Comparison of speech enhancement systems for noise fields in a car environment

Stefan Goetze, Volker Mildner, and Karl-Dirk Kammeyer

University of Bremen, Dept. of Communications Engineering, P.O. Box 330 440, D-28334 Bremen, Germany,

Email: goetze@ant.uni-bremen.de.de

Abstract

For telecommunications in a car environment hand-sets have to be replaced by hands-free units due to an increasing demand for a convenient and natural dialog and for safety in road traffic. The clean speech signal is heavily corrupted by noise from the engine, tire friction or the air conditioning, e.g. The noise field in a car has a strong low-pass characteristic and can be instationary due to changing driving speeds or opening the window, e.g.

Different multi-channel and single-channel speech enhancement algorithms will be compared for a car-environment in this contribution. The well-known Ephraim&Malah algorithm [1] in connection with Martin's Minimum Statistics [2] suppresses the noise without affecting the speech signal too much. Multi-channel combinations of beamformers and post-filters can exploit spacial information and are therefore independent of the statistics of the noise signal. In contrast many multi-channel post-filter algorithms suffer from poor noise reduction in the lower frequency ranges due to strongly correlated noise in the microphone channels.

Hybrid multi-channel noise reduction schemes have the ability to combine the advantages of both systems. In correlated frequency ranges single channel Wiener-Filters or the Ephraim&Malah-algorithm can be used while the multi-channel post-filters exploit spacial information for higher frequencies.

Noise fields in a car environment

In a car environment noise reduction systems have to cope with two major problems. The first one is the strong low-pass characteristic of the noise and secondly the noise field is diffuse. This means that for the low frequencies, where most of the energy of the noise has to be attenuated, the noise in different microphone channels is highly correlated. Thus it can not be suppressed by conventional multi-channel speech enhancement systems like fixed beamformers or post-filters. Figure 1 shows the power spectral density (PSD) $\Phi_{NN}[m]$ of the noise measured in a medium-sized vehicle. m is the discrete frequency index.

The noise field can be considered to be diffuse, as the magnitude squared coherence (MSC) can be approximated by a sinc function $MSC[m] = \Gamma^2[m] = \mathrm{si}^2(2\pi \cdot m \cdot d_{ij}/c)$ [3]. Figure 2 compares the theoretical MSCs with the ones calculated from the measurements in a car.

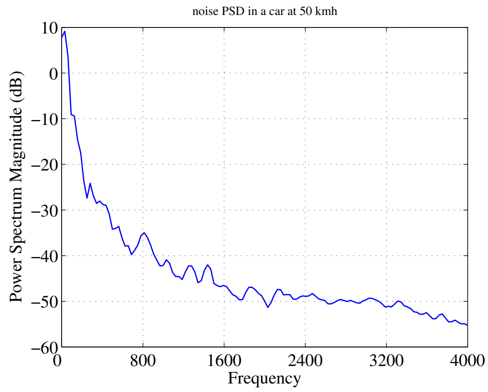

Figure 1: Noise power spectral density measured in a car

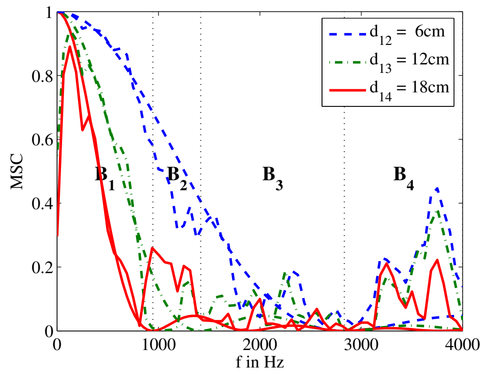

Figure 2: Magnitude squared coherence measured in a car and the theoretically calculated MSCs

Comparison of different post-filters

For this contribution we will compare four post-filter weighting functions for the noise environment in a car described above. The post-filters were compared for an office environment in [3, 4], where they were described in more detail. The Zelinski postfilter in equation (1) was the first multi-channel post-filter [5]:

$$W_Z[m] = \frac{\frac{2}{M(M-2)}\Re\left\{\sum_{i=1}^{M-1}\sum_{j=i+1}^{M} X_i^*[m]X_j[m]\right\}}{\frac{1}{M}\sum_{i=1}^{M} X_i^*[m]X_i[m]} \tag{1}$$

Simmer showed that the Zelinski post-filter suffers from overestimation of the noise and thus leads to signal cancelation. Therefore he modified the weighting rule as [6]

$$W_{SW}[m] = \frac{\frac{2}{M(M-2)}\Re\left\{\sum_{i=1}^{M-1}\sum_{j=i+1}^{M} X_i^*[m]X_j[m]\right\}}{Y^*[m]Y[m]}. \tag{2}$$

Since we have to deal with a diffuse noise field as shown in Figure 2, we defined subband filtering [3, 4], taking only uncorrelated microphone pairs into account:

$$W_{sub1}[m] = \frac{\frac{2}{t(t-2)}\Re\left\{\sum_{i=1}^{t-1}\sum_{j=i+M-t+1}^{M} X_i^*[m]X_j[m]\right\}}{\frac{1}{M}\sum_{i=1}^{M} X_i^*[m]X_i[m]} \quad (3)$$

$$W_{sub2}[m] = \frac{\frac{2}{t(t-2)}\Re\left\{\sum_{i=1}^{t-1}\sum_{j=i+M-t+1}^{M} X_i^*[m]X_j[m]\right\}}{Y^*[m]Y[m]} \quad (4)$$

$t = 1..M$ is the subband index. M is the number of microphones and $\Re\{\cdot\}$ the real part of a complex variable. We can see the lacking noise reduction ability for

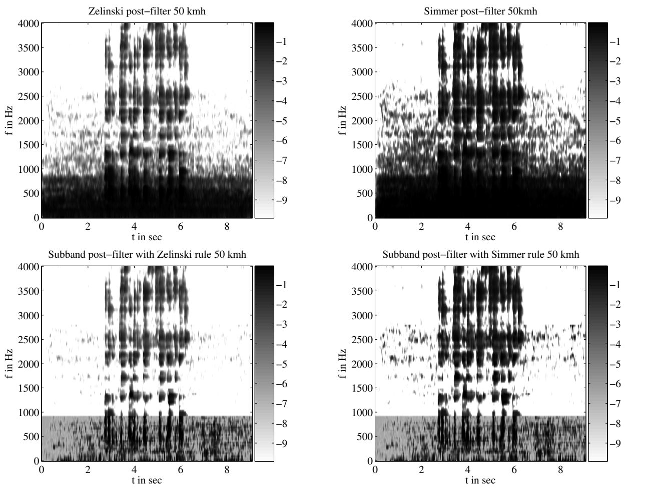

Figure 3: Post-filter weighting as a function of time and frequency

low frequency ranges for the first two post-filters $W_Z[m]$ and $W_{SW}[m]$ in Figure 3 (upper part). In the lower part, we see the transfer-functions of post-filter $W_{sub1}[m]$ and $W_{sub2}[m]$ with a single channel Ephraim&Malah algorithm working in the lowest subband B_1. Comparing the post-filters in the left and the right part of Figure 3 we realize that the the approaches $W_Z[m]$ and $W_{sub1}[m]$ introduce more signal degradation compared to the post-filters relying on Simmers approach.

Simulation Results and Conclusions

Figure 4 compares the post-filters by means of the noise reduction (NR), the SNR enhancement (SNRE) and the Perceptual Similarity Measure (PSM) from PEMO-Q [7] which takes the human auditory system into account and tries to estimate the speech quality. The noise reduction and SNRE plots qualitatively show the same results as Figure 3. Without the subband approach the reduction of the noise is insufficient due to the strong lowpass characteristic. The PSM shows that the subband approaches have more affect to the desired signal and thus lead to some signal cancelation because of the Ephraim&Malah filter for subband B_1.

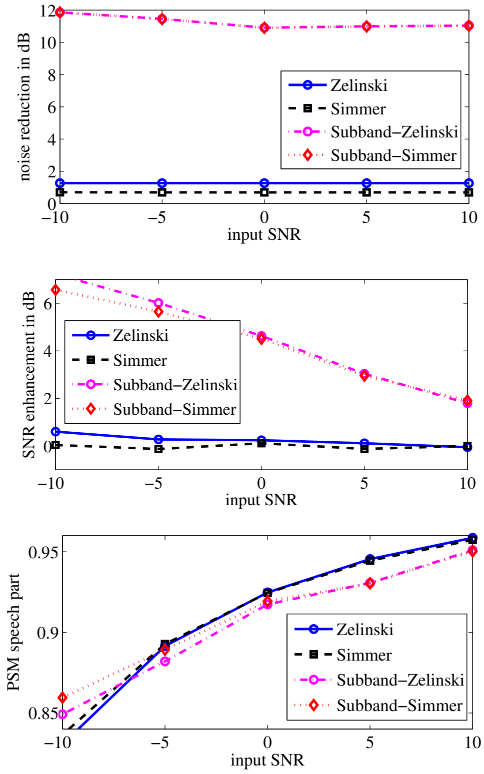

Figure 4: Comparison of the post-filters by means of noise reduction, SNR enhancement and perceptual similarity measure

References

[1] Y. Ephraim and D. Malah, “Speech enhancement using a minimum mean–square error log-spectral amplitude estimator,” *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 33, no. 2, pp. 443–445, 1985.

[2] R. Martin, “Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics,” *IEEE Trans. on Speech and Audio Processing*, vol. 9, no. , July 2001.

[3] S. Goetze, V. Mildner, and K.-D. Kammeyer, “A Psychoacoustic Noise Reduction Approach for Stereo Hands-Free Systems,” in *Audio Engineering Society (AES), 120th Convention*, Paris, France, 20.-23. May 2006.

[4] V. Mildner, S. Goetze, and K.-D. Kammeyer, “Multi-Channel Speech Enhancement using a Psychoacoustic Approach for a Post-Filter,” in *German ITG-Symposium on Speech Communication*, Kiel, Germany, 26.-28. April 2006.

[5] R. Zelinski, “A microphone array with adaptive post–filtering for noise reduction in reverberant rooms,” in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, New York City, New York, Apr. 1988, pp. 2578–2581.

[6] K. U. Simmer and A. Wasiljeff, “Adaptive microphone arrays for noise suppression in the frequency domain,” in *Second Cost 229 Workshop on Adaptive Algorithms in Communications*, Bordeaux, Frankreich, Oct. 1992, pp. 185–194.

[7] R. Huber, *Objective Assessment of Audio Quality Using an Auditory Processing Model*, Ph.D. thesis, University of Oldenburg, 2003.