NOISE SUPPRESSION WITH ADAPTIVE ADJUSTMENT OF THE MAXIMUM ATTENUATION

Mohamed Krini, Gerhard Schmidt

Harman/Becker Automotive Systems
Acoustic Speech Enhancement – Research
Söflinger Str. 100, 89077 Ulm, Germany
Email: mkrini@harmanbecker.com

ABSTRACT

In this paper a noise suppression method with adaptive adjustment of the maximum allowed signal attenuation is presented. In contrast to conventional noise reduction schemes, where often a so-called *spectral floor* is specified, we start with the specification of a desired residual background noise in terms of its power spectral density (PSD). Then the maximum filter attenuation of the proposed method is dynamically adjusted such that the PSD of the residual noise matches that of the desired background noise. Since a straight forward approach for achieving this objective results in an unnatural residual noise signal, three extensions that overcome these difficulties are presented. These new methods, which are particularly suitable as post-processing for beamformers, help to overcome residual non-stationary noise components. Furthermore, residual noise shaping can be utilized. Evaluations have shown that highly non-stationary noises are suppressed considerably without degrading the speech quality.

Index Terms— Speech enhancement, noise suppression, maximum attenuation, non-stationary noise, beamforming

1. INTRODUCTION

In several applications such as hands-free telephony or speech-dialog systems the recording of a speech signal takes place in a noisy environment. For systems, e.g., installed in cars, used in sidewalk cafes or in train stations, the local speech is often corrupted by background noise. Therefore, noise suppression algorithms are used to attenuate the distorted components while keeping the speech signal as natural as possible.

In literature a wide variety of different noise reduction characteristics, such as the approaches proposed by Ephraim/Malah [2, 3], Wiener filtering in its direct or recursive way [5], or the method proposed by Lotter [6] (to mention just a few), exist. These approaches differ mainly in their optimization criteria (e.g. MMSE or MAP) and in the statistical models which are used for the speech and the noise signal, respectively. One major limitation that exists in common noise suppression techniques is that these schemes are usually unable to attenuate non-stationary noise components. The interfering sound fields in a cafeteria or a station for instance can be considered to be non-stationary. In automotive applications the non-stationary disturbances result, e.g., from overtaking cars or from cars on the opposite lane. Often single-channel speech enhancement methods classify abrupt increases of disturbance level as speech onsets – as a consequence the filter "opens" and non-stationary noise bursts are not attenuated at all.

A large improvement can be achieved by using more than one microphone in conjunction with beamforming schemes followed by a post-filter that exploits spatial information (see, e.g., [1] for an overview). However, even if multi-channel approaches clearly improve the behavior for non-stationary noise, the residual background noise usually still follows mainly the original noise – attenuated by a fixed or at most by a slowly changing maximum attenuation (see, e.g., [7]) of about 6 to 20 dB. The residual non-stationary noise, after applying a maximum allowed attenuation, is often still annoying.

The methods that will be presented in the following utilize an adaptive adjustment of the maximum attenuation. This makes it possible to transform a highly non-stationary residual noise into a stationary one or in a better sounding type of disturbance which follows the spectral behavior of an a priori specified residual noise spectrum. Also residual noise shaping to enhance speech recognition systems can be applied with that methods. The contribution is organized as follows: first the main idea of classical noise suppression will be presented shortly, followed by the derivation of four proposed methods. The paper concludes with simulation results and a summary.

2. CONVENTIONAL NOISE SUPPRESSION RULES

In the following it is assumed that the microphone signal $y(n)$ consists of speech $s(n)$ and of undesired background noise components $b(n)$:

$$y(n) = s(n) + b(n)\,. \tag{1}$$

The signal $y(n)$ might also be regarded as the output of a multi-channel preprocessing, e.g. a beamformer. For signal enhancement often the noisy speech signal is first split up into overlapping block segments of appropriate size. The segmentation can be described by extracting the M last recent samples of the input signal and combining them to a vector. Typically successive segments are subsampled by a factor $r = M/4$ or $r = M/2$. In order to separate the desired and undesired signal components each signal block is multiplied by a window function h_k and transformed into the frequency domain using a filterbank or a DFT :

$$Y\left(e^{j\Omega_\mu}, n\right) \;=\; \sum_{k=0}^{M-1} y(nr-k)\, h_k\, e^{-j\Omega_\mu k}\,. \tag{2}$$

The frequency supporting points Ω_μ can be distributed equidistantly over the normalized frequency range as $\Omega_\mu = 2\pi\mu/M$ with $\mu \in \{0, \ldots, M-1\}$. Depending on the current SNR in each frequency subband μ an attenuation factor is computed using a noise suppression characteristic. If, e.g., the recursive Wiener approach [5] is chosen the corresponding weights are determined as follows:

$$G\big(e^{j\Omega_\mu}, n\big) = \max\left\{G_{\min},\, 1 - \beta\big(e^{j\Omega_\mu}, n\big)\frac{\widehat{S}_{bb}(\Omega_\mu, n)}{\big|Y(e^{j\Omega_\mu}, n)\big|^2}\right\},$$

$$\text{with } \beta\big(e^{j\Omega_\mu}, n\big) = \min\left\{\beta_{\max},\, \frac{1}{G\big(e^{j\Omega_\mu}, n-1\big)}\right\}. \qquad (3)$$

The quantities $G_{\min}$ and $\beta_{\max}$ are the maximum filter attenuation and the maximum over-estimation factor, respectively. In order to estimate the PSD of the noise $\widehat{S}_{bb}(\Omega_\mu, n)$ simply first-order IIR smoothing can be applied during speech pauses. An enhanced method to estimate $\widehat{S}_{bb}(\Omega_\mu, n)$ for multi-channel applications can be found, e.g., in [8]. This method, which is optimized in the MAP sense, exploits spatial information using the output signals of a GSC-type adaptive beamformer and of a blocking matrix.

Finally, the weighting factors are applied to the noisy input short-term spectrum (STS) to get the enhanced output STS:

$$\widehat{S}\big(e^{j\Omega_\mu}, n\big) \quad = \quad Y\big(e^{j\Omega_\mu}, n\big)\, G\big(e^{j\Omega_\mu}, n\big)\,. \qquad (4)$$

The synthesized output signal $\hat{s}(n)$ is computed by performing first an inverse DFT to the weighted STS $\widehat{S}(e^{j\Omega_\mu}, n)$ followed by appropriate windowing and adding of the overlapping output frames.

With multi-channel approaches one can additionally increase the SNR of the incoming signal and also the robustness of the background noise estimation can be improved considerably by exploiting spatial information. However, due to the limitation inserted either directly in current filter characteristic ($G_{\min}$ in Eq. 3) or within the SNR estimation, the output spectrum follows in noise-only periods the (attenuated) spectrum of the disturbance:

$$\widehat{S}\big(e^{j\Omega_\mu}, n\big)\big|_{\text{Noise-only periods}} \quad = \quad Y\big(e^{j\Omega_\mu}, n\big)\, G_{\min}\,. \qquad (5)$$

3. PROPOSED NOISE SUPPRESSION METHOD

In contrast to conventional noise suppression algorithms the herein proposed method utilizes a predefined PSD of a residual background noise $S_{bb,\text{des}}(\Omega_\mu)$. The maximum attenuations for the filter coefficients are adjusted such that the current residual noise matches with $S_{bb,\text{des}}(\Omega_\mu)$ (at least in terms of the spectral envelope, not necessarily in terms of the probability densities of the individual subbands). However, the a priori specified residual noise $S_{bb,\text{des}}(\Omega_\mu)$ can be, e.g., a more comfortable and better sounding type of disturbance compared to the original one.

Various types of acoustical noises in human environments like car, cafeteria, or station noise are of interest in noise control applications. Starting with car noise, the main components are engine, wind, and rolling noise. Their spectral behavior varies depending on the type of vehicle, the current speed, and the surface of the road. In Fig. 1 the estimated PSDs of three different background noises are shown. The signals were measured in a Mercedes, a BMW, and a Porsche at a speed of 160 km/h. The recordings were performed using the integrated hands-free microphones of the individual cars (including individually optimized equalizations, mainly noticeable at low frequencies). Due to the different shapes depicted in Fig. 1 a residual noise shaping can be performed utilizing a common desired residual background noise for all car types. The specified residual noise can be, e.g., extracted from a well sounding noise of a specific car. Further on, noise shaping can also be employed, for instance, to utilize a type of residual noise like it was previously used to train a speech recognition system. Thereby, it will be guaranteed that the recognizer will "see" the same residual noise for training as well as for application mode. This helps reducing the amount of memory used for modeling noise components.

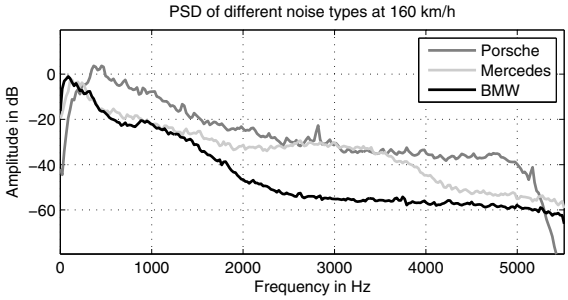

Fig. 1. *Mean PSDs of background noises measured in different vehicles at a speed of 160 km/h.*

Noise shaping can be applied as well for high non-stationary noises measured, e.g., in a cafeteria or in a station environment. The new methods have the capability to transform the highly non-stationary residual noise into a stationary one which follows the spectral behavior of an a priori specified residual noise spectrum.

For the derivation of the new methods it is assumed that a desired shape of the noise PSD $S_{bb,\text{des}}(\Omega_\mu)$ is already found and specified. During initialization a real-valued desired residual background noise spectrum is obtained from $S_{bb,\text{des}}(\Omega_\mu)$ at frame index $n = 0$:

$$B_{\text{des}}\big(e^{j\Omega_\mu}, 0\big) = \sqrt{S_{bb,\text{des}}(\Omega_\mu)}\,. \qquad (6)$$

This desired noise spectrum $B_{\text{des}}(e^{j\Omega_\mu}, n)$ is now continuously adjusted to the current background noise $b(n)$, respectively its estimated PSD. If the current frame is classified as a noise-only one the desired noise spectrum is corrected – during speech activity it is retained unchanged:

$$B_{\text{des}}\big(e^{j\Omega_\mu}, n\big) = \begin{cases} \Delta_{\text{B}}(n)\, B_{\text{des}}\big(e^{j\Omega_\mu}, n-1\big), & \overline{G}(n) < G_0\,, \\ B_{\text{des}}\big(e^{j\Omega_\mu}, n-1\big), & \text{else}\,. \end{cases} \qquad (7)$$

The condition $\overline{G}(n) < G_0$ is a simplified speech activity detector. $\overline{G}(n)$ is defined as the previous mean of the estimated attenuation factors $G\big(e^{j\Omega_\mu}, n-1\big)$ as:

$$\overline{G}(n) = \frac{1}{M}\sum_{\mu=0}^{M-1} G\big(e^{j\Omega_\mu}, n-1\big), \qquad (8)$$

whereas G_0 is a predetermined threshold that can be set, e.g., to -6 dB. Depending on the estimated background noise $\widehat{S}_{bb}(\Omega_\mu, n)$ in a certain frequency range between μ_0 and μ_1, the correction factor $\Delta_{\text{B}}(n)$ is determined as follows:

$$\Delta_{\text{B}}(n) = \begin{cases} \Delta_{\text{inc}}, & \text{if } \sum_{\mu=\mu_0}^{\mu_1} B_{\text{des}}^2\big(e^{j\Omega_\mu}, n-1\big) \\ & \qquad < \widetilde{G}_{\min}^2 \sum_{\mu=\mu_0}^{\mu_1} \widehat{S}_{bb}(\Omega_\mu, n)\,, \\ \Delta_{\text{dec}}, & \text{else,} \end{cases} \qquad (9)$$

$$\text{with} \qquad 0 \ll \Delta_{\text{dec}} < 1 < \Delta_{\text{inc}} \ll \infty\,. \qquad (10)$$

This means that the PSD of the desired background noise keeps its initial shape but follows the power of the real background noise (measured in a small frequency range $[\Omega_{\mu_0}, \Omega_{\mu_1}]$). With the parameter $\widetilde{G}_{\min}$ one can specify the desired noise attenuation in this range. Evaluations have shown, that the following settings are particularly favorable for a setup with a sampling rate $f_{\text{s}} = 11025$ Hz and a block overlap of 75 %:

$$\Delta_{\text{dec}} = 0.98 \quad \text{and} \quad \Delta_{\text{inc}} = 1.02\,. \qquad (11)$$

Furthermore, a reasonable frequency range for Eq. 9 is:

$$\Omega_{\mu_0} \simeq 400\,\text{Hz} \quad \text{and} \quad \Omega_{\mu_1} \simeq 700\,\text{Hz}\,. \tag{12}$$

Due to the slow multiplicative correction fast fluctuations of the true and of the estimated background noise $\widehat{S}_{bb}(\Omega_\mu, n)$ will only influence marginally the desired residual noise $B_{\text{des}}(e^{j\Omega_\mu}, n)$.

In the following we will present four different approaches. All of them can be regarded as a frequency-selective, time-variant version of the limitation for the attenuation factors in Eq. 3. Here the parameter $G_{\min}$ is replaced by $G_{\min,i}(e^{j\Omega_\mu}, n)$, resulting in

$$G\big(e^{j\Omega_\mu}, n\big) = \max\left\{ G_{\min,i}\big(e^{j\Omega_\mu}, n\big),\, 1 - \beta\big(e^{j\Omega_\mu}, n\big)\, \frac{\widehat{S}_{bb}(\Omega_\mu, n)}{\big|Y\big(e^{j\Omega_\mu}, n\big)\big|^2} \right\}, \tag{13}$$

with $i \in \{1,\, 2,\, 3,\, 4\}$ representing one of the four approaches.

A straight forward approach would be to compute the gain factors according to Eq. 3 without any limitation ($G_{\min} = 0$) first, apply them to the subband signals, and check afterwards if the output signal amplitudes are smaller than the desired noise amplitudes. If this is detected the amplitudes should be reset to $B_{\text{des}}(e^{j\Omega_\mu}, n)$ (using the phases of the input signals). An equivalent way to achieve this is to set the adaptive maximum attenuations in Eq. 13 as:

$$G_{\min,1}\big(e^{j\Omega_\mu}, n\big) \;=\; \min\left\{ G_0,\, \frac{B_{\text{des}}\big(e^{j\Omega_\mu}, n\big)}{\big|Y\big(e^{j\Omega_\mu}, n\big)\big|} \right\}. \tag{14}$$

Due to the minimum operator at least a suppression limit of G_0 will be applied. If this is omitted a noise amplification might happen. For G_0 we suggest to use $G_0 = 0.5$ resulting in a minimum attenuation of about 6 dB. Simulations and measurements have shown, that the residual noise of this approach sounds quite annoying due to a very tonal characteristic. These artifacts are caused by the fact that only slow variations of $B_{\text{des}}(e^{j\Omega_\mu}, n)$ are allowed in Eq. 7. Thus, the amplitudes of the output subband signals are changing also very slowly. In combination with quickly changing phases of the individual subbands a buzzy sounding residual noise is created. This can be avoided by allowing some small variations of the output amplitudes.

This can be realized in a simple manner by recursive smoothing of the results of our first approach according to:

$$\begin{aligned} G_{\min,2}\big(e^{j\Omega_\mu}, n\big) \;&=\; \gamma\, G_{\min,2}\big(e^{j\Omega_\mu}, n-1\big) \\ &\quad + (1-\gamma)\, G_{\min,1}\big(e^{j\Omega_\mu}, n\big)\,. \end{aligned} \tag{15}$$

For the smoothing coefficient γ values can be chosen within the range

$$0 \;\le\; \gamma \;<\; 1. \tag{16}$$

A small value of γ will result in low variations of the predetermined noise whereas a large value in higher fluctuations. The latter choice leads to a more pleasant and natural sounding of the residual noise. Several listening tests have shown that $\gamma = 0.8$ was a reasonable choice for the setup described before. It has to be mentioned, however, that due to the smoothing the current adaptive maximum attenuation coefficients are not optimal anymore. Sometimes at *highly* non-stationary noise scenarios, the current residual noise may not match with the a priori specified desired noise. To overcome this, the adaptive maximum attenuations from Eq. 14 are combined with artificial fluctuation weights $G_{\text{n}}(e^{j\Omega_\mu}, n)$ according to:

$$G_{\min,3}\big(e^{j\Omega_\mu}, n\big) \;=\; G_{\min,1}\big(e^{j\Omega_\mu}, n\big)\, G_{\text{n}}\big(e^{j\Omega_\mu}, n\big)\,. \tag{17}$$

Whereas the statistical properties of the random fluctuations should be chosen as:

$$\text{E}\Big\{ G_{\text{n}}\big(e^{j\Omega_\mu}, n\big) \Big\} \;=\; 1.0\,, \tag{18}$$

$$\text{Var}\Big\{ G_{\text{n}}\big(e^{j\Omega_\mu}, n\big) \Big\} \;=\; 0.2 \ldots 0.4\,. \tag{19}$$

The last approach, which leads in our opinion to the most comfortable type of residual noise for automotive applications, consists of an adaptive increment and decrement mechanism that takes the previous attenuation coefficients into account (beside a fixed multiplicative correction):

$$G_{\min,4}\big(e^{j\Omega_\mu}, n\big) = \begin{cases} \gamma_{\text{inc}}\, G_{\min,4}\big(e^{j\Omega_\mu}, n-1\big) + \gamma_G\, G\big(e^{j\Omega_\mu}, n-1\big), \\ \quad \text{if} \quad B_{\text{des}}\big(e^{j\Omega_\mu}, n\big) > G_{\min,4}\big(e^{j\Omega_\mu}, n-1\big)\big|Y\big(e^{j\Omega_\mu}, n\big)\big|, \\ \gamma_{\text{dec}}\, G_{\min,4}\big(e^{j\Omega_\mu}, n-1\big) - \gamma_G\, G\big(e^{j\Omega_\mu}, n-1\big)\,, \\ \quad \text{else}\,. \end{cases} \tag{20}$$

For the same setup that was already mentioned before the following choices for the time constants were selected:

$$\gamma_{\text{inc}} = 1.05, \quad \gamma_{\text{dec}} = 0.995, \quad \text{and} \quad \gamma_G = 0.1\,. \tag{21}$$

Please note that the maximum attenuation values $G_{\min,i}(e^{j\Omega_\mu}, n)$ can be applied also for other filter characteristics than the recursive Wiener filter (described in Eq. 13). Furthermore, the authors suggest to apply the methods as a post-processor for beamformers. Thus spatial information can be exploited for a reliable classification of noise versus desired signal. However, even two-channel approaches are able to achieve such a classification with sufficient reliability.

4. EXPERIMENTAL RESULTS

The proposed methods for adjusting the maximum allowed attenuation in noise suppression schemes can be applied for a variety of speech applications. It can be used, e.g., as a post-processor at the output of a beamformer to avoid residual non-stationary noise components. To demonstrate the effect, time-frequency analyses of a noisy speech signal are depicted in Fig. 2 measured in a sidewalk cafe at an SNR $= 3$ dB. A uniform linear array with 4.2 cm spacing and 4 channels has been employed. In the upper diagram the clean speech signal is shown using a sampling frequency $f_{\text{s}} = 11025$ Hz and a subsampling factor $r = 64$. The microphone signals have been generated using impulse responses and highly non-stationary noise, which were actually measured in an acoustical environment of a sidewalk cafe. The speaker was seated in broadside direction with a distance of 1 meter to the microphone array. A microphone signal (second channel of the array) is depicted in the second graph. The analysis in the third diagram presents the output of a GSC-type adaptive beamformer [1] and a Wiener post-filter with fixed maximum attenuation. As a compromise between speech distortion and noise reduction a fixed maximum attenuation of $G_{\min} = 15$ dB has been applied. A background noise estimate which has been derived for spatial post-filtering (as described in [8]) was used in this scenario. Due to a reliable background noise estimation almost all transient noise components were classified correctly as noise. Although they were not detected as speech no sufficient attenuation has been applied – the residual non-stationary noise components are still audible and perturbing. As it is shown in the third diagram the residual noise spectrum still follows the input spectrum. The analysis at the bottom of Fig. 2 represents the enhanced output signal by applying the same Wiener like post-filtering but now with the proposed adaptive maximum attenuations. An average shape extracted from a short part

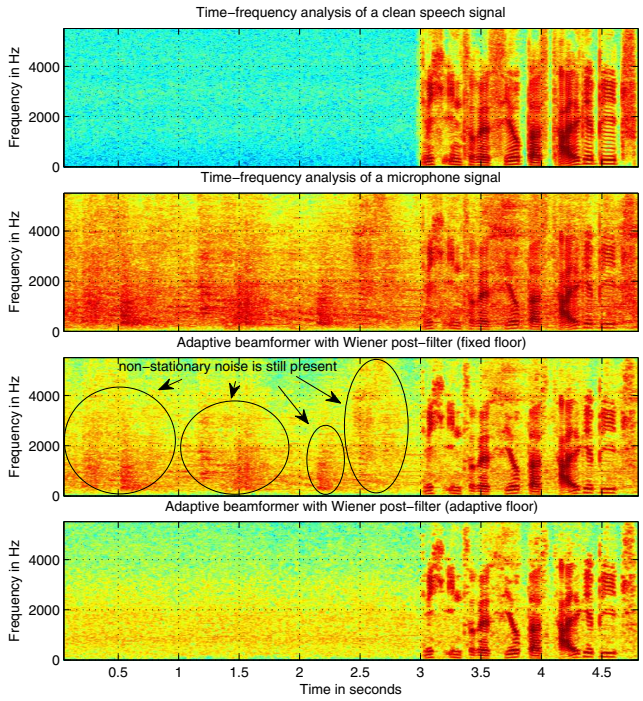

Fig. 2. *Example of time-frequency analyses with input SNR=3 dB in an acoustical environment of a sidewalk cafe.*

of a sidewalk cafe background noise has been utilized as a desired residual noise. The maximum allowed attenuations $G_{\min,3}(e^{j\Omega_\mu}, n)$ and a noise attenuation of $\widetilde{G}_{\min} = 15$ dB (Eq. 9) have been used for this scenario. It can be seen that the non-stationary noise components have been removed completely. At the same time the speech quality has not been affected. Evaluations at different SNRs with the same setup have shown that non-stationary noise components can be eliminated completely while speech distortions are almost kept unchanged compared to the fixed maximum attenuation as shown in Fig. 3. The distortion has been measured using the log-spectral distance according to:

$$LSD = \frac{10}{L}\sum_{n=0}^{N}\sqrt{\sum_{\mu=0}^{M/2}\frac{K_{\mu,n}}{\overline{K}_n}\lg^2\left(\frac{\max\{|S(e^{j\Omega_\mu},n)|^2,\delta_s\}}{\max\{|\widehat{S}(e^{j\Omega_\mu},n)|^2,\delta_{\hat{s}}\}}\right)}. \quad (22)$$

Whereas $|S(e^{j\Omega_\mu}, n)|^2$ is the PSD of the clean speech and the lower bound is defined as $\delta_s = 10^{-5}\max_{\mu,n}\{|S(e^{j\Omega_\mu}, n)|^2\}$. The binary mask $K_{\mu,n} \in \{0, 1\}$ is used to select only components that satisfy the condition: $|S(e^{j\Omega_\mu}, n)|^2 \geq \delta_s$. The corresponding normalization is given by $\overline{K}_n = \max\{\sum_\mu K_{\mu,n}, 0.1\}$. The quantity N represents the number of frames in the signal and L corresponds to the number for which $\overline{K}_n \geq 1$.

Evaluations have also shown that short noise bursts or strong tonal disturbances as they are often produced by cars can be removed almost completely by using the proposed method from Eq. 20 instead of a fixed maximum attenuation (with two-channel processing). The same is true for other transient signals such as they are produced by indicator clicks or by the windshield wiper.

Furthermore, the last method (Eq. 20) has been evaluated with a speech recognizer and a single-channel preprocessing with different noise suppression characteristics. A Lombard [4] data base was utilized consisting of a huge amount of speech signals at different SNR. For the evaluation the recursive Wiener filter [5] and the characteristics according to Ephraim/Malah [2, 3] and Lotter [6] have been employed. The speech recognition system was *trained individually*

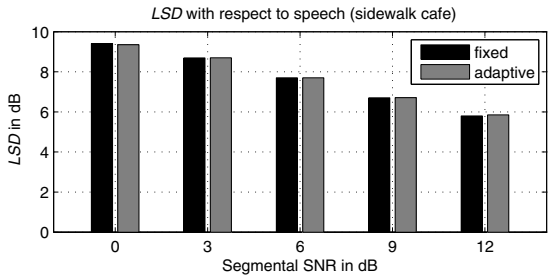

Fig. 3. *Measured log-spectral distance (LSD)* for speech.

for all characteristics and for fixed and adaptive maximum attenuation, respectively. The results have shown that the word accuracy can be increased when using the adaptive maximum attenuation – by approx. 3% relative for Wiener and Ephraim/Malah and 4% for the Lotter characteristic. Although the word accuracies have shown small improvements, it should be mentioned that the proposed method has still the capability for enhancing speech recognition in the future. Due to a predefined residual noise spectrum, less model parameters are needed for the classification of different noise types. The reduced number of model parameters for noise can advantageously be exploited for enhancing the speech model.

5. CONCLUSIONS

A new method for noise suppression and its applications was presented. Unlike conventional noise suppression algorithms the proposed methods utilize a desired residual background noise and the maximum attenuation coefficients are determined adaptively. It was shown that non-stationary noise components such as the ones produced in a sidewalk cafe or by passing or overtaking cars in automotive hands-free systems can be suppressed considerably without degrading the speech quality. Furthermore, residual noise shaping can be utilized – e.g. to transform an unpleasant sounding background noise of a vehicle A into a more pleasant sounding one of a vehicle B. The proposed methods are particularly suitable as post-processing for beamformers. Moreover the new approaches can advantageously be employed for enhancing speech recognition systems.

6. REFERENCES

[1] M. Brandstein, D. Ward (eds.), *Microphone Arrays,* Springer, Berlin, Germany, 2001.

[2] Y. Ephraim, D. Malah, "Speech Enhancement Using a MMSE Short-Time Spectral Amplitude Estimator," *IEEE Trans. Acoust. Speech Sig. Proc.,* vol. 32, no. 6, pp. 1109–1121, 1984.

[3] Y. Ephraim, D. Malah, "Speech Enhancement Using a MMSE Log-Spectral Amplitude Estimator," *IEEE Trans. Acoust. Speech Sig. Proc.,* vol. 33, no. 2, pp. 443–445, 1985.

[4] J. C. Junqua, "The Influence of Acoustics on Speech Production, a Noise-Induced Stress Phen. Known as the Lombard Reflex," *Speech Communication,* vol. 20, no. 1, pp. 13–22, 1996.

[5] K. Linhard, T. Haulick, "Spectral Noise Subtraction with Recursive Gain Curves," *Proc. ICSLP*, Sydney, Australia, 1998.

[6] T. Lotter, P. Vary "Speech Enhancement by MAP Spectral Amplitude Estimation Using a Super-Gaussian Speech Model," *EURASIP Journal on Applied Signal Processing*, vol. 2005, no. 7, pp. 1110–1126, 2005.

[7] H. Puder, "Single Channel Noise Reduction Using Time-Frequency Dependent Voice Activity Detection," *Proc. IWAENC*, Pocono Manor, NY, USA, 1999.

[8] T. Wolff, M. Buck "Spatial MAP Post-Filtering for Arbitrary Beamforming," *Proc. HSCMA*, Trento, Italy, 2008.