Audio Engineering Society

Convention Paper

Presented at the 120th Convention
2006 May 20–23 Paris, France

A Psychoacoustic Noise Reduction Approach for Stereo Hands-Free Systems

Stefan Goetze, Volker Mildner, and Karl-Dirk Kammeyer

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

Correspondence should be addressed to Stefan Goetze (goetze@ant.uni-bremen.de)

ABSTRACT

One demand for comfortable high quality hands-free video conferencing systems is the transmission of a spatial acoustical impression. Therefore a major task is the transmission of stereo speech signals from a noisy environment. The suppression of the noise components must not corrupt the stereo effect. In this context different single channel, multi-channel and hybrid speech enhancement systems will be evaluated in this contribution. The problem of musical noise in post-filter-algorithms is addressed. Therefore a psychoacoustic masking threshold for the noise reduction algorithms is considered.

1. INTRODUCTION

Post-filtering or short time spectral attenuation (STSA) is a common enhancement technique for conventional beamformers. Zelinski [1] proposed the first multichannel noise reduction post-filter which considers statistically independent noise components in the different microphone paths. Unfortunately this assumption is not fulfilled for practical environments [2, 3] and therefore this post-filter leads to insufficient noise reduction, especially in the low frequency ranges where the noise field is highly correlated.

Furthermore the Zelinski design rule leads to an amount of signal distortion due to violation of the Wiener criterion since the filter function is calculated in dependence of the beamformer's input signals but is applied to the beamformer's output. Bitzer and Simmer developed post-filter structures with less signal distortion which are calculated by the power spectral densities from the beamformer's input signals as well as the beamformer's output signals [3, 4].

The insufficient suppression of the noise signal in the low frequency ranges can be improved by a subband approach [4, 5]. In practical environments the noise field can often be considered to be diffuse. By exploiting this assumption the frequency region can be divided into subbands where the noise signals are either correlated or uncorrelated depending on the intermicrophone distance only. For every intermicrophone distance in the array a subband can be defined by the analysis of the magnitude squared coherence (MSC) function. Highly correlated microphone pairs can be elided from the averaging of the spectral densities. For regions where the noise is correlated in all microphone pairs, to be precise in the lowest subband, a single channel approach can be applied, e.g. a Wiener Filter or algorithms which take statistical or psychoacoustic aspects into account. In this contribution the Ephraim&Malah [6] algorithm in combination with Martin's Minimum Statistics [7] will be used as an example for a statistically motivated algorithm. The psychoacoustic approach by Gustafsson [8], which exploits the masking threshold of the speech signal, will be evaluated as well.

The psychoacoustic filter design rule introduced by Gustafsson leads to a single channel noise reduction filter which has the advantage of not suffering from the musical noise problem during speech pauses. The masking threshold is calculated as a function of the clean speech signal's power spectral density, which has to be obtained by prefiltering of the noisy signal. The multi-channel post-filter for the higher frequency regions and the single-channel Ephraim&Malah algorithm for the lowest subband are combined to form a hybrid post-filter. It exploits as well spacial information in the higher subbands as statistical information in frequency regions where the noise field is correlated. The problem of musical noise is reduced by the use of spectral masking.

Microphone arrays can be employed for a stereo setup, but the acoustic scenario must not be corrupted by the signal processing. Under the circumstance of limited space for the microphone array two overlapping subarrays can be applied to create the stereo signals for transmission.

The comparison of the different multi-channel, single-channel and hybrid noise reduction schemes will be presented in this paper especially under consideration of the distortion of the desired signal and the overall speech quality for a stereophonic setup.

The remainder of this paper ist organized as follows: Section 2 reviews the conventional Zelinski post-filter and the modifications of Bitzer and Simmer concerning the correlated noise and the over-estimation of the noise power spectral density (PSD). Section 3 introduces the psychoacoustically motivated weighting rule after Gustafsson. In section 4 the stereo system model is explained and in section 5 the objective measures used in this contribution to evaluate and compare the presented design rules are introduced. Section 6 presents our simulation results and a final conclusion is given in section 7.

2. POST-FILTERING FOR MULTI-CHANNEL NOISE REDUCTION

The first multi-channel post-filter was proposed by Zelinski [1]. Figure 1 depicts a generalized scheme which includes the Zelinski post-filter. The signal in each microphone path $x_i[k] = s_i[k] + n_i[k]$ consists of the desired signal $s_i[k]$ and an additive noise $n_i[k]$. k is the discrete-time index and $i = 0..M$ the channel index. After the time delay compensation (TDC) the system is steered towards the desired source.

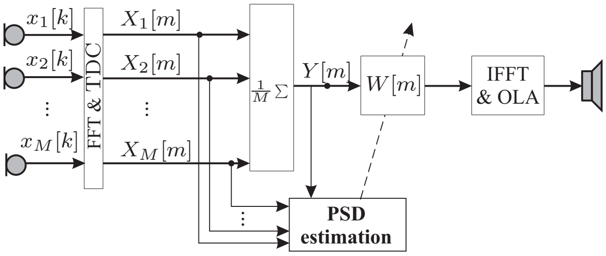

Figure 1: Multi-channel post-filter design rule, e.g. the Zelinski post-filter after (1)

The transfer function of the Zelinski post-filter in the frequency domain is given by

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

$\Re\{\cdot\}$ is the real part of a complex variable, $(\cdot)^*$ is the conjugate complex and m is the discrete frequency index.

With the assumption

- that the speech signal $S[m]$ and the noise signals $N_i[m]$ are uncorrelated for each microphone path $\mathrm{E}\{S^*[m]N_i[m]\} = 0, i = 1..M$ and with the assumption of
- mutually uncorrelated noise signals $\mathrm{E}\{N_i^*[m]N_j[m]\} = 0$, $\forall i \neq j$.

the Zelinski-rule can be considered to fulfill the Wiener criterion

$$W_W[m] = \frac{\Phi_{SS}[m]}{\Phi_{SS}[m] + \Phi_{NN}[m]}. \quad (2)$$

$\Phi_{SS}[m] = \mathrm{E}\{S^*[m]S[m]\}$ and $\Phi_{NN}[m] = \mathrm{E}\{N^*[m]N[m]\}$ are the (auto) power spectral densities of the speech signal and the noise signal respectively. $\mathrm{E}\{\cdot\}$ is the expectation operator.

There are two major problems in the argumentation above. Firstly the noise in the microphone channels is seldom uncorrelated and secondly equation (2) disregards the noise reduction of the beamformer and thus leads to an overestimation of the noise. Two approaches to overcome these problems are discussed in section 2.1 and 2.2.

2.1. Subband-filtering

For practical environments the assumption of an uncorrelated noise field does not hold. As shown in [2, 5] for example, the noise field often can be assumed to be diffuse. For a diffuse noise field the MSC

$$MSC = \Gamma^2_{X_iX_j}[m] = \frac{\left|\Phi_{X_iX_j}[m]\right|^2}{\Phi_{X_iX_i}[m]\Phi_{X_jX_j}[m]}, \quad (3)$$

which is a common measure for the description of the noise field, can be calculated by the $\mathrm{si}^2(\cdot)$-function, which depends on the intermicrophone spacing only.

$$\Gamma^2_{X_iX_j}[m] = \mathrm{si}^2(2\pi \cdot m \cdot d_{ij}/c) \quad (4)$$

$\mathrm{si}(x) = \sin(x)/x$ is called the sinc-function. c is the speed of sound and d_{ij} is the distance between the microphones i and j. From equation (4) we see, that the noise field is highly correlated for low frequencies and has a low correlation for high frequencies. Figure 2 shows the theoretical MSC after (4) for a microphone distance of 16 cm and the calculated MSC after (3) for the office environment which is described in section 4.

We see from Figure 2 that the assumption of a diffuse noise field is fulfilled, even if we take into account that only one noise source is present in our simulations (see left part of Figure 7). For an increased number of noise sources the MSC in the regions above 900 Hz would decrease further.

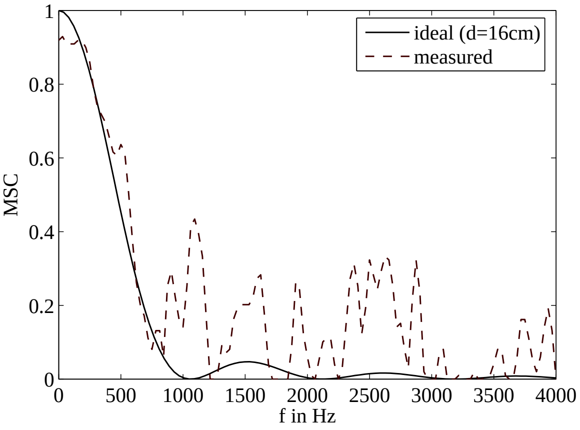

Figure 2: Theoretical MSC between the microphone channels for a microphone distance of 16 cm and the measures MSC for the room described in section 4

As proposed in [9] and relying on [5] we now define frequency subbands B_t with bandlimits at the first zero of the MSC:

$$f_{t,\{i,j\}} = \frac{c}{2d_{ij}}, \quad \text{for } 1 \leq t \leq M \quad (5)$$

t is the number of the subband. The resulting bandlimits and the corresponding MSCs are depicted in Figure 3 for a linear array consisting of $M = 4$ microphones and a intermicrophone distance of 8 cm. The uncorrelated microphone pairs are written down in Table 1.

Table 1: Subbands B_t and the corresponding uncorrelated microphone pairs in a diffuse noise field for a microphone spacing d_{12}=8cm and M=4

Subband	uncorrelated microphone pairs
$B_1 = 0..708$Hz	no uncorrelated microphone pairs
$B_2 = 708..1063$Hz	{1,4}
$B_3 = 1063..2125$Hz	{1,3},{1,4},{2,4}
$B_4 = 2125..4000$Hz	{1,2},{1,3},{1,4},{2,3},{2,4},{3,4}

As we see from Table 1, the condition of an uncorrelated noise field, which is the underlying assumption of Zelinski's post-filter in equation (1) is only true for the highest subband. Therefore the Zelinski post-filter hardly achieves noise suppression for low frequencies.

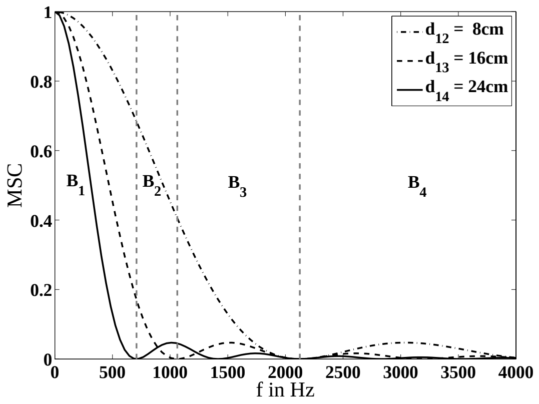

Figure 3: Theoretical MSC between the microphone channels for a microphone distance of 8 cm and the resulting subbands ($M = 4$)

To overcome this problem we define a subband post-filter which only takes uncorrelated cross power spectral densities after Table 1 into account for the estimate of the power spectral density of the speech:

$$W_{sub1}[m] = \frac{\frac{2}{t(t-2)}\Re\left\{\sum_{i=1}^{t-1}\sum_{j=i+M-t+1}^{M} X_i^*[m]X_j[m]\right\}}{\frac{1}{M}\sum_{i=1}^{M} X_i^*[m]X_i[m]} \tag{6}$$

$t = 1..M$ is the number of the subband. It should be mentioned that the post-filter rule in equation (6) can not be calculated for the lowest subband B_1, because no uncorrelated microphone pairs are available (see Table 1). Therefore a single channel post-filtering rule has to be chosen for subband B_1. Spectral subtraction was used in [2] and a Wiener-Filter with a decision directed approach in [5]. As in [9] we apply the log-STSA weighting rule of Ephraim&Malah in conjunction with Martin's Minimum Statistics [7].

2.2. Overestimation of the noise

The second drawback of Zelinski's design rule (1) which is the reason that it does not meet the Wiener criterion (2) for the complete noise reduction system is the disregard of the beamformers influence to the noise. As shown in [3] the optimum weighting function for a post-filter after a beamformer depends on the coherence function of the noise field:

$$W_{opt}[m] = \frac{\Phi_{SS}[m]}{\Phi_{SS}[m] + \frac{1}{M}\Phi_{NN}[m]\left(1 + \tilde{\Gamma}[m]\right)} \tag{7}$$

with

$$\tilde{\Gamma}[m] = \frac{2}{M}\sum_{i=1}^{M-1}\sum_{j=i+1}^{M}\Re\{\Gamma_{N_iN_j}[m]\} \tag{8}$$

and the complex coherence function of the noise

$$\Gamma_{N_iN_j}[m] = \frac{\Phi_{N_iN_j}[m]}{\sqrt{\Phi_{N_iN_i}[m]\Phi_{N_jN_j}[m]}}. \tag{9}$$

For uncorrelated regions of the noise field ($\tilde{\Gamma}[m] \approx 0$ for $f > c/2d_{ij}$), equation (7) simplifies to

$$W_{opt}[m] = \frac{\Phi_{SS}[m]}{\Phi_{SS}[m] + \frac{1}{M}\Phi_{NN}[m]}. \tag{10}$$

As equation (10) shows, the PSD of the noise is over-estimated by a factor of M compared to the Wiener criterion in equation (2). Therefore Simmer proposed to replace the denominator of (1) by the the beamformers output [3]:

$$W_{SW}[m] = \frac{\frac{2}{M(M-2)}\Re\left\{\sum_{i=1}^{M-1}\sum_{j=i+1}^{M} X_i^*[m]X_j[m]\right\}}{Y^*[m]Y[m]} \tag{11}$$

For $f > c/2d_{ij}$ the weighting rule (11) meets the Wiener criterion for a post-filter following a beamformer (10). It should be mentioned that equation (11) does not solve the problem of insufficient noise reduction for frequencies below $f > c/2d_{ij}$. Therefore we modify our subband approach from section 2.1:

$$W_{sub2}[m] = \frac{\frac{2}{t(t-2)}\Re\left\{\sum_{i=1}^{t-1}\sum_{j=i+M-t+1}^{M} X_i^*[m]X_j[m]\right\}}{Y^*[m]Y[m]}. \tag{12}$$

3. THE PSYCHOACOUSTICALLY MOTIVATED NOISE REDUCTION

For speech enhancement purposes, achieving a maximum of noise reduction or the Wiener criterion, which is the optimal tradeoff between noise reduction and speech distortion, is not necessarily the optimum design method. If the speech signal, which was enhanced by a noise reduction algorithm, is presented to a human listener, these conventional mathematical minimization criteria needn't be the best choice. For this scenario an improvement of speech intelligibility or speech quality could be more appropriate. Therefore Gustafsson proposed a psychoacoustic weighting rule based on the auditory masking effect in [8] which is reviewed in the following.

3.1. Auditory masking

Exploiting the auditory masking effect is widely used in audio coding [10]. In presence of an acoustic stimulus (such as a sinusoidal waveform or a narrowband noise signal) the absolute threshold of hearing is raised for adjacent frequency regions. The absolute threshold of hearing is the minimum sound pressure which is necessary for the human auditory system to perceive a signal for a given frequency or more precisely to cause a sufficient excitation of the nerves of the cochlea.

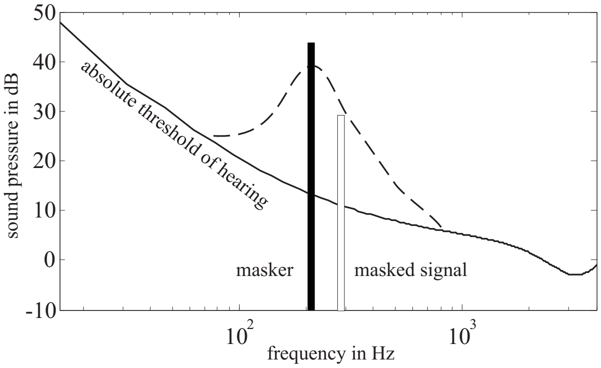

Figure 4: Rising of the absolute threshold of hearing caused by a masker

The effect of auditory masking is illustrated in Figure 4. The "masker" rises the absolute threshold of hearing and the "masked signal" becomes inaudible for the human auditory system. If we transform the frequency axis into the Bark domain [10] the so called spreading function (the dashed line in Figure 4) can be approximated by a triangle or the somewhat sophisticated function of [10]. Figure 5 shows the speech power spectral density $\Phi_{SS}(f)$ and the calculated masking threshold $\Phi_{TT}(f)$ as a function of the frequency. The masking threshold can be calculated as a function of the clean speech signal after [10]

$$\Phi_{TT}[m] = f(\Phi_{SS}[m]). \tag{13}$$

Thus an initial estimate for $\Phi_{SS}[m]$ has to be calculated by prefiltering the noisy signal. Gustafsson used the Ephraim&Malah weighting rule. It should be noted that if noise would be present at about 1100 Hz or 2900 Hz for example, it needn't be attenuated below the masking threshold, because it would be inaudible for a human listener anyway. Based on this assumption, Gustafsson developed his weighting rule trying to preserve the characteristics of the noise.

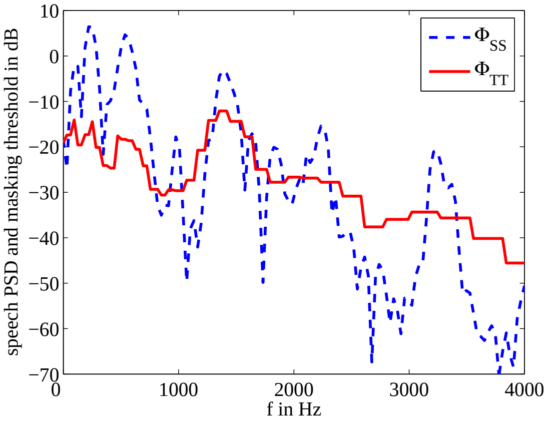

Figure 5: Speech PSD $\Phi_{SS}(f)$ and masking threshold $\Phi_{TT}(f)$

3.2. Psychoacoustically motivated weighting rule

As mentioned before a complete removal of the noise is neither necessary nor desirable. We define the desired signal at the output of the speech enhancement algorithm $\breve{S}[m]$, that should consist of the clean speech signal $S[m]$ and a constantly attenuated noise $N[m]$. Thus the spectral characteristics of the noise should be preserved, but with an lowered amplitude.

$$\breve{S}[m] = S[m] + \zeta_N N[m] \tag{14}$$

$\zeta_N < 1$ is defined as the noise attenuation factor. Since a spectral weighting has to be applied to the noisy signal $X[m] = S[m] + N[m]$, the actual output of the filter is

$$\hat{S}[m] = H[m] \left(S[m] + N[m] \right). \tag{15}$$

The quadratic error between (14) and (15)

$$\Phi_{QQ}[m] = \mathrm{E}\left\{ \left| \breve{S}[m] - \hat{S}[m] \right|^2 \right\} \tag{16}$$

can be decomposed into the distortion of the speech part $\Phi_{Q_SQ_S}[m]$ and the "distortion" of the noise part $\Phi_{Q_NQ_N}[m]$ respectively [8], which both are quadratic functions of $H[m]$:

$$\Phi_{QQ}[m] = \Phi_{Q_SQ_S}[m] + \Phi_{Q_NQ_N}[m] \tag{17}$$

$$\Phi_{Q_SQ_S}[m] = (1 - H[m])^2 \, \Phi_{SS}[m] \tag{18}$$

$$\Phi_{Q_NQ_N}[m] = (\zeta_N - H[m])^2 \, \Phi_{NN}[m] \tag{19}$$

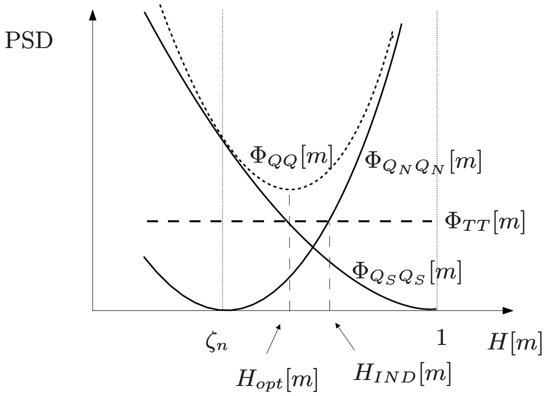

Figure 6: Distortion of the speech component $\Phi_{Q_SQ_S}[m]$ and "distortion" of the noise component $\Phi_{Q_NQ_N}[m]$ in dependance of the filter coefficient $H[m]$ and the masking threshold $\Phi_{TT}[m]$

Figure 6 shows the influence of different minimization criteria on $\Phi_{Q_SQ_S}[m]$ and $\Phi_{Q_NQ_N}[m]$. The minimization of the speech distortion would be achieved by $H[m] = 1$ but this would not reduce the noise. The minimization of $\Phi_{Q_NQ_N}[m]$ which could be interpreted as the power of the difference of the desired noise amplitude and the actual noise amplitude would lead to $H[m] = \zeta_N$. Thus the actual filter coefficient has to be $\zeta_N \leq H[m] \leq 1$.

Since the masking of the total distortion $\Phi_{QQ}[m]$ can not be guarantied, because $\Phi_{TT}[m]$ is often lower than $\Phi_{QQ}[m]$, the masking of $R_{Q_NQ_N}$ was chosen for the filter design [8], which means that the perceived noise will be constant.

$$\Phi_{Q_NQ_N}[m] = (\zeta_N - H[m])^2 \, \Phi_{NN}[m] \overset{!}{=} \Phi_{TT}[m] \tag{20}$$

With a constraining to values smaller than 1 the weighting rule $W_{IND}[m]$ after [8] can be calculated as

$$W_{IND}[m] = \min\left(\sqrt{\frac{\Phi_{TT}[m]}{\Phi_{NN}[m]}} + \zeta_n, \ 1 \right) \tag{21}$$

IND stands for *inaudible noise distortion.*

4. THE STEREO SYSTEM

The stereophonic office environment used in this contribution is shown in Figure 7. Two overlapping subarrays of 4 microphones for each output channel are formed by grouping from the 6 available microphones. Thus the two microphones in the middle position can be reused for the left and the right channel. The dashed lines in the left part of Figure 7 indicate the steering of the subarrays towards the desired source which is achieved by proper delays in the TDC-block in the right part of Figure 7. The microphone signals $X_i[m]$ are calculated by convolution of the clean speech signal $S[m]$ (male speaker) and the noise source $N[m]$ with simulated impulse responses after [11]. The room reverberation time is $\tau_{60} = 400\text{ms}$ and the sampling frequency $f_s = 8000\text{Hz}$. The sampled time signal was block-processed with a block length of $L_{\mathrm{Bl}} = 128$, an overlapping of 50%, and a Hann-window was applied to each block. After processing by the frequency-domain speech-enhancement-system the signal was synthesized by the overlap-add-method (OLA) [12] as depicted in Figure 7.

5. OBJECTIVE MEASURES

The algorithms, which were presented in this contribution, all belong to the class of STSA-algorithms, which try to reconstruct the spectral envelope of the desired signal by a suitable spectral weighting. They will be compared by means of the widely used measures of noise reduction (NR) and signal to noise ratio enhancement (SNRE) and by the perceptual similarity measure (PSM) which takes the auditory system into account [13, 14].

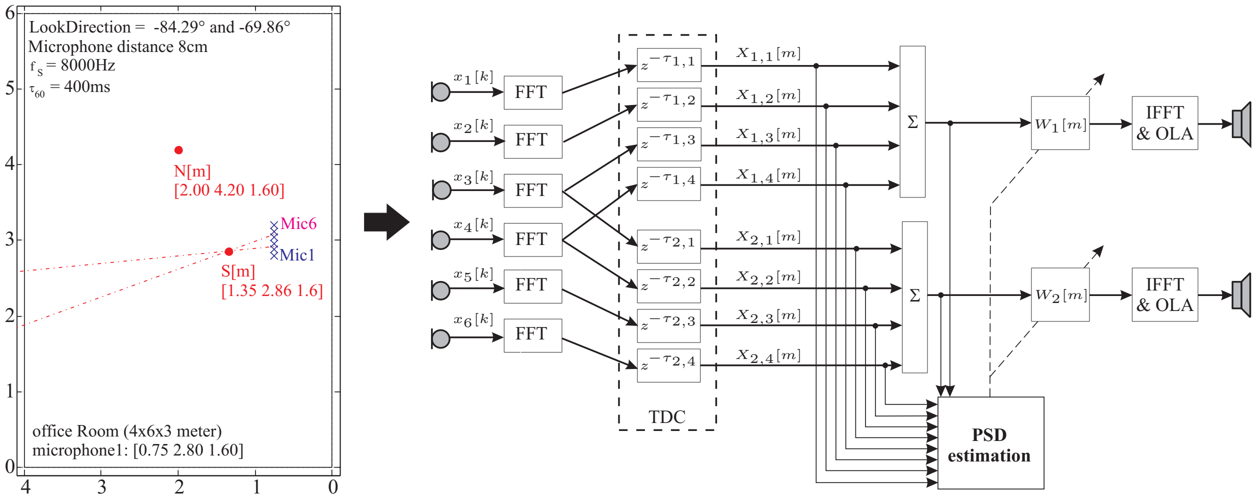

Figure 7: Block diagram of the stereo noise reduction system

The noise reduction is defined as

$$NR = \frac{1}{|\mathcal{A}|} \sum_{\ell \in \mathcal{A}} 10 \cdot \log_{10} \left(\frac{\sum_{k=1}^{K} x^2[\ell K + k]}{\sum_{k=1}^{K} \hat{s}^2[\ell K + k]} \right) \qquad (22)$$

with the block length K, the noisy speech signal at the input of the analysed algorithm $x[k]$ and the enhanced signal after the algorithm $\hat{s}[k]$, which should be an estimate of the desired signal. $\mathcal{A}$ represents the set of frames where speech is absent and $|\mathcal{A}|$ its cardinality. The NR measure does not necessarily indicate a good speech quality because it does not take a possible cancelation of the desired signal into account.

Therefore we also evaluate the presented algorithms in terms of the SNRE

$$\text{SNRE} = \text{SSNR}_{out} - \text{SSNR}_{in} \qquad (23)$$

with SSNR_{in} and SSNR_{out} as the segmental signal to noise ratio (SSNR) at the input and the output of the tested algorithm respectively:

$$\text{SSNR} = \frac{1}{|\mathcal{B}|} \sum_{\ell \in \mathcal{B}} 10 \cdot \log_{10} \left(\frac{\sum_{k=1}^{K} s^2[\ell K + k]}{\sum_{k=1}^{K} n^2[\ell K + k]} \right) \qquad (24)$$

$\mathcal{B}$ is the set of frames where speech is present and $|\mathcal{B}|$ its cardinality.

The technical measures presented above, like most of the other commonly used measures, do not provide information whether the remaining noise or a distortion introduced by the algorithm is disturbing for a human listener. These technical measures do not provide sufficient information about the perceivability of the distortion [14]. Therefore we furthermore use the perceptual quality measure PSM from PEMO-Q, which is based on a model of the auditory system of a human listener [13].

6. SIMULATION RESULTS

The signals used for the simulations are depicted in Figure 8. The desired speech signal was a male speaker and speech shaped noise with slightly instationary (pulsing) components was used.

6.1. Subband-filtering

As we can see from Figure 9 the Zelinski design rule in equation (1) is not able to suppress noise in the low frequency regions. Comparing with the subband-limits in Table 1 we can identify the subband limits at 700 Hz, 1000 Hz and 2000 Hz in the upper part of Figure 9 where the corresponding sinc-function has its first minimum.

The same is true for the Simmer design rule in equation (11) which is illustrated in the middle part of

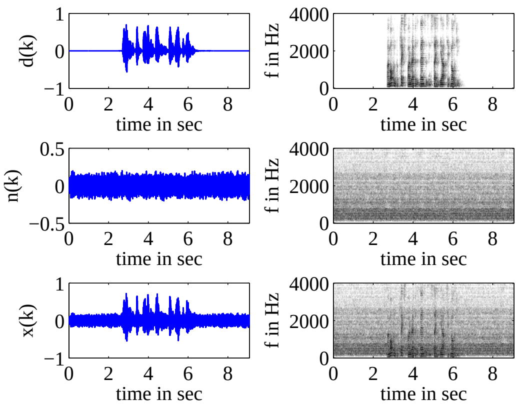

Figure 8: Signals used for simulation: desired speaker $s[k]$, noise signal $n[k]$ and microphone signal $x[k]$ at the first microphone and the related PSDs

Figure 9. Even worse, the filter is close to 1 for the low frequency regions. For this reason the Zelinski filter is often preferred by different authors. However we have to keep in mind, that the higher noise reduction of the Zelinski filter is due to an overestimation of the noise. This means that the cost for the better noise reduction is a higher speech distortion, which reduces speech intelligibility.

The lower subplot shows the modified Zelinski design rule in subbands after equation (6). It can be realized, that this post-filter performs better in the subbands B_2 and B_3. Furthermore the Ephraim&Malah rule removes correlated noise parts in the lowest subband.

6.2. Comparison of the weighting rules

Figure 10 and Figure 11 compare the different weighting rules:

- W_Z: The Zelinski post-filter after equation (1)
- W_{sub1}: The subarray post-filter based on the Zelinski weighting rule after equation (6)
- E&M after BF: An Ephraim&Malah weighting after a Delay&Sum beamformer
- W_{SW}: The Simmer weighting rule after equation (11)
- W_{sub2}: The subarray post-filter based on the Simmer weighting rule after equation (12)
- W_{IND}: The psychoacoustic motivated weighting rule after (21) with prefiltering by (12) and an estimate of the noise PSD by means of Minimum Statistics

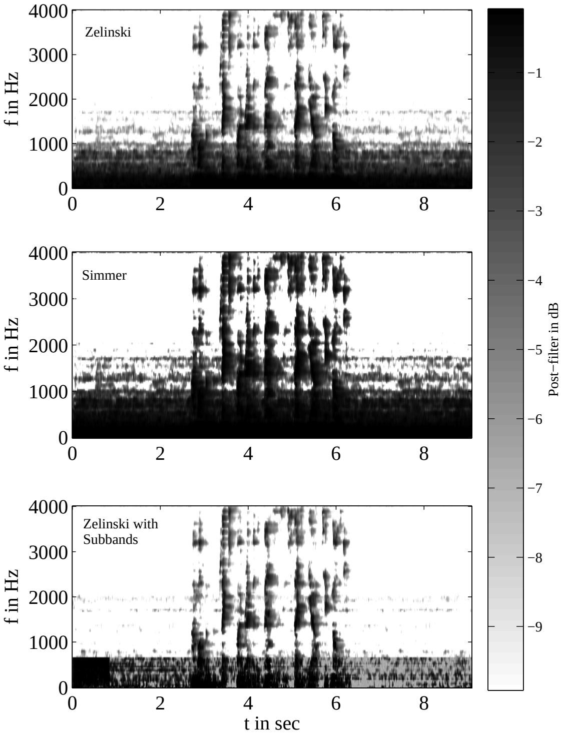

Figure 9: Post-filter weighting rules for the Zelinski-filter (1), the Simmer-filter (11) and the Zelinski post-filter with subbands (6). Input SNR is 0dB

The measures are averaged for the left and the right channel of the stereo system. If we take a look at the traditional measures NR and SNRE we realize the poor noise reduction and SNR enhancement of the Zelinski- and Simmer weighting functions due to their missing ability to reduce the correlated noise in the lower frequency regions. The noise has a lowpass characteristic and therefore the subarray-approaches show improved performance. They hardly differ from each other concerning the NR and SNR enhancement. The psychoacoustic weighting rule shows by far the best results.

The PSM curves in Figure 11 are separately calcu-

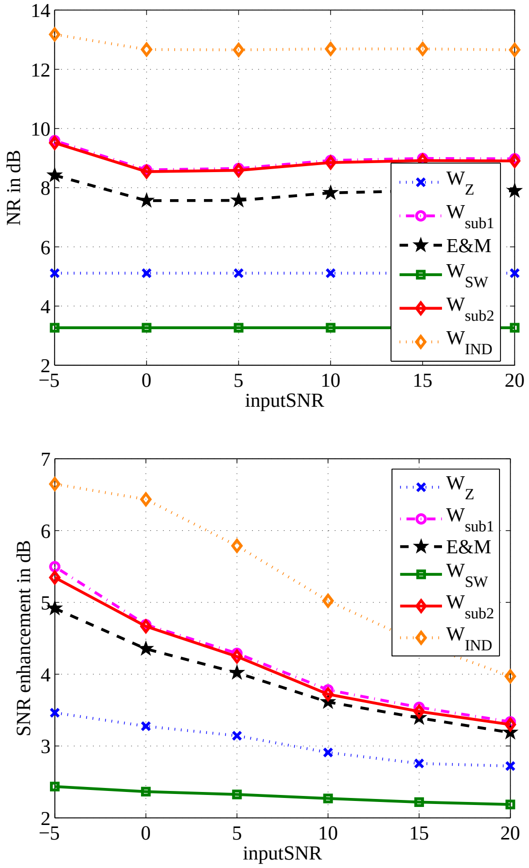

Figure 10: Comparison of the different weighting rules by means of NR and SNRE enhancement

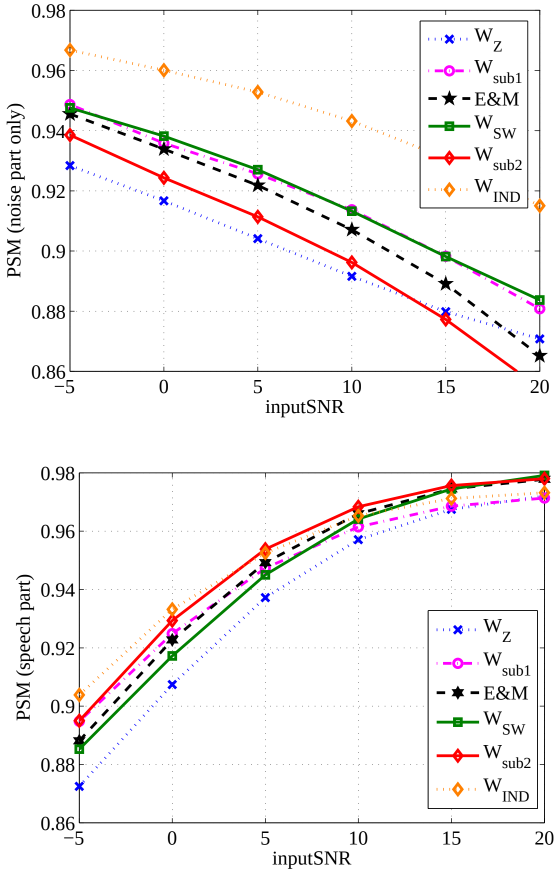

Figure 11: Comparison of the different weighting rules by means of the perceptual similarity measure (PSM)

lated for parts with speech activity and for parts which contain only noise. The area of speech activity is from 3 to 6 seconds as you can see from the signal plots in Figure 8 and the part from 6.5 to 9 seconds was chosen as noise only. For the noise only part the psychoacoustic motivated weighting rule W_{IND} clearly outperforms the other weighting rules which is due to the design criterion that tries to preserve the noise characteristics. Also for the speech part W_{IND} shows very good performance.

If we compare the Zelinski post-filter and the Simmer post-filter by means of PSM, we see that we get better results for the Simmer post-filter. The reason for this is the overestimation of the noise by the Zelinski filter which was explained in section 2.2. On the other hand the Simmer filter leads to a lower noise reduction and SNR enhancement.

Comparing the original design rules of Zelinski and Simmer with their subband extensions we recognize that the subband approaches perform better.

The filter rules which were compared here regarding speech and noise quality were compared in [15] for speaker recognition.

The best design rule for all objective measures used in this contribution was the new multi-channel psychoacoustic filter, which extends the approach of Gustafsson by a multi-microphone system to exploit spacial information.

7. CONCLUSIONS

Different multi-channel noise reduction approaches for stereo-speech enhancement were compared in this contribution. The psychoacoustically motivated weighting rule after Gustafsson was extended by a multi-channel case with prefilter which is able to exploit as well spatial information for incoherent frequency regions as statistical information for regions with highly correlated noise. The psychoacoustic filter outperforms all other weighting rules by means of NR, SNRE and PSM which is based on the human auditory system. Furthermore it does not suffer from the musical noise phenomenon.

8. REFERENCES

[1] 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.

[2] J. Meyer (Bitzer) and K. U. Simmer, "Multichannel speech enhancement in a car environment using Wiener filtering and spectral subtraction," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, München, Deutschland, Apr. 1997, pp. 1167–1170.

[3] 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.

[4] J. Bitzer, K. U. Simmer, and K. D. Kammeyer, "Multi-microphone noise reduction by post-filter and superdirective beamformer," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Pocono Manor, Pennsylvania, Sept. 1999, pp. 100–103.

[5] J. Li and M. Akagi, "A Hybrid Microphone Array Post-Filter in a Diffuse Noise Field," in *Proc. Eurospeech 2005*, Lisbon, Portugal, September 2005.

[6] 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.

[7] 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.

[8] S. Gustafsson, *Enhancement of Audio Signals by Combined Acoustic Echo Cancellation and Noise Reduction*, Ph.D. thesis, Aachen University of Technology, Wissenschaftsverlag Mainz, Aachen, June 1999, Aachener Beiträge zu digitalen Nachrichtensystemen, Band 11.

[9] 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.

[10] International Organization for Standardization, *Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 MBit/s, Audio Part (11172-3)*, Nov. 1992.

[11] J. B. Allen and D. A. Berkley, "Image Method for Efficiently Simulating Small–Room Acoustics," *J. Acoust. Soc. Amer.*, vol. 65, pp. 943–950, 1979.

[12] L.R. Rabiner and R.W. Schafer, *Digital Processing of Speech Signals*, Englewood Cliffs, Prentice Hall, 1978.

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

[14] T. Rohdenburg, V. Hohmann, and B. Kollmeier, "Objective Measures for the Evaluation of Noise Reduction Schemes," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, 2005.

[15] V. Mildner, S. Goetze, and K.-D. Kammeyer, "Multi-Channel Noise-Reduction-Systems for Speaker Identification in an Automotive Acoustic Environment," in *Audio Engineering Society (AES), 120th Convention*, Paris, France, 20.-23. May 2006.

The authors would like to thank Dr. Rainer Huber from Hoertech, Oldenburg as well as Thomas Rohdenburg and Stefan Strahl from the University of Oldenburg for the assistance with the PSM calculations.