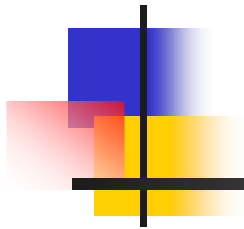


Speech Enhancement for Hands-Free Terminals



Nedelko Grbic,
Sven Nordholm and Anders Johansson



Contents

- Handsfree Telephony Principles
- Handsfree problem
- Optimal Beamformers
 - Linearly Constrained Minimum Variance Beamformer
 - Optimal Signal-to-Noise plus Interference
 - Diffuse Noise Field Beamformer
 - Minimum Mean Square Error Beamformer
- Results in a real environment
- Conclusions

Handsfree Telephony

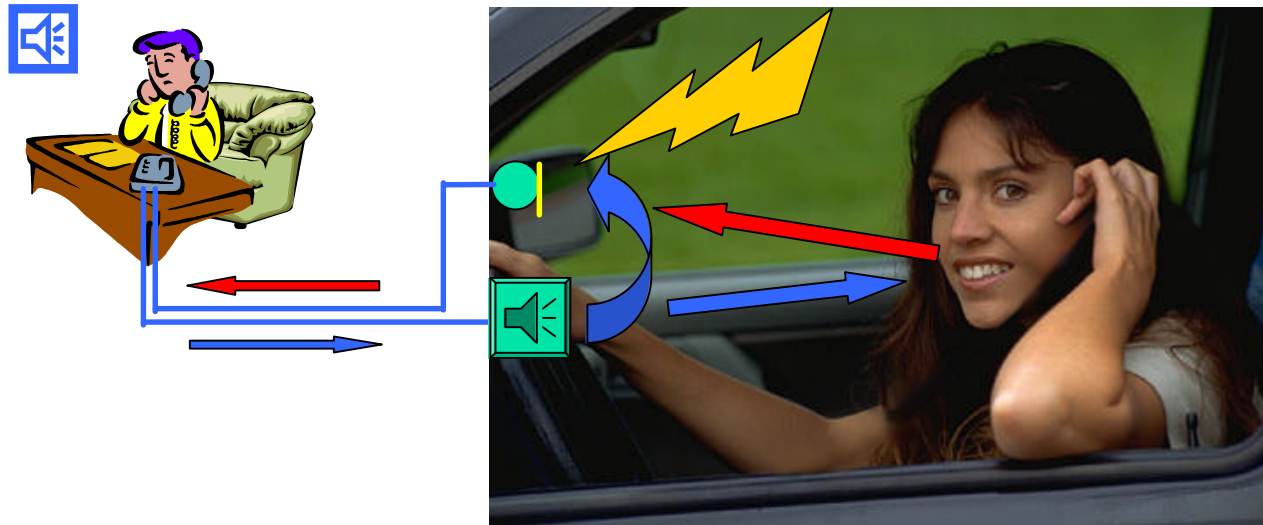
- Safety problems in cars
- Inconvenience of conversation
- Prohibited by legislation in some regions



Handsfree Telephony

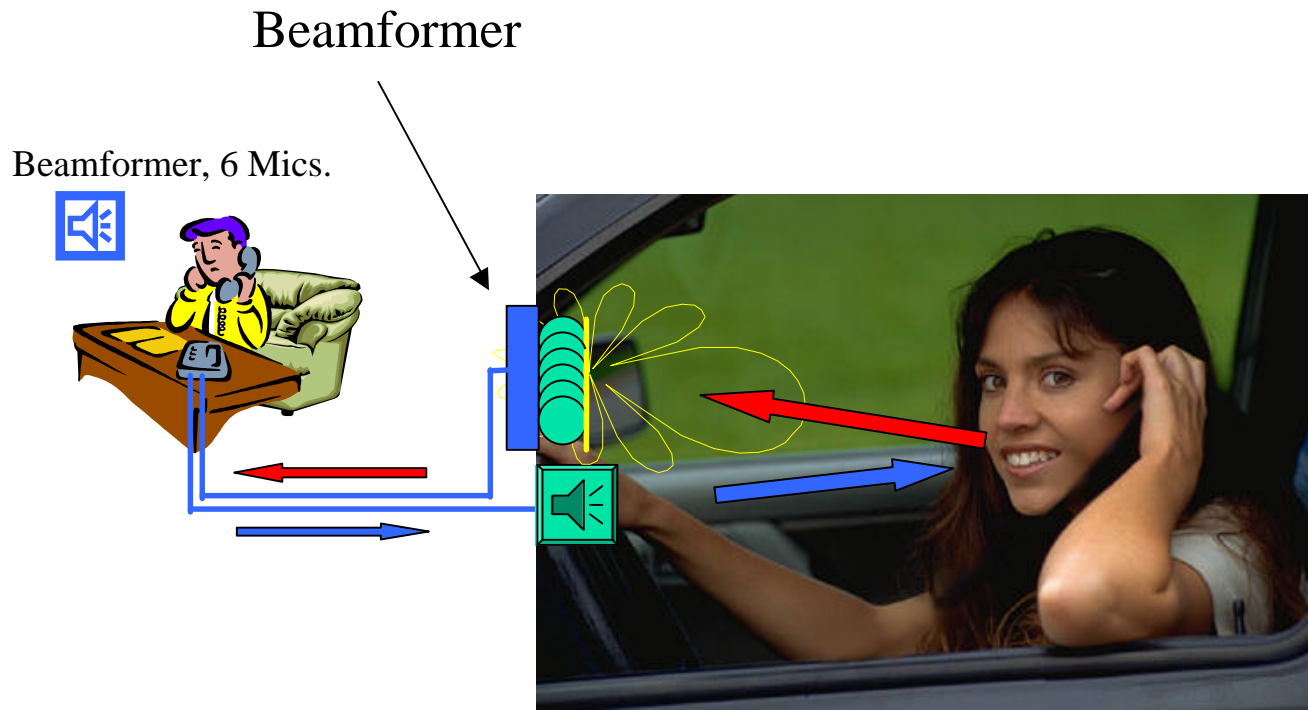
- Perception problems
 - Acoustic feedback
 - Wind and Tire friction in cars
 - Engine and Fan noise

Single Mic.

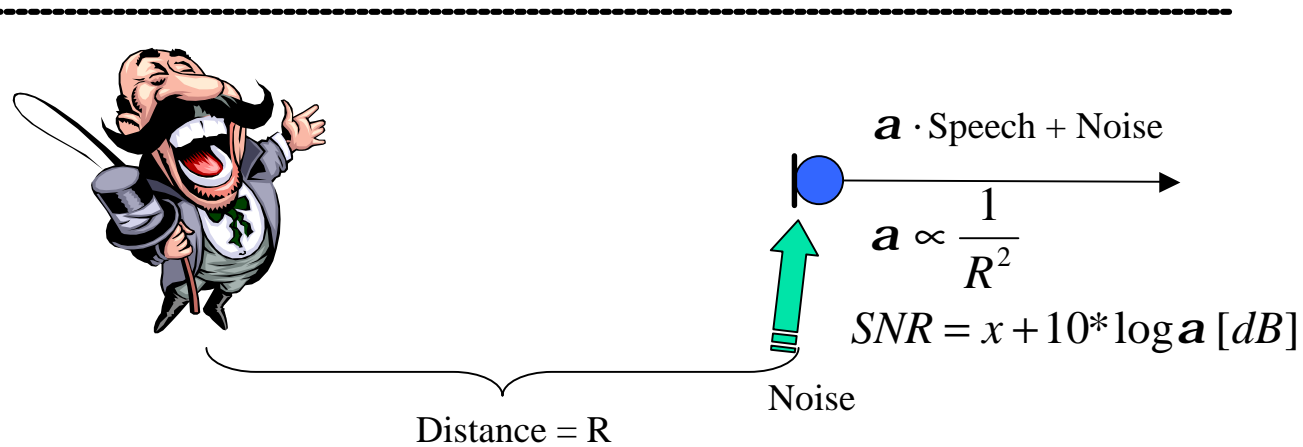
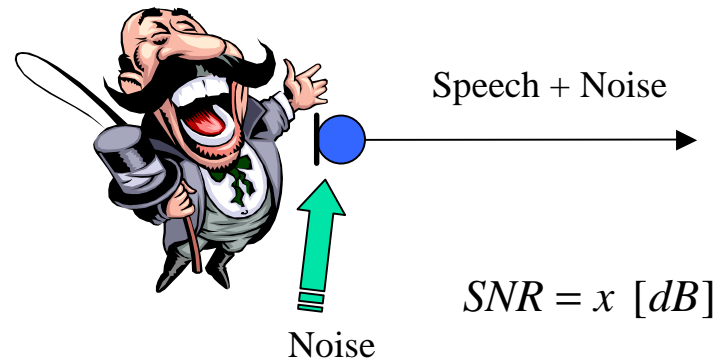


Handsfree Telephony

- Speech enhancement by means of beamforming

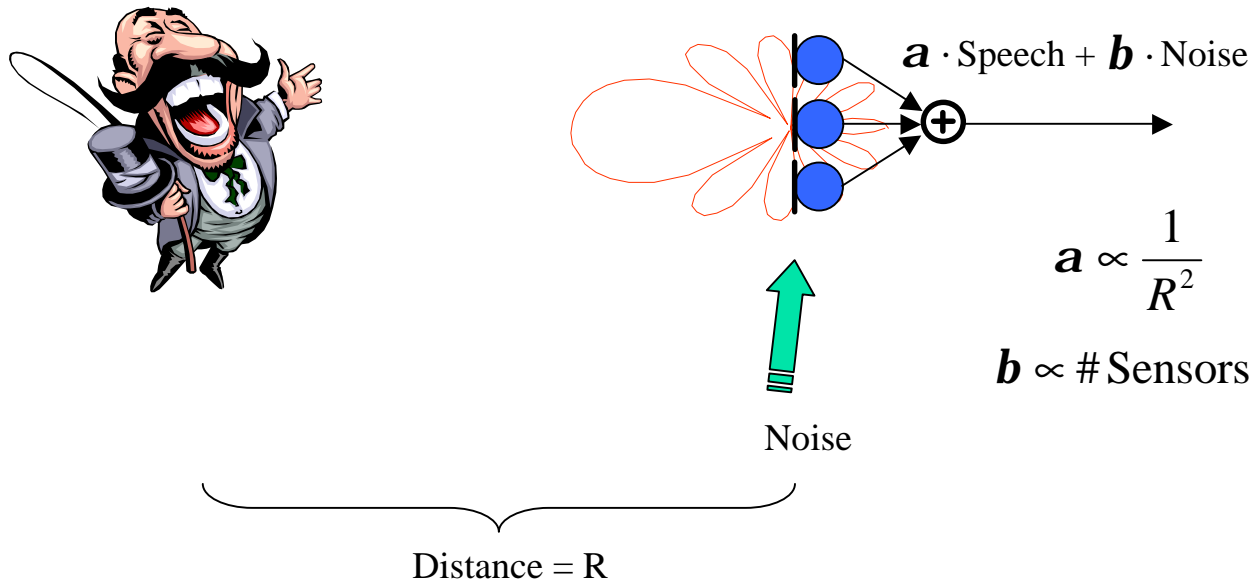


Handsfree problem

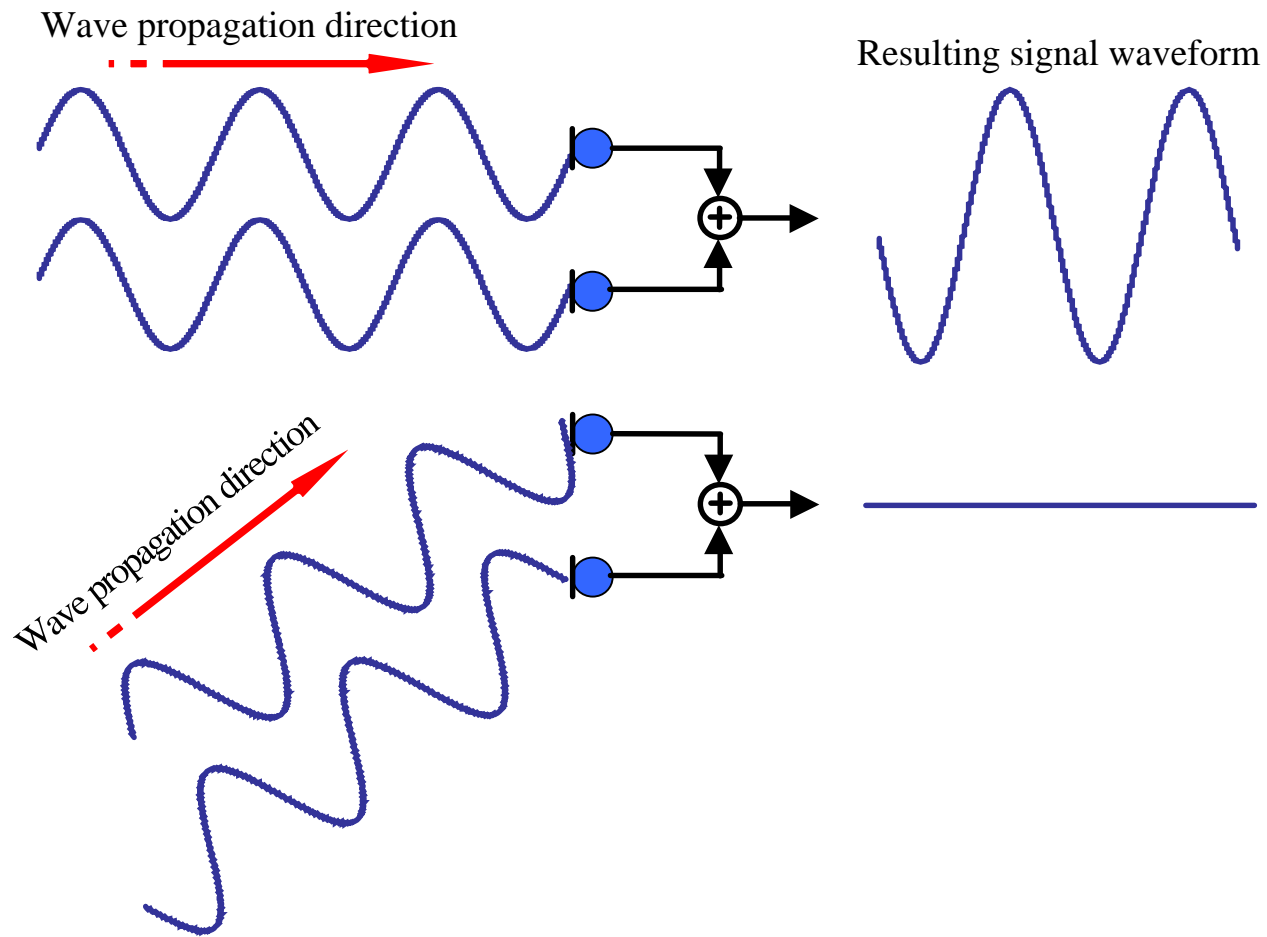


Handsfree Improvement

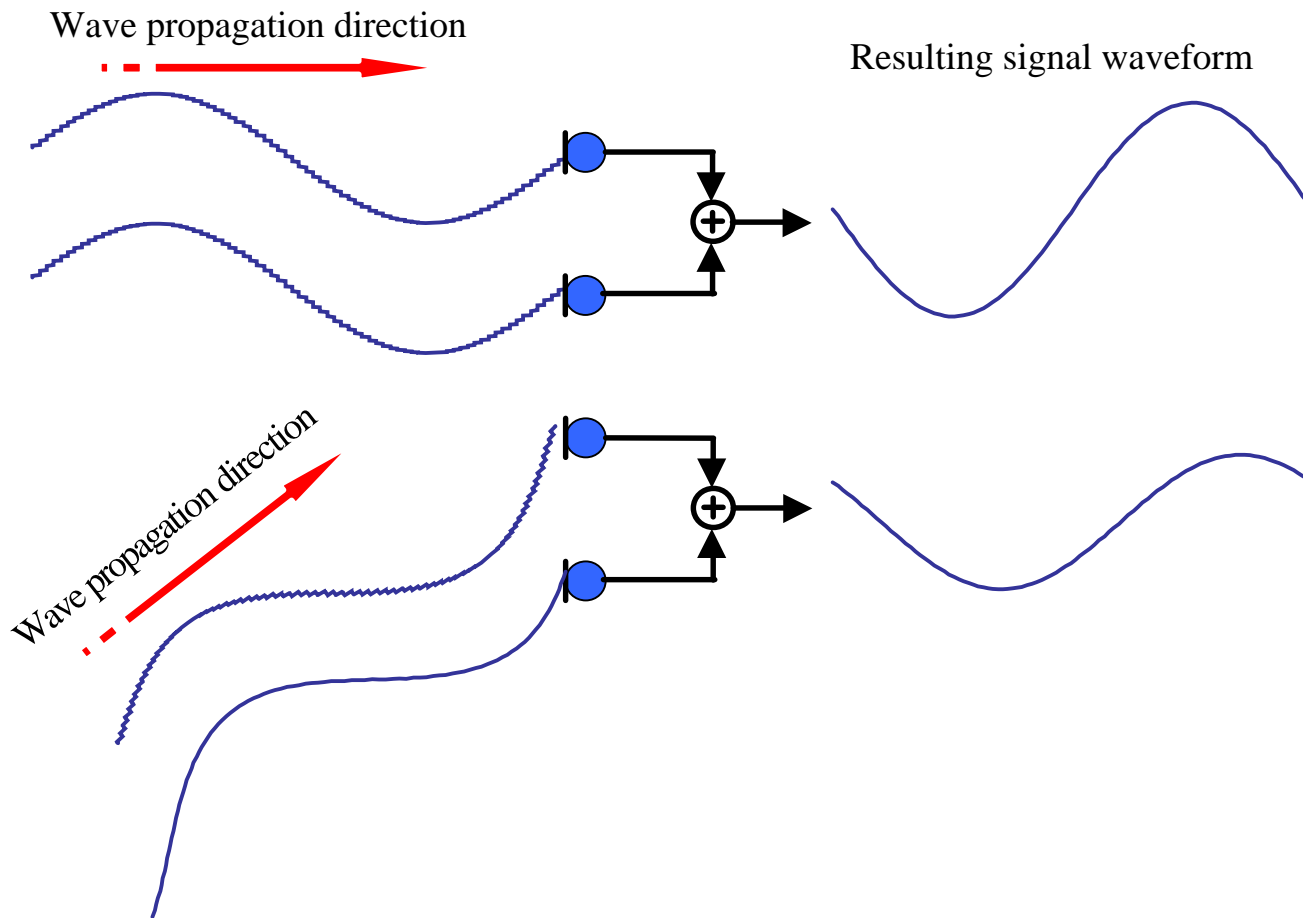
$$SNR = x + 10 \cdot \log\left(\frac{a}{b}\right) \text{ [dB]}$$



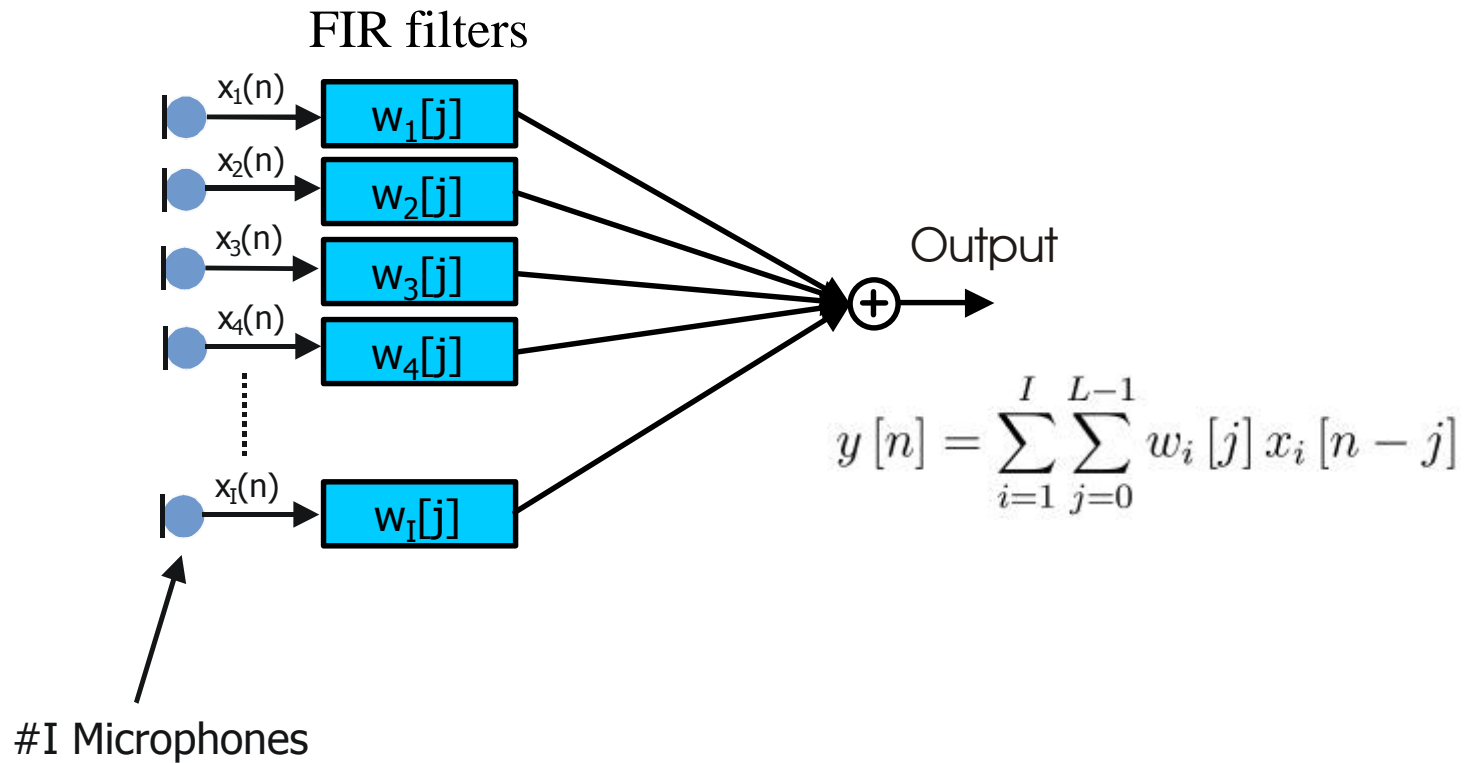
Spatial Selectivity



Spatial Selectivity

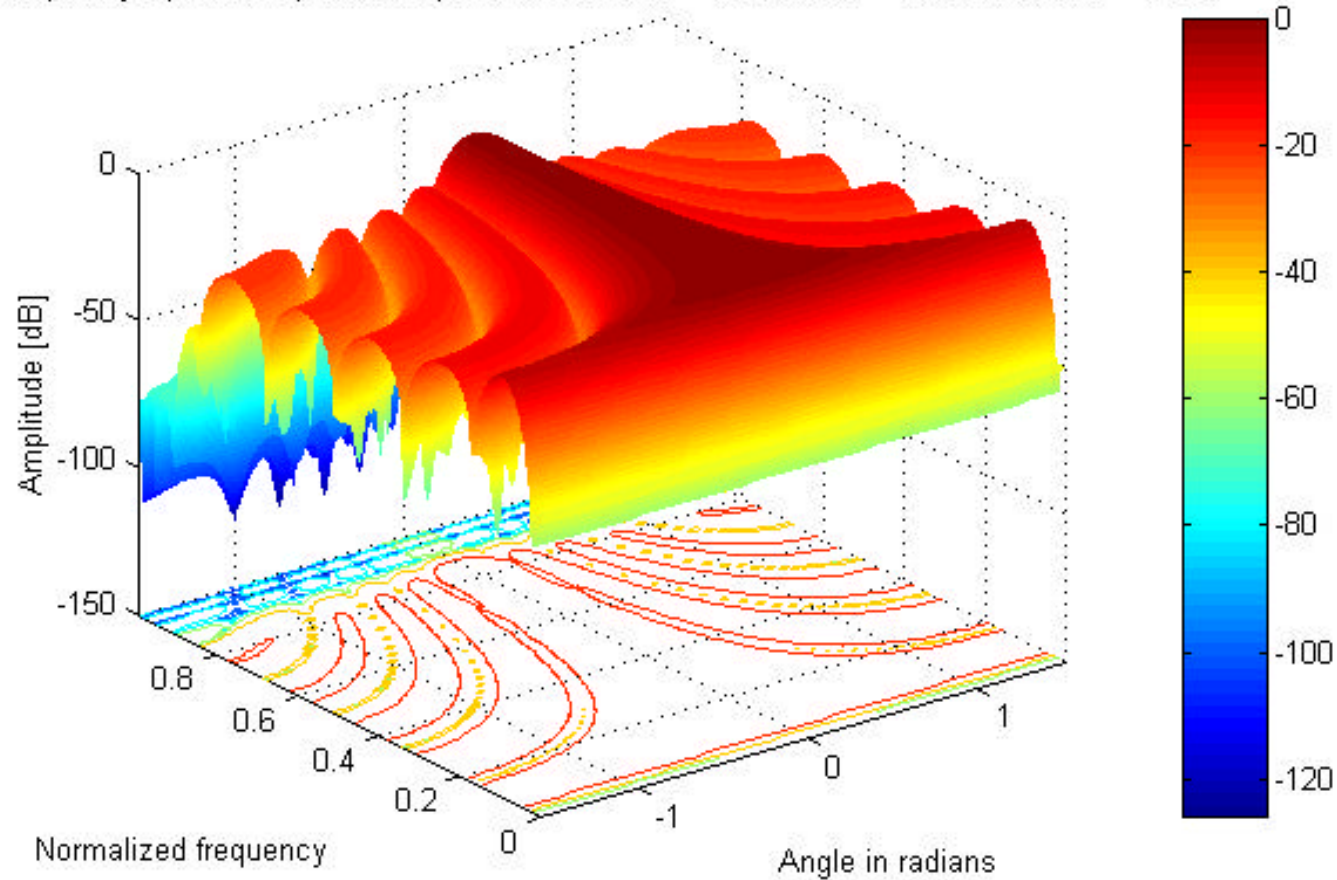


Broadband Beamformer



Ex. Broadband response

Frequency-Spatial Amplitude response , #Sensors = 10, Radius = 10, Sensor dist. = 0.05





Beamforming approaches

Data independent Beamformers

- The Delay and Sum Beamformer
- Multidimensional Filter designed Beamformers

Statistical Beamformers

- Linearly Constrained Minimum Variance Beamforming
- The Optimal Signal-to-Noise plus Interference (SNIB) Beamformer
- Minimum Mean Square Beamformer
- Diffuse Noise Field Beamformer

Linearly Constrained Minimum Variance Beamformer (LCMV)

For each frequency, the weights are found from:

$$\left. \begin{array}{l} \min_{\mathbf{w}} \mathbf{w}^H \mathbf{R}_{\mathbf{xx}}(\Omega) \mathbf{w} \\ \text{Subject to:} \\ \mathbf{d}^H(R, \theta, \Omega) \mathbf{w} = g^* \end{array} \right\} \Rightarrow$$

$$\Rightarrow \mathbf{w}_{opt} = g^* \frac{\mathbf{R}_{\mathbf{xx}}^{-1}(\Omega) \mathbf{d}(R, \theta, \Omega)}{\mathbf{d}^H(R, \theta, \Omega) \mathbf{R}_{\mathbf{xx}}^{-1}(\Omega) \mathbf{d}(R, \theta, \Omega)}$$

$$\text{where } \mathbf{d}(R, \theta, \Omega) = \left[\frac{1}{R}, \frac{1}{R_1} e^{-j\Omega\tau_1(R, \theta)}, \dots, \frac{1}{R_{L-1}} e^{-j\Omega\tau_{L-1}(R, \theta)} \right]$$

The correlation matrix $\mathbf{R}_{\mathbf{xx}}(\Omega)$ contains contributions from all sources

If $g=1$, then LCMV equals the Minimum Variance Distortionless Response, (MVDR)



Optimal SNIB Beamformer

$$Q = \frac{\text{average signal output power}}{\text{average noise-plus-interference output power}}$$

$$\mathbf{w}_{opt} = \arg \max_{\mathbf{w}} \frac{\mathbf{w}^H \mathbf{R}_{ss} \mathbf{w}}{\mathbf{w}^H \mathbf{R}_{nn} \mathbf{w}}$$

The correlation matrix \mathbf{R}_{ss} contains contributions from the source of interest and \mathbf{R}_{nn} contains contributions from all other sources

The weights that maximizes the quote, are found from the Generalized Eigenvalue relation, i.e.,

$$\begin{aligned} \mathbf{w}_{opt} = \arg \max_{\mathbf{w}} \left\{ \frac{\mathbf{w}^H \mathbf{R}_{ss} \mathbf{w}}{\mathbf{w}^H \mathbf{R}_{nn} \mathbf{w}} \right\} &\Rightarrow \mathbf{R}_{ss} \mathbf{w}_{opt} = \lambda_{max} \mathbf{R}_{nn} \mathbf{w}_{opt} \\ &\Rightarrow \mathbf{R}_{nn}^{-1} \mathbf{R}_{ss} \mathbf{w}_{opt} = \lambda_{max} \mathbf{w}_{opt} \end{aligned}$$



MMSE Beamformer

$$\mathbf{w}_{opt} = \arg \min_{\mathbf{w}} E \left\{ |y[n] - s_r[n]|^2 \right\} \quad r \in [1, 2, \dots, I]$$

$$\mathbf{w}_{opt} = \arg \min_{\mathbf{w}} \left\{ E \left[|\mathbf{w}^H \mathbf{s}[n] - s_r[n]|^2 + |\mathbf{w}^H \mathbf{x}[n]|^2 \right] \right\}$$

$$\mathbf{w}_{opt} = \arg \min_{\mathbf{w}} \left\{ \mathbf{w}^H [\mathbf{R}_{ss} + \mathbf{R}_{xx}] \mathbf{w} - \mathbf{w}^H \mathbf{r}_s - \mathbf{r}_s^H \mathbf{w} + r_{s_r} \right\}$$

$$\mathbf{w}_{opt} = [\mathbf{R}_{ss} + \mathbf{R}_{xx}]^{-1} \mathbf{r}_s$$



Diffuse Noise Field beamformer

For each frequency, the weights are found from:

$$\mathbf{w}_{opt} = \mathbf{R}_{nn}^{-1}(\Omega) \mathbf{d}(R, \theta, \Omega)$$

where

$$\mathbf{d}(R, \theta, \Omega) = \left[\frac{1}{R}, \frac{1}{R_1} e^{-j\Omega\tau_1(R,\theta)}, \dots, \frac{1}{R_{L-1}} e^{-j\Omega\tau_{L-1}(R,\theta)} \right]$$

and

$$\mathbf{R}_{n_i n_j}(\Omega) = \sigma^2 \frac{\sin(kd_{ij})}{kd_{ij}}$$

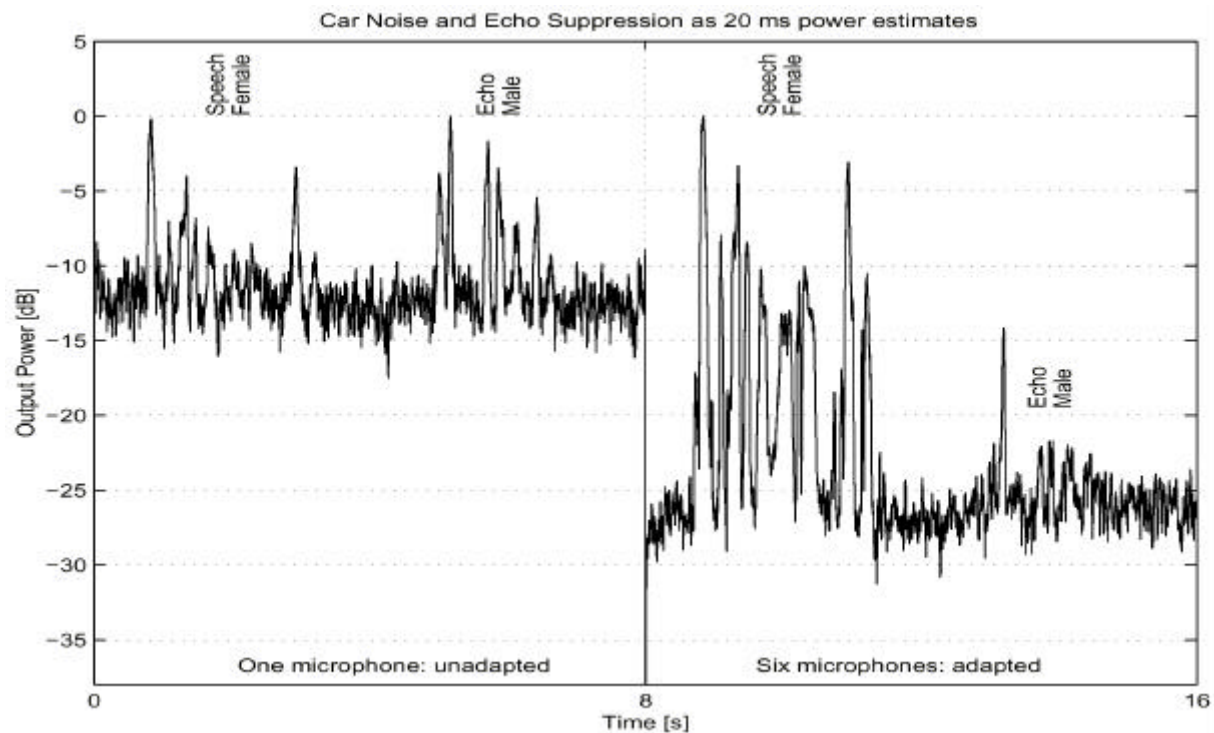


Evaluation Conditions

- Environment in car running at 110 km/h
- Linear sensor array
- 6 sensors with 12 kHz sampling rate
- Evaluation on real speech signals

Results

Performance [dB]	Speech Distortion	Noise Suppression	Interference Suppression
SNIB	-19.4	18.1	30.7
MMSE	-30.6	15.2	17.2
Diffuse Noise Field	-26.5	4.0	1.9





Conclusions

- Multisensor techniques are efficient in a terminal handsfree situation
- An SNR improvement of 15-18 dB can be achieved with six sensors
- The SNIB has better noise suppression than the MMSE but also more distortion
- The Diffuse Noise field model is inaccurate in a car handsfree environment