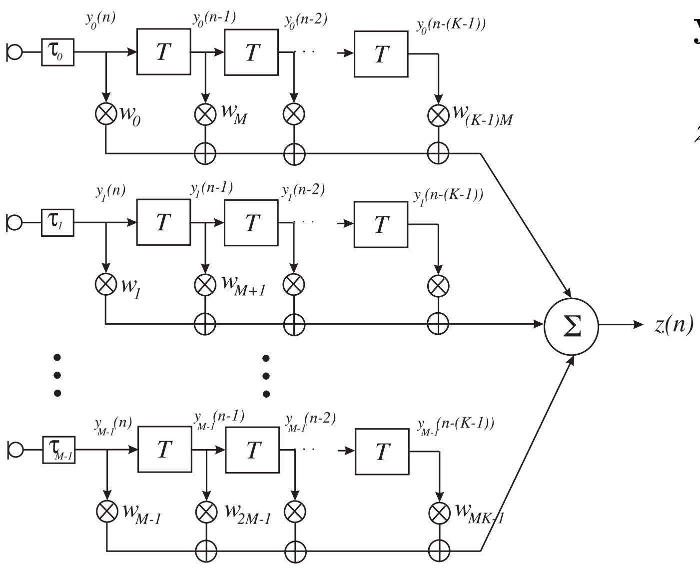






⇒ requires in general many sensors

(b) adaptive beamforming:



$$\mathbf{w} = [w_0, w_1, \ldots, w_{MK-1}]^T$$

$$\mathbf{y}(n) = [y_0(n), y_1(n), \ldots, y_{M-1}(n-K+1)]^T$$

$$z(n) = \mathbf{w}^T\mathbf{y}(n) \qquad \mathbf{R_{yy}} = \mathfrak{E}\{\mathbf{y}(n)\mathbf{y}^T(n)\}$$

$$\boxed{\begin{array}{ll} \min\limits_{\mathbf{w}} & \mathbf{w}^T\mathbf{R_{yy}}\mathbf{w} \\ \text{subject to:} & \mathbf{C}^T\mathbf{w} = \mathbf{f} \end{array}}$$

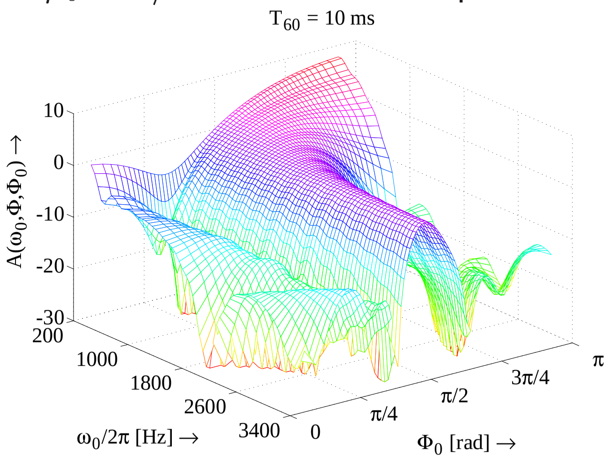
$$\Rightarrow \mathbf{w_{opt}} = \mathbf{R_{yy}^{-1}C}\left(\mathbf{C}^T\mathbf{R_{yy}^{-1}C}\right)^{-1}\mathbf{f}$$

Achievable noise reduction (NR):

coherent noise field:

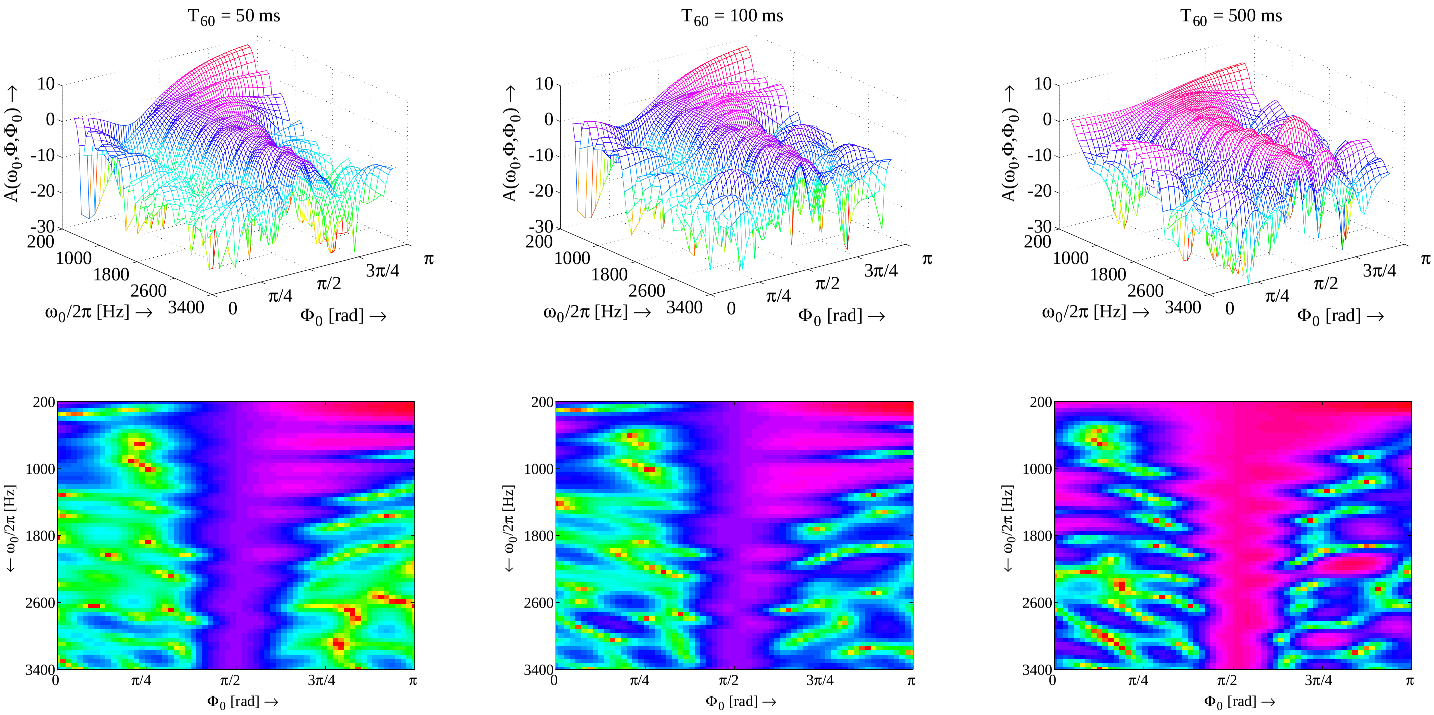
- broadband null in the beampattern in the direction of the noise source

Example:
(M=7, d=0.05 m, single noise source at $\varphi_0 = \pi/4$, 1.5 seconds adaptation time)

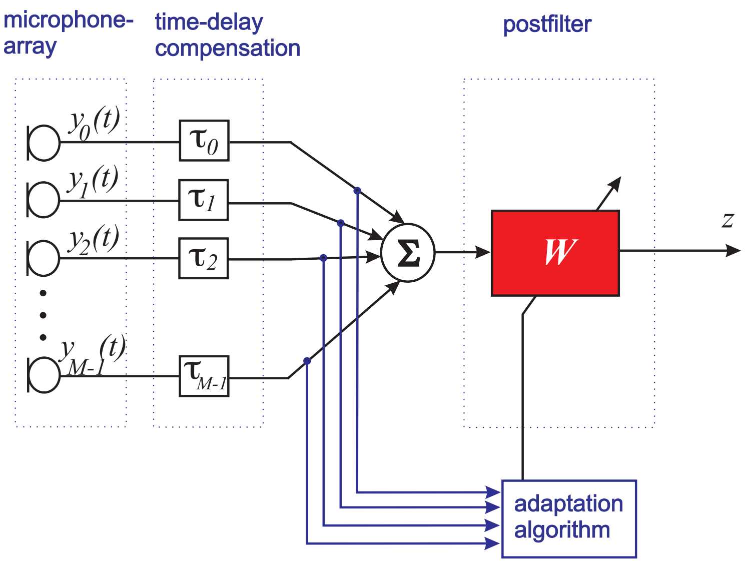


diffuse noise field:

- performance not better than conventional (delay&sum) beamformer



(c) microphone arrays with adaptive postfiltering:



postfilter:
adaptive Wiener–Filter for noise reduction

$$y(n) = s(n) + n(n)$$

$$\mathfrak{E}\{(z(n) - s(n))^2\} \to \text{min.}$$

$$\Rightarrow W(e^{j\Omega}) = \frac{\Phi_{ss}(e^{j\Omega})}{\Phi_{yy}(e^{j\Omega})}$$

$$= \frac{\Phi_{ss}(e^{j\Omega})}{\Phi_{ss}(e^{j\Omega}) + \Phi_{nn}(e^{j\Omega})}$$

Estimation of the Wiener–Filter:

- $\Phi_{yy}(e^{j\Omega})$ can be estimated directly from the disturbed microphone signals
- $\Phi_{ss}(e^{j\Omega})$ can be estimated from the spatial cross–spectrum of the disturbed microphone signals if $C_{ij}(\omega) \approx 0$:

$$\begin{aligned}\mathfrak{E}\{y_i(n)y_j(n+m)\} &= \mathfrak{E}\{[s(n)+n_i(n)]\,[s(n+m)n_j(n+m)]\} \\ &= \mathfrak{E}\{s(n)s(n+m)\} + \underbrace{\mathfrak{E}\{s(n)n_j(n+m)\}}_{=0} \\ &\quad + \underbrace{\mathfrak{E}\{n_i(n)s(n+m)\}}_{=0} + \underbrace{\mathfrak{E}\{n_i(n)n_j(n+m)\}}_{=0}\end{aligned}$$

$\Rightarrow \Phi_{y_iy_j}(e^{j\Omega}) = \Phi_{ss}(e^{j\Omega})$ **for** $i \neq j$ **if** $\mathfrak{E}\{n_i(n)n_j(n+m)\} = 0$**.**

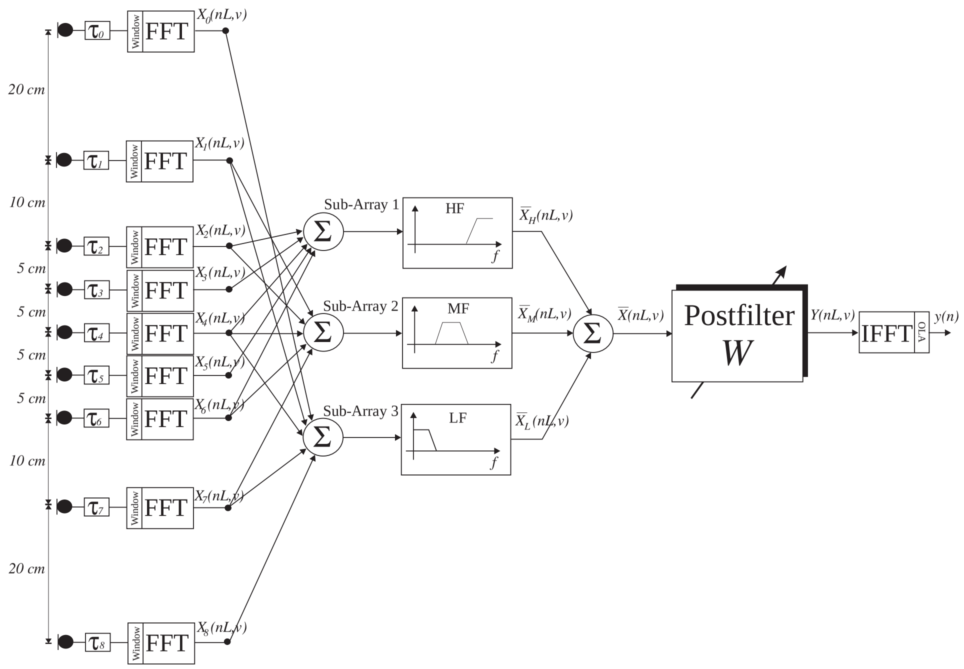
$$\boxed{\widehat{W}(e^{j\Omega}) = \frac{\frac{2}{M(M-1)} \sum_{i=0}^{M-2} \sum_{j=i+1}^{M-1} \widehat{\Phi}_{y_iy_j}(e^{j\Omega})}{\widehat{\Phi}_{yy}(e^{j\Omega})}}$$

⇒ To obtain spatially uncorrelated noise signals, great sensor distances are required in diffuse noise fields

⇒ But:

- great sensor distances yield grating lobes in the visible region of the array (⇒ poor performance in coherent noise fields)
- large interelement spacing results in very narrow beamwidth at higher frequencies (⇒ sensitive to steering misadjustment)

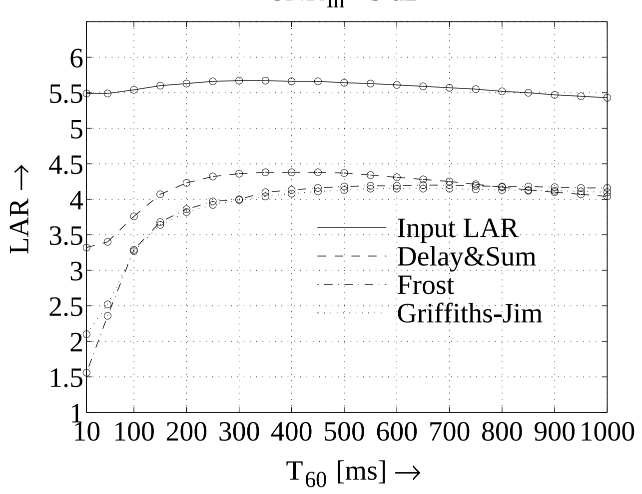
⇒ Remedy:
combination of subarray with postfilter
(Marro et al., 1996; Fischer & Kammeyer, 1997)



⇒ great sensor distances are available for postfilter estimation
⇒ no grating lobes in the visible region of the array

adaptive beamformer:

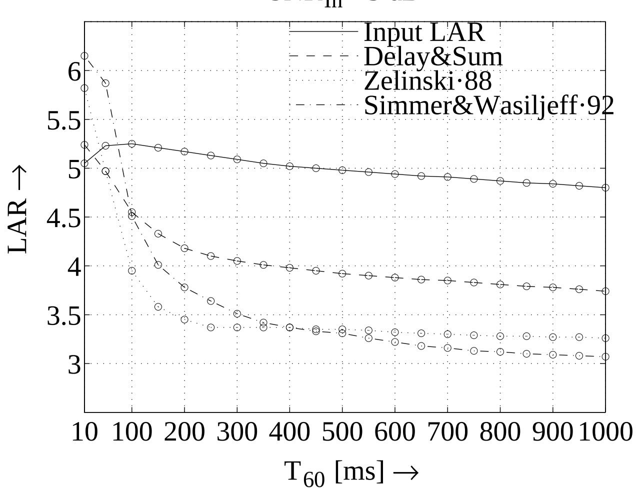
- good for **coherent noise** (low reverberation times)
- performance in **diffuse noise** fields (high reverberation times) not better than delay&sum



T_{60}: Reverberation Time

microphone arrays with postfilter:

- bad for **coherent noise** (low reverberation times)
- good for **diffuse noise** (high reverberation times) better than delay&sum



LAR: Log Area Ratio Distance

Linearly Constrained Beamforming with Adaptive Constraint Values

(Fischer & Simmer, 1996)

optimization criterium:

minimize output power subject to the constraint that the signal from look–direction is the least square estimate of the desired speech signal

constraint values ($\widehat{=}$ impulse response in look–direction) are given by the normal equations

$$\boxed{\begin{array}{ll} \underset{\mathbf{w}}{\textbf{minimize}} & \mathbf{w}^T\mathbf{R}_{\mathbf{yy}}\mathbf{w} \\ \textbf{subject to:} & \mathbf{C}^T\mathbf{w} = \mathbf{R}_{\chi\chi}^{-1}\mathbf{r}_{ss} \end{array}}$$

$$\Rightarrow \mathbf{w}_{\mathbf{opt}} = \mathbf{R}_{\mathbf{yy}}^{-1}\mathbf{C}\left(\mathbf{C}^T\mathbf{R}_{\mathbf{yy}}^{-1}\mathbf{C}\right)^{-1}\left(\mathbf{R}_{\chi\chi}^{-1}\mathbf{r}_{ss}\right)$$

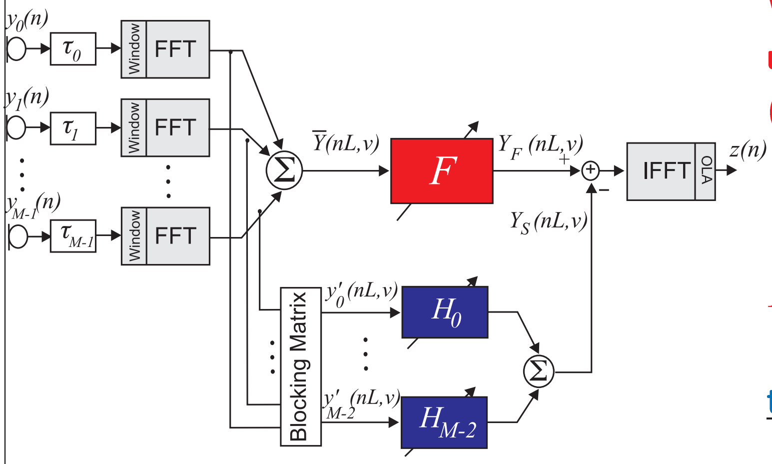
⇒ supression of **coherent noise** due to null steering

⇒ reduction of **diffuse noise** due to Wiener filtering of the signal from look direction

$\mathbf{R}_{\chi\chi}$: autocorrelation matrix of the input signal of the equivalent FIR filter in look–direction
$\mathbf{r}_{ss}$: autocorrelation vector of desired speech signal (estimated by spatial cross correlation)

Realization in open–loop GSC form:

- Implementation in frequency domain with short–time fourier transform and overlap–add
- GSC with two adaptive portions



transfer function in look–direction:

Wiener–Filter in the case of spatially uncorrelated noise

($\widehat{F} \approx 1$ in the case of coherent noise)

$$\widehat{F}(e^{j\Omega}) = \frac{\frac{2}{M(M-1)} \sum_{i=0}^{M-2} \sum_{j=i+1}^{M-1} \left| \widehat{\Phi}_{y_i y_j}(e^{j\Omega}) \right|}{\widehat{\Phi}_{\bar{y}\bar{y}}(e^{j\Omega})}$$

transfer function of sidelobe canceller:

$$H_i(e^{j\Omega}) = \frac{\Phi_{y'_i y_F}(e^{j\Omega})}{\Phi_{y'_i y'_i}(e^{j\Omega})}$$

Estimation of transfer function in look–direction via exponentially averaged Nuttall/Carter periodogram (Nuttall & Carter, 1982):

$$\widehat{\widehat{\Phi}}_{ss}(nL,v) = \alpha \widehat{\widehat{\Phi}}_{ss}((n-1)L,v) + \frac{2}{M(M-1)} \sum_{i=0}^{M-2} \sum_{j=i+1}^{M-1} Y_i^*(nL,v) Y_j(nL,v)$$

$$\widehat{\widehat{\Phi}}_{\bar{y}\bar{y}}(nL,v) = \alpha \widehat{\widehat{\Phi}}_{\bar{y}\bar{y}}((n-1)L,v) + |\overline{Y}(nL,v)|^2$$

$$\left.\begin{matrix} \widehat{\widehat{R}}_{ss}(nL,m) \\ \widehat{\widehat{R}}_{\bar{y}\bar{y}}(nL,m) \end{matrix}\right\} \Rightarrow \begin{matrix} \widehat{R}_{ss}(nL,m) = \widehat{\widehat{R}}_{ss}(nL,m) w_{lag}(m) \\ \widehat{R}_{\bar{y}\bar{y}}(nL,m) = \widehat{\widehat{R}}_{\bar{y}\bar{y}}(nL,m) w_{lag}(m) \end{matrix}$$

$$\widehat{\Phi}_{ss}(nL,v)$$
$$\widehat{\Phi}_{\bar{y}\bar{y}}(nL,v)$$

$\Rightarrow$ smoothing in time and frequency

$\Rightarrow$ convolution constrained to be linear by proper selection of lag window w_{lag}

Signal cancellation:

If the target signal is not completely blocked out by the blocking matrix, the adaptive filters H_i adapts to the target signal; this results in (partial) cancellation of the target signal.

Reasons for signal leakage through blocking matrix are:

- steering misadjustment
- different attenuation of the speech signal due to spherical wave propagation
- non ideal microphones
- multipath propagation / reverberation
- ⋮

⇒ "robust" beamforming methods are recommended (Cox et al., 1987; Greenberg & Zurek, 1992; Hoffman et al., 1994)

Simple approach to limit signal cancellation:

Bound the transfer function magnitude of the sidelobe–cancelling filters H_i between values of 0 and 1

⇒ signal leakage through blocking matrix will not be amplified

⇒ amount of signal cancellation is limited

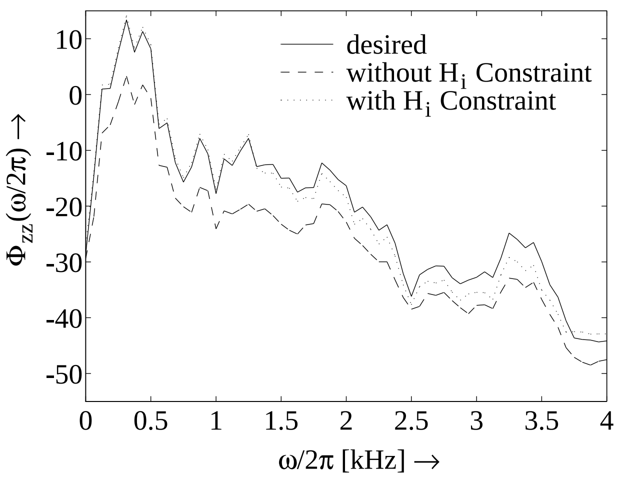
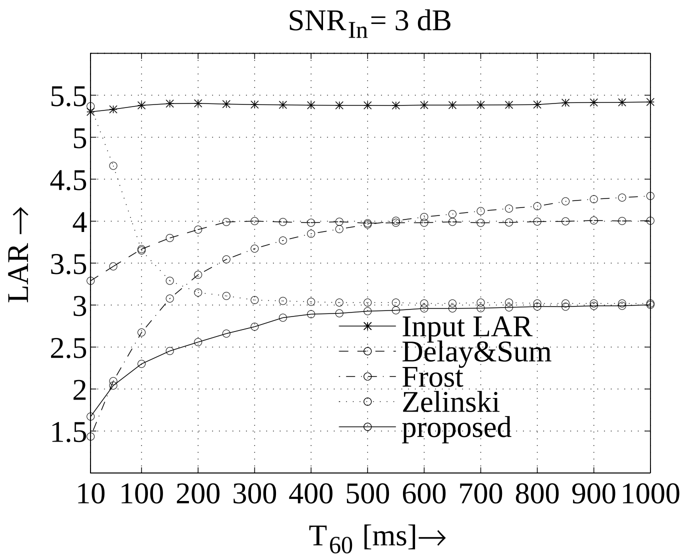
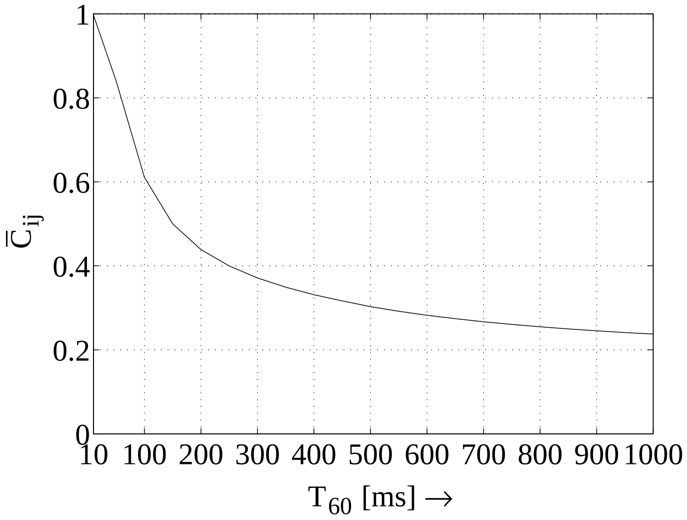


Fig.: Power density of output signal with speech only input and steering misadjustment $\approx 5°$

Objective quality measure:

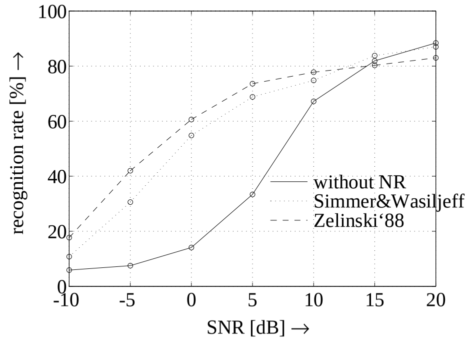


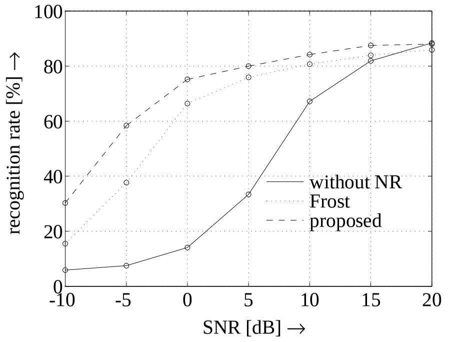


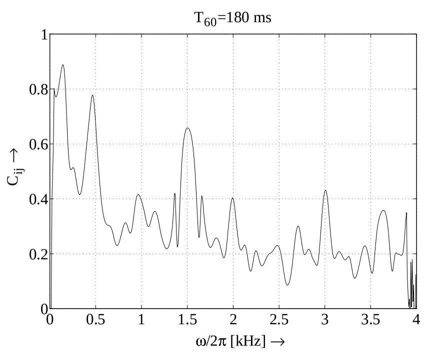
⇒ Performance of proposed method in **coherent noise** field (low reverberation time) same as for adaptive beamformer

⇒ Performance in **diffuse noise** field (high reverberation time) same as microphone array with adaptive postfiltering

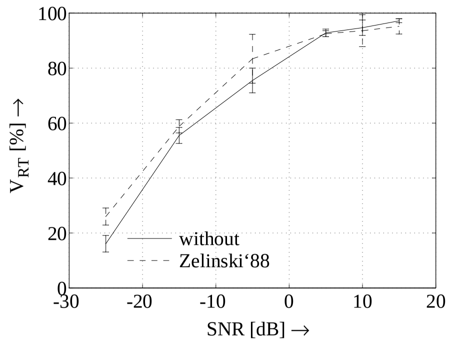
Recognition rate of speaker independent isolated word recognition system:

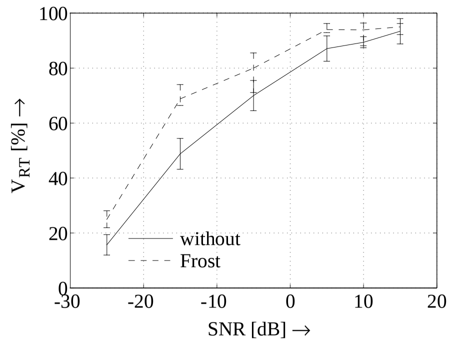


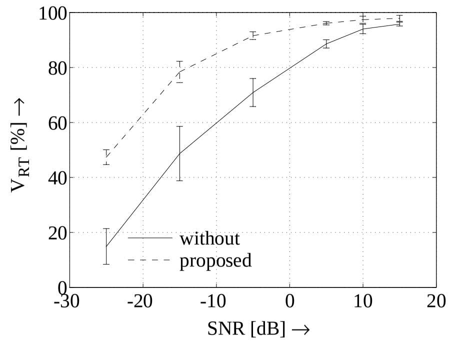




Intelligibility measurements (rhyme test):







Conclusions

Microphone Arrays for Noise Reduction

conventional (delay&sum) beamforming

- requires in general many sensors

adaptive methods to increase noise reduction performance:

adaptive beamforming

- requires **coherent** noise field
- in **diffuse** noise field not better than delay&sum

microphone arrays with adaptive postfiltering

- requires **diffuse** noise field
- in **coherent** noise field distortions may appear

adaptive beamforming with adaptive constraint values

- works well in **coherent and diffuse** noise fields
- no a priori assumptions about correlation properties of the noise field required

References

Cox, H., Zeskind, R., & Owen, M. (1987). Robust adaptive beamforming. *IEEE Trans. Acoust. Speech, Signal Processing*, ASSP–35(10), 1365–1376.

Fischer, S. (1996). *Adaptive Mehrkanalgeräuschunterdrückung bei gestörten Sprachsignalen unter Berücksichtigung der räumlichen Kohärenz des Geräuschfeldes*. PhD thesis, Universität Bremen. Verlag Shaker, Aachen (in German).

Fischer, S. & Kammeyer, K. (1997). Broadband beamforming with adaptive postfiltering for speech acquisition in noisy environments. Submitted to ICASSP–97.

Fischer, S. & Simmer, K. (1996). Beamforming microphone arrays for speech acquisition in noisy environments. *Speech Communication*. Dec. To be published.

Flanagan, J., Berkley, D., Elko, G., West, J., & Sondhi, M. (1991). Autodirective microphone systems. *Acustica*, 73, 58–71.

Flanagan, J., Johnston, J., Zahn, R., & Elko, G. (1985). Computer–steered microphone arrays for sound transduction in large rooms. *J. Acoust. Soc. Amer.*, 78(5), 1508–1518.

Frost, O. (1972). An algorithm for linearly constrained adaptive array processing. *Proc. IEEE*, 60(8), 926–935.

Greenberg, J. & Zurek, P. (1992). Evaluation of an adaptive beamforming method for hearing aids. *J. Acoust. Soc. Amer.*, 91(3), 1662–1676.

Griffiths, L. & Jim, C. (1982). An alternative approach to linearly constrained adaptive beamforming. *IEEE Trans. Antennas Propagat.*, AP-30(1), 27–34.

Hoffman, M., Trine, T., & Van Tasell, D. (1994). Robust adaptive microphone array processing for hearing aids: realistic speech enhancement. *J. Acoust. Soc. Amer.*, 96(2), 759–770.

Kaneda, Y. & Ohga, J. (1986). Adaptive microphone–array system for noise reduction. *IEEE Trans. Acoust. Speech, Signal Processing*, ASSP-34(6), 1391–1400.

Kellermann, W. (1991). A self–steering digital microphone array. In *Proc. of the Internat. Conference on Acoustics, Speech and Signal Processing ICASSP–91* (pp. 3581–3584).

Khalil, F., Jullien, J., & Gilloire, A. (1994). Microphone array for sound pickup in teleconference systems. *J. Audio Eng. Soc.*, 42(9), 691–700.

Marro, C., Mahieux, Y., & Simmer, K. (1996). Performance of adaptive dereverberation techniques using directivity controlled arrays. In *Proc. European Signal Processing Conf. EUSIPCO–96* (pp. 1127–1130). Trieste, Italy.

Nuttall, A. H. & Carter, G. (1982). Spectral estimation using combined time and lag weighting. *Proc. IEEE*, 70(9), 1115–1125.

Simmer, K. & Wasiljeff, A. (1992). Adaptive microphone arrays for noise suppression in the frequency domain. In *Second Cost 229 Workshop on Adaptive Algorithms in Communications* (pp. 185–194). Bordeaux, France.

Sondhi, M. & Elko, G. (1986). Adaptive optimization of microphone arrays under a nonlinear constraint. In *Proc. of the Internat. Conference on Acoustics, Speech and Signal Processing ICASSP–86*, volume 2 (pp. 981–984). Tokio, Japan.

Zelinski, R. (1988). A microphone array with adaptive post–filtering for noise reduction in reverberant rooms. In *Proc. of the Internat. Conference on Acoustics, Speech and Signal Processing ICASSP–88* (pp. 2578–2581). New York.

Zelinski, R. (1990). Noise reduction based on microphone array with LMS adaptive post–filtering. *Electronics Letters*, 26(24), 2036–2037.