Microsoft
AI and Research

Directional Interference Suppression using a Spatial Relative Transfer Function Feature

Sebastian Braun, Ivan Tashev

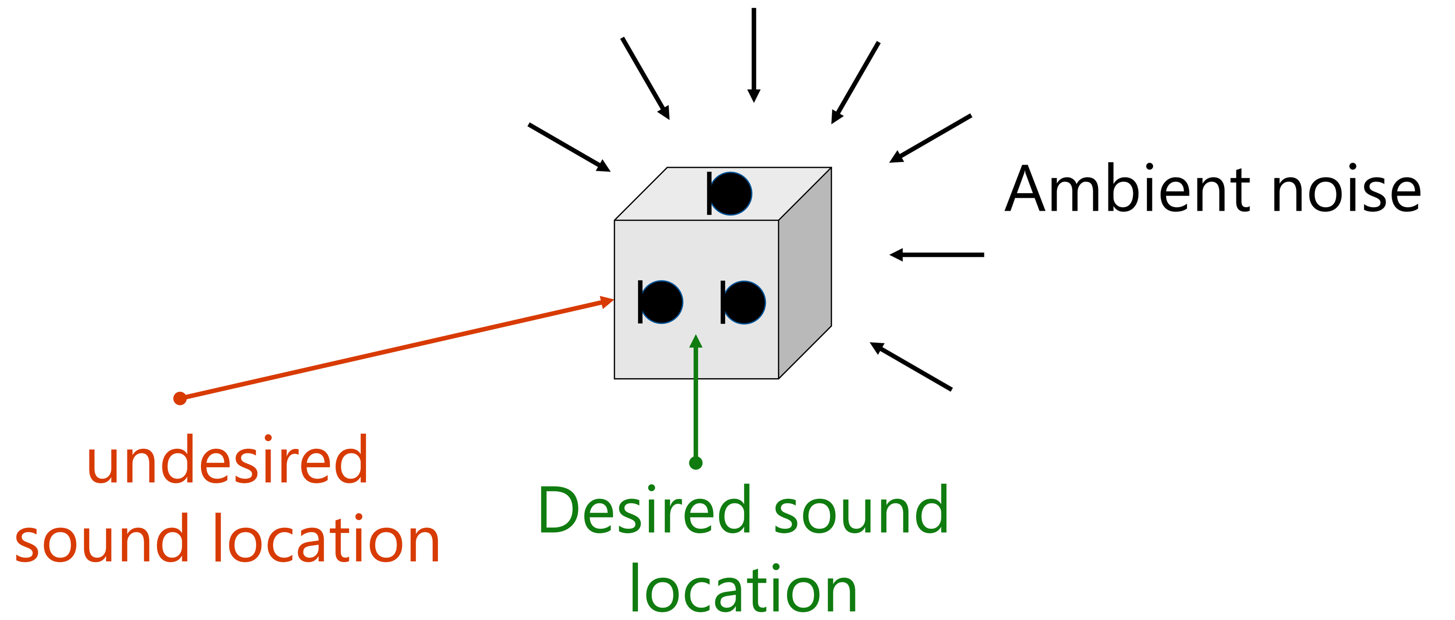
Problem description

Many speech enhancement systems consist of a beamformer and a spectral postfilter. Such spectral suppressors are typically designed for non-directional noise fields, and therefore are suboptimal to suppress highly directional interference such as interfering talkers. We present a spatial suppressor based on the spatial presence probability, taking magnitude and phase of relative transfer functions.

Signal model

STFT domain model assuming one dominant directional sound due to speech sparsity

$$Y_m(k,n) = A_m(k,\mathbf{r})\, S_\mathbf{r}(k,n) + V_m(k,n)$$



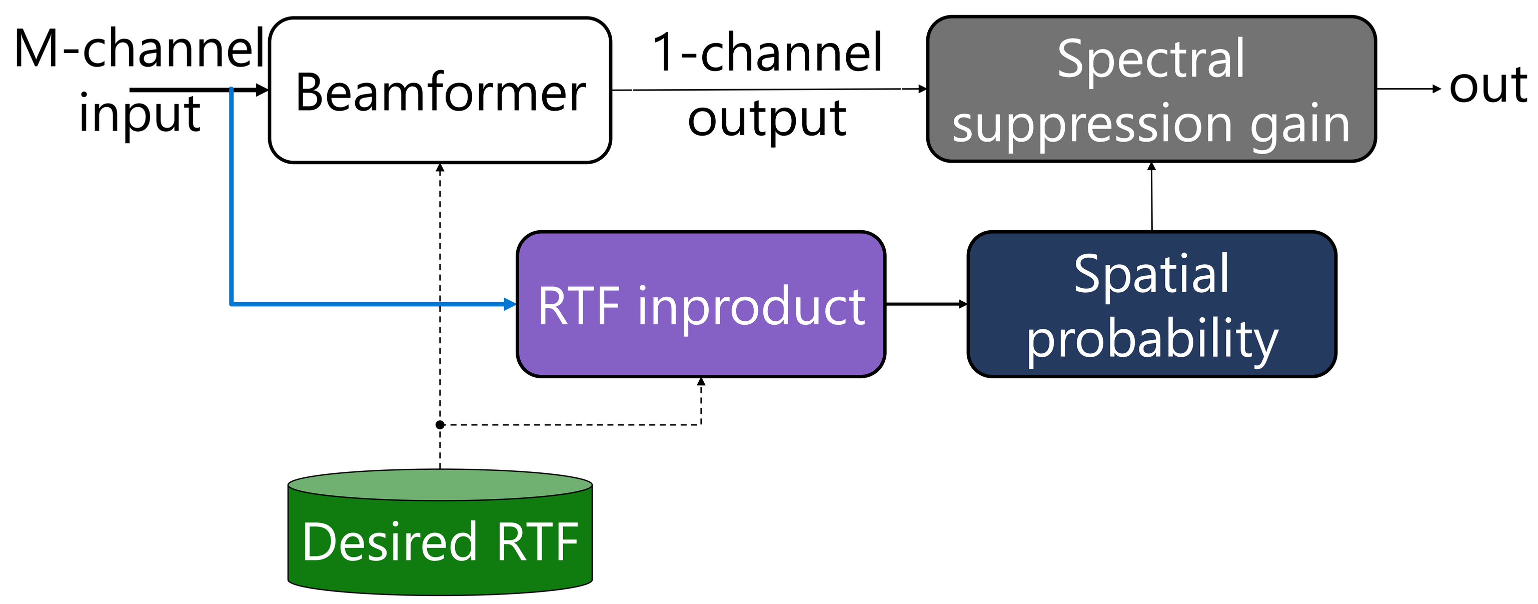
Desired signal

$$X_1(k,n) = \begin{cases} A_1(k,\mathbf{r}) S_\mathbf{r}(k,n) & \text{if} \quad \mathbf{r} = \mathbf{r}_\mathrm{d} \\ 0 & \text{if} \quad \mathbf{r} \neq \mathbf{r}_\mathrm{d} \end{cases}$$

Relative transfer function (RTF)

$$B_{m,1}(k,n) = \frac{A_m(k,\mathbf{r})}{A_1(k,\mathbf{r})} \approx \frac{\mathrm{E}\left\{Y_m(k,n) Y_1^*(k,n)\right\}}{\mathrm{E}\left\{|Y_1(k,n)|^2\right\}}$$

RTF-based Spatial Suppressor

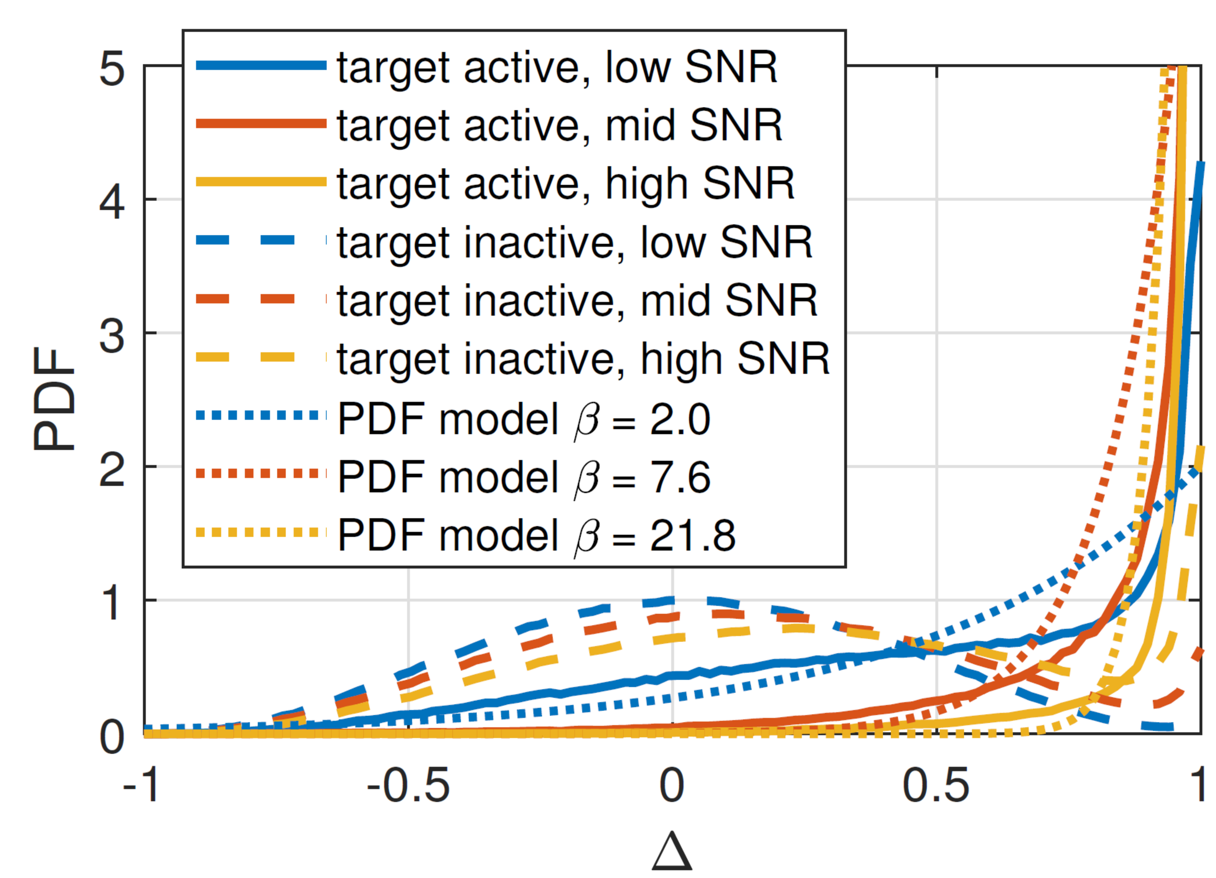


RTF inproduct a.k.a. cosine of Hermitian angle

$$\Delta = \cos\left\langle \mathbf{b}_\mathrm{d}(k), \widehat{\mathbf{b}}(k,n) \right\rangle = \frac{\Re\left\{\mathbf{b}_\mathrm{d}^H(k)\widehat{\mathbf{b}}(k,n)\right\}}{\|\mathbf{b}_\mathrm{d}(k)\|\, \|\widehat{\mathbf{b}}(k,n)\|}$$

Spatial probability of speech presence at $\mathbf{r}_\mathrm{d}$

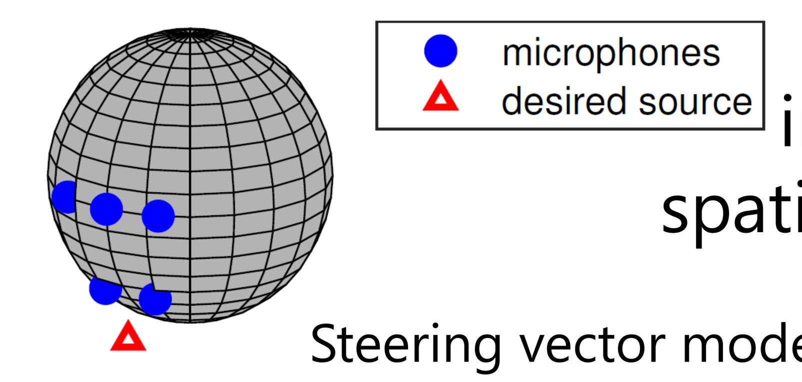
$$P(H_\mathrm{d}|\Delta) = \frac{P(H_\mathrm{d}) p(\Delta|H_\mathrm{d})}{P(H_0)p(\Delta|H_0) + P(H_\mathrm{d})p(\Delta|H_\mathrm{d})}$$



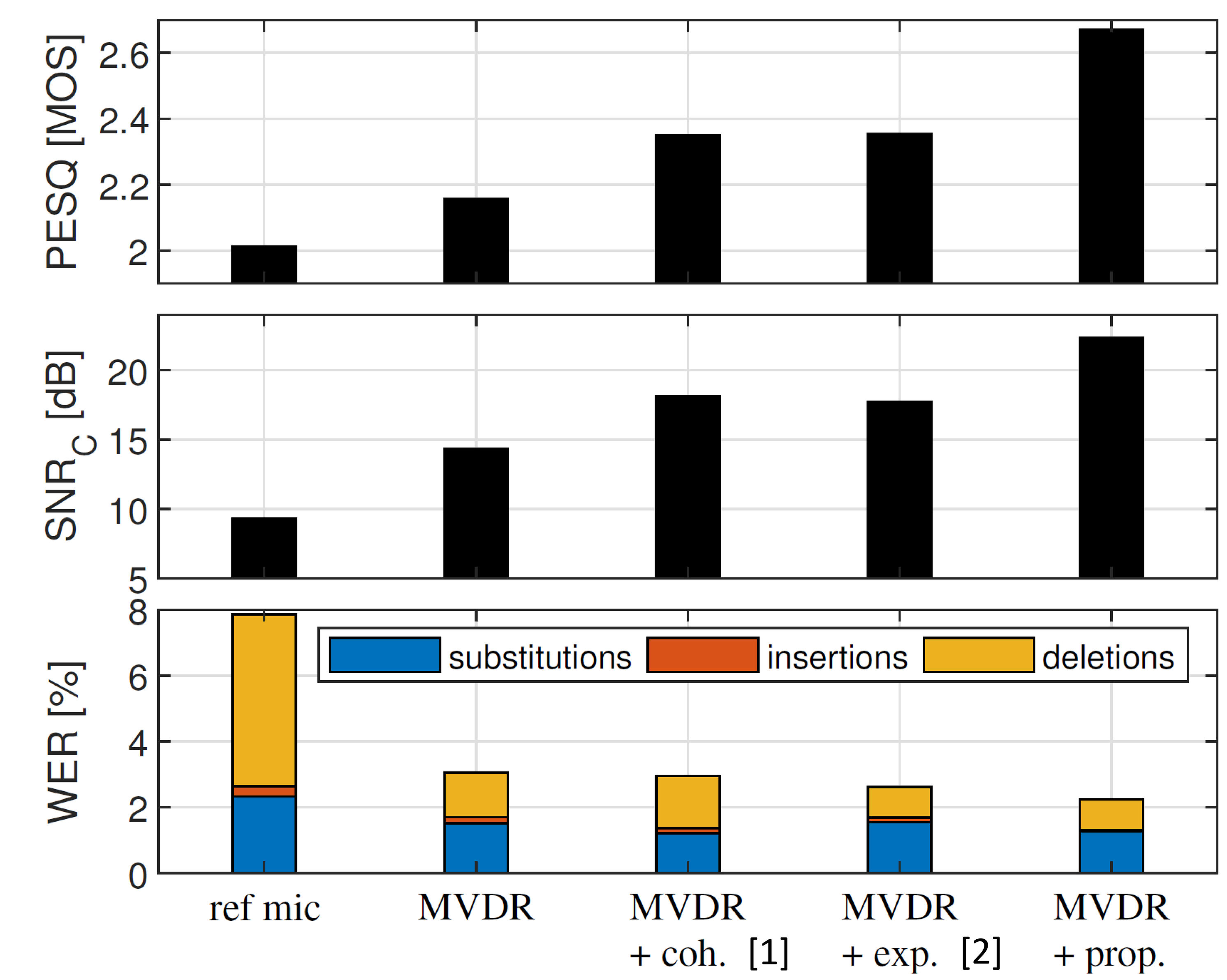
Fitted likelihoods models for spatial presence and absence $p(\Delta|H_\mathrm{d}),\ p(\Delta|H_0)$

Evaluation

Nearfield source extraction for HMD
user speech 80 dB
interfering talkers 90 dB
spatial noise 45 – 75 dB SPL



Steering vector model:
mics on rigid sphere



Conclusions

- Spatial presence probability using magnitude and phase (important in nearfield)
- No requirement for DOA estimator
- Efficient suppression of directional interfering talkers and non-directional noise
- Large reduction of ASR word insertions

[1] A. Kuklasinski, S. Doclo, S.H. Jensen, and J. Jensen, "Maximum likelihood based multi-channel isotropic reverberation reduction for hearing aids," in Proc. EUSIPCO, Sept. 2014, pp. 61–65.
[2] S. Jovicic and Z. Saric, "Adaptive microphone array free of the desired speaker cancellation combined with postfilter," J. Acoust. Soc. Am., vol. 123, no. 5, pp. 3739–3739, 2008.