Energy-decay Based Postfilter for ICC Systems with Feedback Cancellation

Marco Gimm[1], *Philipp Bulling*[2], *Gerhard Schmidt*[1]

[1]*Christian-Albrechts-Universität zu Kiel,* [2]*Daimler AG*
mgi@tf.uni-kiel.de

Abstract: In so called In-Car Communication (ICC) systems the most challenging problem is feedback that occurs when a recorded and amplified signal is played back and recorded again. This electro-acoustic loop restricts the maximum gain that can be applied in such system until it gets unstable. There are different solutions to increase the stability of ICC systems. One is to use model-based feedback suppression. Another is to use cancellation approaches as they are well established in hands-free systems to reduce echo.
In this paper, we propose a postfilter based on a modified energy-decay model that is used in combination with feedback cancellation.

1 Introduction

Todays ICC systems are often using feedback cancellation based on adaptive filter approaches. Therefore, larger gains can be applied, making such systems convinient for practical usage. Theoretically, this method leads to an complete removal of all feedback componets but practically, there are several limitations. This means that there will always be a residual feedback. One example is the length of the adaptive filter that is restricted due to computational complexity. Another one is the estimation of the impulse response in the adaptive filter that always has uncertainties.

A different method to increase the maximum gain in an ICC system is to estimate the power spectral density (PSD) of the feedback and use it for a Wiener filter. The advantage of this method is that the PSD estimation can be done recursively using an infinite impulse response (IIR) model, so there is no length limitation. By just using the PSD, the phase information of the feedback is neglected. This leads to a spectral attenuation which also affects the speech signal.

In this paper, we investigate different possibilities to combine the two approaches. Such combinations allow to use shorter cancellation filters, since the succeeding model-based suppression approach can attenuate remaining feedback components.

The paper is structured as following: In the next section previous work is summarized. The model-based feedback suppression is described in Sec. 3. The model is then adapted to be used as a postfilter for the feedback cancellation in Sec. 4. A short overview of the model parameters is presented in Sec. 5. In Sec. 6, we evaluate the method by using objective and subjective methods. Finally, Sec. 7 provides the conclusion.

2 Previous Work

The problem of acoustic feedback cancellation is of interest in various applications, such as hearing aids and public-address systems. Some of the most important approaches are collected in [1]. In recent years, different methods, especially designed for ICC systems, were published.

An overview about ICC systems and the feedback problem is given in [2]. In [3] an adaptive filter is combined with a Wiener filter based residual echo suppressor to cancel the acoustic feedback in an ICC system. This approach is based on echo cancellation methods used for hands-free systems, as for example explained in [4]. However, the problem is a strong correlation between the local speech and the loudspeaker signal which causes adaptive filters not to converge in a closed electro-acoustic loop. Therefore, acoustic feedback cancellation requires an additional decorrelation stage and more sophisticated control mechanisms. Meanwhile, different approaches to overcome this problem exist. In [5] the loudspeaker signal is decorrelated from the local speech by means of a frequency shift. In [6] the stepsize of the adaptive filter is controlled depending on reverberation. To reduce computational complexity, the length of the adaptive filter is limited in real applications and therefore a residual feedback remains. A method to suppress feedback by means of spectral subtraction is given in [7]. This approach is based on dereverberation by means of a statistical model of the impulse response [8, 9].

3 Energy Decay Model for Feedback Suppression in ICC Systems

To increase the stability limit in ICC systems a feedback suppression as it is described in [7] can be used. The whole system is shown in Fig. 1.

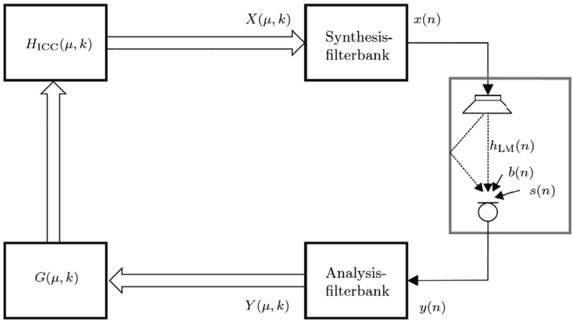

Figure 1 – Structur of a single channel ICC system with feedback suppression.

Here $h_{\mathrm{LM}}(n)$ is the impulse response of the car cabin, $s(n)$ is the clear speech signal and $b(n)$ is additional noise, with n being the discrete time index. The microphone signal $y(n)$ can be described as

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

$$r(n) = \sum_{i=0}^{\infty} x(n-i)\, h_{\mathrm{LM}}(i), \tag{2}$$

with $r(n)$ being the feedback and $x(n)$ the loudspeaker signal. Statistically, acoustic impulse responses are often described as an exponentially decaying noise sequence according to

$$h_{\mathrm{LM,mod}}(n) = \begin{cases} 0, & \text{for } n < p, \\ g(n)\, e^{\alpha(n-p)}, & \text{for } p \leq n, \end{cases} \tag{3}$$

with $g(n)$ being a Gaussian distributed random process and p being the latency between the loudspeaker and the microphone. α describes the decay behavior in dependency of the reverberation time T_{60} and the sample rate f_{S}

$$\alpha = \frac{3\ln 10}{T_{60} f_{\mathrm{S}}}. \tag{4}$$

Using the modeled impulse response, the expected value of the energy envelope in the subband domain can be derived:

$$E\left\{\left|H_{\mathrm{LM,mod,A}}(\mu,k)\right|^2\right\} = \begin{cases} 0, & \text{for } k < P(\mu), \\ A(\mu)\, e^{-\gamma(\mu)(k-P(\mu))}, & \text{for } P(\mu) \le k. \end{cases} \tag{5}$$

$P(\mu)$ is the latency at the subband index μ. $A(\mu)$ are coupling factors and $\gamma(\mu)$ describes the decay behavior. The convolution of this model with the PSD of the loudspeaker signal $S_{xx}(\mu,k)$ leads to an estimation of the feedback's PSD $S_{rr}(\mu,k)$. This calculation can be done recursively

$$\hat{S}_{rr,\mathrm{A}}(\mu,k) = A(\mu)\, S_{xx}(\mu,k-P(\mu)) + e^{-\gamma(\mu)}\, \hat{S}_{rr}(\mu,k-1). \tag{6}$$

The estimation of the feedback's PSD can now be used for a frequency-domain Wiener filter

$$G(\mu,k) = 1 - \frac{\hat{S}_{rr}(\mu,k)}{S_{yy}(\mu,k)} \tag{7}$$

that can be used to suppress the feedback in the recorded signal

$$Y(\mu,k) = S(\mu,k) + B(\mu,k) + H_{\mathrm{LM}}(\mu,k)\, X(\mu,k) \tag{8}$$

$$\hat{S}(\mu,k) + \hat{B}(\mu,k) = Y(\mu,k)\, G(\mu,k). \tag{9}$$

4 Residual Feedback Suppression in a System using Feedback Cancellation

When an adaptive filter is added to the system as depicted in Fig. 2, it will affect the energy decay behavior.

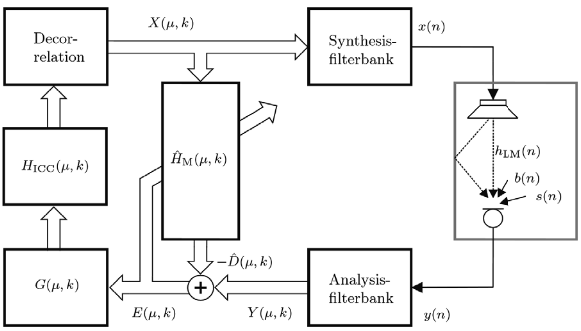

Figure 2 – Structur of a single channel ICC system with feedback cancellation.

If a model as described in the previous section is used for the feedback estimation within a postfilter, it has to be adapted with respect to the cancellation. In this case $G(\mu,k)$ can be used as a postfilter to suppress the residual feedback after the cancellation approach. The signal $e(n)$ is the microphone signal after subtracting the estimated feedback $\hat{d}(n)$

$$e(n) = s(n) + b(n) + \mathbf{x}^T(n)\,\mathbf{h}_{\mathrm{LM}}(n) - \hat{d}(n) \tag{10}$$
$$= s(n) + b(n) + \mathbf{x}^T(n)\,\mathbf{h}_{\mathrm{LM}}(n) - \mathbf{x}_m^T(n)\,\hat{\mathbf{h}}_{\mathrm{LM},m}(n) \tag{11}$$
$$= s(n) + b(n) + \mathbf{x}^T(n)\,\mathbf{h}_{\mathrm{LM,Res}}(n). \tag{12}$$

The impulse response of the vehicle cabin $h_{\mathrm{LM}}(n)$ and the one of the adaptive filter $\hat{h}_{\mathrm{LM},m}$ can be combined to $h_{\mathrm{LM,Res}}(n)$ according to

$$h_{\mathrm{LM,Res}}(n) = \begin{cases} h_{\mathrm{LM}}(n) - \hat{h}_{\mathrm{LM},m}(n), & \text{for } n < m, \\ h_{\mathrm{LM}}(n), & \text{else,} \end{cases} \tag{13}$$
$$= \begin{cases} h_{\mathrm{LM},\Delta}(n), & \text{for } n < m, \\ h_{\mathrm{LM}}(n), & \text{else.} \end{cases} \tag{14}$$

This can now be used for a new energy decay model in the subband domain

$$E\left\{\left|H_{\mathrm{LM,mod,B}}(\mu,k)\right|^2\right\} = \begin{cases} \left\|\mathbf{H}_\Delta(\mu,k)\right\|^2, & \text{for } 0 < M-1, \\ A(\mu)\,e^{-\gamma(\mu)(k-P(\mu))}, & \text{for } M \le k. \end{cases} \tag{15}$$

Here M is the filter length in frames and $\left\|\mathbf{H}_\Delta(\mu,k)\right\|^2$ is the system distance. A convolution of the model with the PSD of the loudspeaker signal leads to the estimated PSD of the feedback. This can be divided into two parts

$$\hat{S}_{rr,\mathrm{Res,B}}(\mu,k) = \sum_{i=0}^{M-1} S_{xx}(\mu,k-i)\left\|\mathbf{H}_\Delta(\mu,k)\right\|^2 + \sum_{i=M}^{\infty} S_{xx}(\mu,k-i)\,A(\mu)\,e^{-\gamma(\mu)\left(i-P(\mu)\right)} \tag{16}$$
$$= \hat{S}_{hh}(\mu,k) + \hat{S}_{mm}(\mu,k). \tag{17}$$

Each of these parts can be calculated recursively and added to obtain $\hat{S}_{rr,\mathrm{Res}}(\mu,k)$

$$\hat{S}_{hh}(\mu,k) = \left\|\mathbf{H}_\Delta(\mu,k)\right\|^2 \left(S_{xx}(\mu,k) - S_{xx}(\mu,k-M)\right) + \hat{S}_{hh}(\mu,k-1) \tag{18}$$
$$\hat{S}_{mm}(\mu,k) = S_{xx}(\mu,k-M)\,A(\mu)\,e^{-\gamma(\mu)(M-P(\mu))} + e^{-\gamma(\mu)}\,\hat{S}_{mm}(\mu,k-1). \tag{19}$$

Another approach is to use the old model with an attenuation in the part $k \in [0, M-1]$ when the adaptive filter needs to be taken into account

$$E\left\{\left|H_{\mathrm{LM,mod,C}}(\mu,k)\right|^2\right\} = \begin{cases} 0, & \text{for } k < P(\mu), \\ \left\|\mathbf{H}_\Delta(\mu,k)\right\|^2 A(\mu)\,e^{-\gamma(\mu)(k-P(\mu))}, & \text{for } P(\mu) \le k < M, \\ A(\mu)\,e^{-\gamma(\mu)(k-P(\mu))}, & \text{for } M \le k. \end{cases} \tag{20}$$

Using this model the PSD of the residual feedback can be calculated as follows:

$$\hat{S}_{rr,\mathrm{Res,C}}(\mu,k) = \sum_{i=P(\mu)}^{M-1} S_{xx}(\mu,k-i)\left\|\mathbf{H}_\Delta(\mu,k)\right\|^2 A(\mu)\,e^{-\gamma(\mu)\left(i-P(\mu)\right)}$$
$$+ \sum_{i=M}^{\infty} S_{xx}(\mu,k-i)\,A(\mu)\,e^{-\gamma(\mu)\left(i-P(\mu)\right)} \tag{21}$$
$$= e^{-\gamma(\mu)}\hat{S}_{rr}(\mu,k-1) + S_{xx}\big(\mu,k-P(\mu)\big)\left\|\mathbf{H}_\Delta(\mu,k)\right\|^2 A(\mu)$$
$$+ S_{xx}(\mu,k-M)\left(A(\mu) - \left\|\mathbf{H}_\Delta(\mu,k)\right\|^2 A(\mu)\right)e^{-\gamma(\mu)(M-P(\mu))}. \tag{22}$$

The third approach we investigated is the assumption that the adaptive filter ideally identifies the first m samples of the car's impulse response. In this case, the first m samples of the measured impulse response can be set to zero which leads to an impulse response with a longer latency and a smaller coupling. In this case the model described in Equation 5 can be adapted

$$\hat{S}_{rr,\mathrm{Res,D}}(\mu,k) = e^{-\gamma(\mu)(M-P(\mu))} A(\mu)\, S_{xx}(\mu,k-M) + e^{-\gamma(\mu)}\, \hat{S}_{rr}(\mu,k-1). \tag{23}$$

The following figure shows the energy decay behavior as it is described by the different models, where the first one is the model proposed in [7] and the other ones are the three adapted versions we investigated.

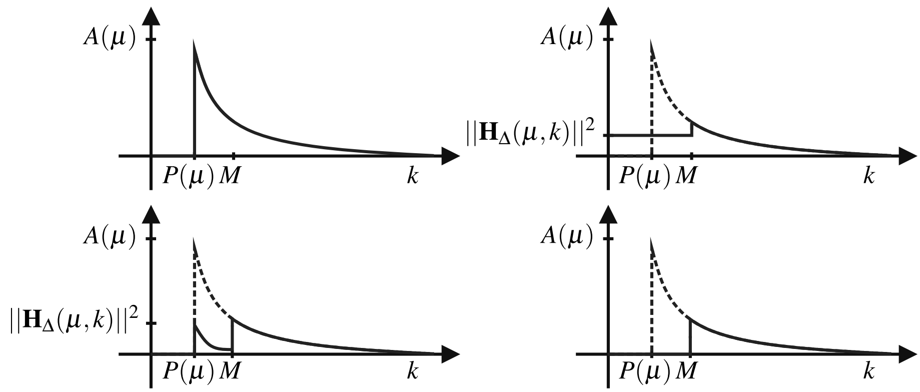

Figure 3 – Scheme of different models. With model A in the top left, model B in the top right, model C in the bottom left and model D in the bottom right.

5 Derivation of the Model Parameters

To implement model-based feedback suppression, the model parameters $A(\mu)$, $P(\mu)$ and $\gamma(\mu)$ can be determined using the measured impulse response.

For simplification $P(\mu)$ and $\gamma(\mu)$ can be assumed to be constant for all frequencies. Thus, they can be derived from the energy decay curve (EDC) of the measured impulse response $h_{\mathrm{LM,mes}}$. The latency T_{D} is given at the time instant, when the EDC begins to drop. The latency in frames $P(\mu)$ is obtained from T_{D} via the relation

$$P(\mu) \approx P = \left\lceil T_{\mathrm{D}} \cdot \frac{f_{\mathrm{S}}}{R} \right\rceil \; \forall \mu \tag{24}$$

where $\lceil \cdot \rceil$ means rounding up to the next integer and R is the frameshift in samples. The reverberation time T_{60} is obtained by subtracting T_{D} from the time instant, when the EDC reaches -60 dB. According to [7], the decay behavior $\gamma(\mu)$ is then given by

$$\gamma(\mu) \approx \gamma = \frac{2 \cdot 3 \ln 10 \cdot R}{T_{60} f_{\mathrm{S}}} \; \forall \mu. \tag{25}$$

The coupling factors $A(\mu)$ are obtained from the smoothed PSD of the measured impulse response. Both, the energy decay curve and the coupling factors of a measured impulse response are shown in Fig. 4. The latency between loudspeaker and microphone is $T_{\mathrm{D}} \approx 6$ ms. At approximately -40 dB the measurement noise dominates the EDC (black solid line). Therefore, it is interpolated linearly (dashed line) to calculate the reverberation time, which is $T_{60} \approx 119.9$ ms. The resulting spectrograms of the modeled and the measured impulse response's PSD are shown in Fig. 5.

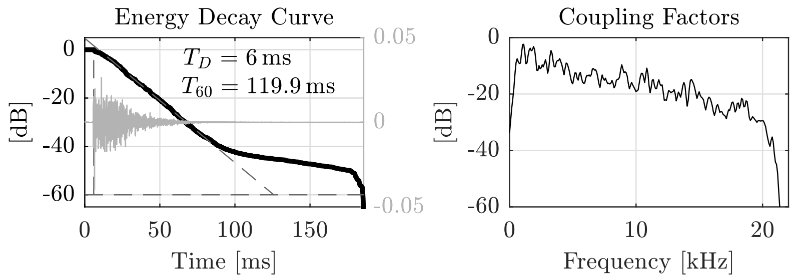

Figure 4 – The parameters of a measured impulse response. This impulse response is also used for simulations.

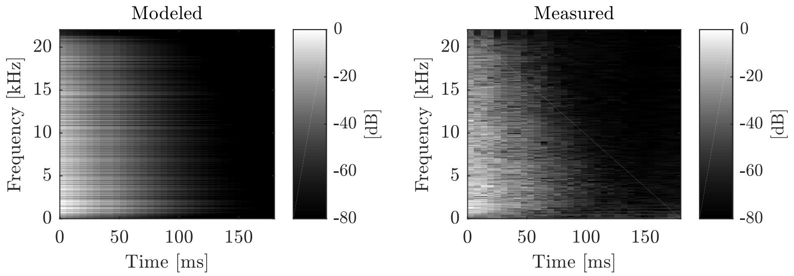

Figure 5 – Measured and modeled PSD over time of the impulse response, shown in Fig 4.

6 Evaluation

In this section, the results of the proposed approaches are presented by means of simulations and a listening test. In both cases, the car cabin is modeled with an impulse response measured in a van. The parameters of this impulse response are shown in Fig. 4. The van is equipped with four loudspeakers in the passenger compartment and a microphone in the headliner above the driver's seat. The impulse response describes the acoustic path from the rear loudspeakers to the microphone.

The local speech is the recording of a male speaker, sitting on the driver's seat. To regard the suppression of the feedback solely, clean speech without background noise was used for simulations.

An overlap-add filterbank with a Hanning window is used to transfer the signals to the frequency domain. The filterbank length is $N = 512$, with frameshift $R = N/2$, at a sampling frequency of $f_S = 44.1\,\mathrm{kHz}$. With $M = 4$, the first 23.2 ms of the impulse response are covered by the adaptive filter. During simulation, the adaptive filter is set to a fixed impulse response, so the system distance is held constant at $\|\mathbf{H}_\Delta(\mu,k)\|^2 = -32.8\,\mathrm{dB}$.

6.1 Simulation Results

In the simulation, we investigated the impairment of the local speech by the Wiener filter compared to the amount of feedback suppression. Therefore, the real feedback $R(\mu,k) \bullet\!\!-\!\!\circ\, r(n)$ and the local speech $S(\mu,k) \bullet\!\!-\!\!\circ\, s(n)$ are weighted by the Wiener filter $G(\mu,k)$. Afterwards, two binary masks are applied. The first mask $M_S(\mu,k)$ is based on a speech activity detection. The elements of the mask are one during periods of local speech and zero during speech pauses or reverberation. The values of the second mask $M_R(\mu,k)$ are zero during speech activity and one during reverberation. Both, the filtered and the unfiltered local speech signal, are weighted with $M_S(\mu,k)$, while the feedback signals are weighted with $M_R(\mu,k)$. Subsequently, the mean power in the range between 260 and 3000 Hz, where the main components of human speech are located, is calculated for all signals. Finally, the unfiltered power is divided by the filtered power.

In Table 1 the resulting powers are shown. Ideally, P_S should be 0 dB, which means that local speech is not attenuated by the Wiener filter at all. Whereas, P_R should be as large as possible, resulting in a strong attenuation of the feedback. The relation P_R/P_S shows the amount of attenuation of feedback and speech. Values close to 0 dB indicate that the attenuation of speech and feedback is equal. It can be seen that the ratio P_R/P_S is approximately equal for the three adapted impulse response models B, C and D, while it is approx 1.2 dB smaller for the original model A. This results in a reduced speech quality, which also emerges in the listening tests.

Table 1 – Average amount of attenuation during speech activity and reverberation at a system distance of $||\mathbf{H}_\Delta(\mu,k)||^2 =$ -32.8 dB.

Model	P_S	P_R	P_R/P_S
A	2.43 dB	8.23 dB	5.80 dB
B	0.51 dB	7.55 dB	7.04 dB
C	0.47 dB	7.49 dB	7.02 dB
D	0.47 dB	7.49 dB	7.02 dB

6.2 Listening Test

To evaluate the subjective meaning about the different models, a listening test with 26 participants was made. In this test, the unprocessed clean speech signal was presented first as a reference. After that, the simulated signal processed in the closed electro acoustic loop with either one of the models and a cancellation filter (A-D) or just the cancellation filter (0) was shown in a random order. The participants had to rate each signal with respect to a MOS scale between 1 and 5.

5. Excellent – Speech sounds like the unprocessed signal
4. Good – Speech is slightly impaired, but sounds natural
3. Fair – Speech is impaired, but the artifacts are not disturbing
2. Poor – Speech quality degrades, interfering artifacts are clearly audible
1. Bad – Speech is heavily impaired

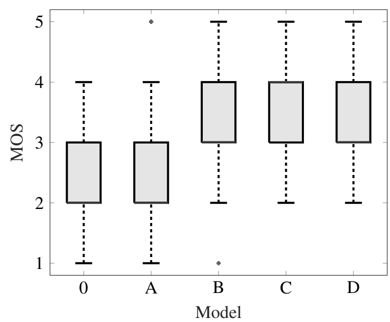

Figure 6 – Result of the listening test combined over all speakers.

We used five different speakers saying non-sense sentences. Three speakers were female and two were male. One of the female speakers was used for a test run which we did not take into account. So the participants had the opportunity to get familiar with the test procedure. The overall results of the test can be seen in Fig. 6.

The results show that the signals without any postfiltering and the signals processed with model A were rated with a median of two. All other signals which were processed with adapted versions of the energy decay model, were rated with a median between three and four.

The three adapted models differ barely. A possibile explanation is that we used a small system distance for our simulation. In this case, the first part of the energy decay model where this parameter is of interest does not contain much energy. This leads to the fact that the difference between the adapted models is just slightly audible.

7 Conclusion and Outlook

In this paper, we presented a postfilter for an acoustic feedback canceler. The postfilter is capable of suppressing the residual feedback, remaining in the signal after feedback cancellation. On the one hand, residual feedback occurs due to misalignment of the adaptive filter. On the other hand, it is caused by the limited length of the adaptive filter. The postfilter is designed in such a way, that both issues are considered. Therefore, different energy decay models are developed with respect to the adaptive filter. Evaluation by means of objective measurements as well as a listening test shows that the proposed postfilter clearly helps to improve the sound, compared to existing approaches. Further enhancement of the algorithm could involve a dynamical update of the model parameters, based on the adaptive filter's impulse response.

References

[1] WATERSHOOT, T. V. and M. MOONEN: *Fifty years of acoustic feedback control: State of the art and future challenges*. *Proceedings of the IEEE*, 99(2), pp. 288–327, 2011.

[2] LINHARD, K. and J. FREUDENBERGER: *Passenger in-car communication enhancement*. In *12th European Signal Processing Conference (EUSIPCO)*, pp. 21–24. Wien, Austria, 2004.

[3] ORTEGA, A., E. LLEIDA, E. MASGRAU, and F. GALLEGO: *Cabin car communication system to improve communications inside a car*. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 2. Orlando, Florida, 2002.

[4] HÄNSLER, E. and G. SCHMIDT: *Hands-free telephones - joint control of echo cancellation and postfiltering*. *Signal Proceesing*, 80(11), pp. 2295–2305, 2000.

[5] WITHOPF, J. and G. SCHMIDT: *Estimation of time-variant acoustic feedback paths in in-car communication systems*. In *14th International Workshop on Acoustic Signal Enhancement (IWAENC)*. Antibes, France, 2014.

[6] BULLING, P., K. LINHARD, A. WOLF, and G. SCHMIDT: *Stepsize control for acoustic feedback cancellation based on the detection of reverberant signal periods and the estimated system distance*. In *Conference of the International Speech Communication Association (Interspeech)*. Stockholm, Sweden, 2017.

[7] WOLF, A. and B. ISER: *Energy decay model based feedback suppression: Theory and application*. In *5th Biennial Workshop on DSP for In-Vehicle Systems*. Kiel, Germany, 2011.

[8] HABETS, E. A. P.: *Multi-channel speech dereverberation based on a statistical model of late reverberation*. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 4, pp. iv/173–iv/176. Philadelphia, USA, 2005.

[9] LEBART, K., J.-M. BOUCHER, and P. N. DENBIGH: *A new method based on spectral subtraction for speech dereverberation*. *Acta Acustica united with Acustica*, 87(3), pp. 359–366, 2001.