

RESEARCH

Open Access

Residual feedback suppression with extended model-based postfilters



Marco Gimm^{1†} , Philipp Bulling^{2†} and Gerhard Schmidt^{1*}

Abstract

When designing closed-loop electro-acoustic systems, which can commonly be found in hearing aids or public address systems, the most challenging task is canceling and/or suppressing the feedback caused by the acoustic coupling of the transducers of such systems. In many applications, feedback cancelation based on adaptive filters is used for this purpose. However, due to computational complexity such a feedback canceler is often limited in the length of the filter's impulse response. Consequently, a residual feedback, which is still audible and may lead to system instability, remains in most cases. In this work, we present enhancements for model-based postfilters based on a priori knowledge of the feedback path which can be used cooperatively with the adaptive filter-based feedback cancelation system to suppress residual feedback and improve the overall feedback reduction capability. For this, we adapted an existing reverberation model such that our model additionally considers the presence and the performance of the adaptive filter. We tested the effectiveness of our approach by means of both objective and subjective evaluations.

Keywords: Residual feedback suppression, Public address system, In-car communication system, Feedback cancelation, Postfilter

1 Introduction

Signal processing in a closed electro-acoustic loop is a challenging task. It occurs in various applications such as hearing aids [1, 2], public address (PA) systems [3, 4] or so-called in-car communication (ICC) systems [5, 6]. In all these systems, feedback occurs because the signal that is played back using a loudspeaker is recorded by a microphone, processed, and then played back again using the same loudspeaker. This may lead to an instability of the system, namely when the loop gain for at least one frequency is larger than 0 dB and the phase is a multiple of 2π . Even if the system is in fact stable, the additional reverberation may make the signals sound unnatural or—more generally—degraded with respect to quality. To reduce the described effects, different methods already exist. The state of the art approach is to use an adaptive filter to estimate the acoustic path utilizing methods like

the normalized least mean square (NLMS) algorithm or a Kalman filter [7]. However, besides the fact that operating in a closed acoustic loop requires a sophisticated control mechanism for a robust application of adaptive filters, there are some more limitations. One is that the filter converges towards a bias due to the high correlation between the local and the excitation signal. This makes an additional decorrelation stage essential for many approaches. Another limitation is that usually a filter with limited length will be used when implementing the adaptive filter. Consequently, the filter must be designed in such a way that its length covers at least the most important part of the room impulse response (RIR). Sometimes this is not possible, especially in cases of multichannel applications where a multitude of filters have to be implemented or in cases when the reverberation time is long, e.g., for large rooms like concert halls. The other limitation that leads to feedback never being completely removed is that there will always be a residual misalignment in the adaptive filter, which in turn leads to an error in the estimated feedback signal. Figure 1 depicts an example for a true

*Correspondence: gus@tf.uni-kiel.de

[†]Marco Gimm and Philipp Bulling contributed equally to this work.

¹Digital Signal Processing and System Theory, Kiel University, Kaiserstr. 2, 24143 Kiel, Germany

Full list of author information is available at the end of the article

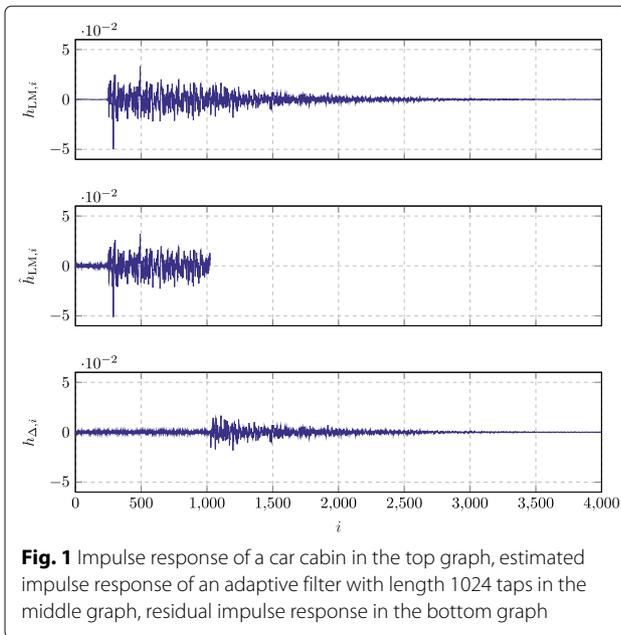


Fig. 1 Impulse response of a car cabin in the top graph, estimated impulse response of an adaptive filter with length 1024 taps in the middle graph, residual impulse response in the bottom graph

impulse response $h_{LM,i}$ as well as the part $\hat{h}_{LM,i}$ that an adaptive filter has estimated. The bottom plot shows the difference between the impulse response $h_{LM,i}$ and $\hat{h}_{LM,i}$. The estimation error $h_{\Delta,i}$ is the impulse response which causes the residual feedback in such a system.

A different method to increase the stability gain in electro-acoustic loops is to estimate the short-term power spectral density (PSD) of the feedback by using the energy envelopes of the room's subband impulse responses. These envelopes can be obtained by a priori or online measurements as well as simulations. With this information, a model can be derived which is then used for a convolution with the loudspeaker subband power signal. This results in an estimate of the feedback's short-term PSD. This estimate is then used within a so-called Wiener filter (or a variant of it) to attenuate the feedback components within the microphone signal. Except from online measurements, the envelope is assumed to be constant. However, it can be shown that the model-based methods are robust against room changes and that the envelopes vary only slightly over time. The main advantage of this method is that the model can be implemented recursively and, thus, very efficiently in terms of computational complexity. There will not be any length limitations as described when using adaptive finite impulse response (FIR) filters. However, there are disadvantages, too. The main one lies in the derivation of the Wiener filter, which assumes that both the desired and the undesired signals are orthogonal. In the presented application this is not the case, since the feedback signal (undesired) is only a delayed and processed version of the local speech signal (desired). This means that not only feedback will be

reduced, the model-based approach will also affect the desired signal. However, due to the fact that speech is assumed to be short-time stationary and there is a delay in the processing and also in the path between loudspeaker and microphone, it is usually observed that the attenuation of the desired signal is small compared to the attenuation of the feedback signal. Hence, this method is able to increase the stability in closed-loop systems.

In [8], we presented a method that makes use of the advantages of both described systems. Therefore, we introduced three ways to estimate the residual feedback PSD recursively, taking an adaptive filter into account. We also compared this with the model-based feedback suppression which was presented in [9]. In this work we made some improvements regarding the models. Furthermore, we show more implementation details. The objective evaluation was improved by adjusting the features. Additional simulations were also performed to investigate the performance during room changes and the convergence of the adaptive filter in the presence of a postfilter. In addition, further acoustic paths were simulated.

1.1 Organization of this paper

The paper is organized as follows: after this introduction, previous research work is summarized in Section 2. Afterwards the model-based feedback suppression approach is explained in Section 3. In Section 4, we show how we adapted the model-based approach to use it as a postfilter. After that, we present different methods to derive the required model parameters in Section 5. Finally, we show the evaluation procedure in Section 6 before a conclusion is provided in Section 7.

1.2 Notation

Throughout this contribution the notation will follow some basic rules:

- Scalar quantities such as time-domain signals are written in lowercase, non-bold letters such as $s(n)$ for a signal at time index n .
- Short-term frequency-domain quantities are described by upper case letters such as $X(\mu, k)$, with k being the frame index and μ as frequency index.
- Vectors are noted as bold letters, e.g., $\mathbf{H}(\mu, k)$ represents a vector containing filter coefficients in subband μ at frame index k .
- Smoothed signals are noted by over-lined letters such as $\bar{x}(n)$ and estimated signals are written as letters with a hat such as $\hat{x}(n)$.
- All signals are represented in discrete time.

2 Previous and related work

Electro-acoustic feedback is a challenge in various technical systems. The most prominent ones are hearing aids,

public address systems, and in-car communication systems. Therefore, lots of research has been done in those domains in recent years. A comprehensive overview of different approaches regarding feedback suppression can be found in [10]. In this work, we will focus on room modeling methods.

To fully erase the feedback and, therefore, to allow arbitrary gains, the impulse response of the feedback path must be estimated by means of an adaptive filter. Early approaches use a *standard* echo canceler to fulfill this task [11–13]. Here, the impulse response is estimated e.g. with a normalized least mean square (NLMS) algorithm in the time domain. If the local signal and the excitation signal are correlated, the problem is that adaptive filters converge to a biased solution. This is strongly the case in closed-loop electro-acoustic systems [14].

One solution to overcome this problem is to decorrelate the signals. This can, for example, be realized by frequency shifting. It is shown in [15–17] that a slight frequency shift within the frequency range of speech is sufficient to improve the convergence of the adaptive filter. The signals can also be decorrelated with linear prediction or pre-whitening [18]. In addition to the decorrelation of the signals, a special step-size control can further improve the convergence of the adaptive filter. In [1, 19], the decorrelation methods frequency shift and pre-whitening are compared and combined with a step-size control, based on a derivation of the so-called *pseudo-optimal step size*. Another step-size control that is able to improve the convergence of the adaptive filter without the need of any decorrelation method is described in [20, 21]. Here, the reverberation of the system is exploited to adapt the filter, since signals are not correlated during reverberation. With this step-size control, both stability and speed of convergence can be improved also for high system gains.

One drawback of the feedback cancellation approaches is that the adaptive filter must cover the relevant length of the room's impulse response. Otherwise, residual feedback is audible and may even cause the system to become unstable. Since long filters increase the computational complexity as well as the convergence time, short filter lengths are often preferred.

In the field of acoustic echo cancellation (AEC), postfilters based on frequency-domain Wiener filters are commonly used [22–25]. The idea is that the residual echo is nothing but the undisturbed error signal which is the signal after the subtraction of the AEC took place assuming the absence of any local speech and noise signals. A very similar approach was already used for residual feedback suppression [13]. The downside of this technique is that the PSD estimation should only be done in remote single talk conditions [26]. Such a situation does not exist in case of closed-loop systems. There is however one exception

which is the end of a speech segment when there is still some power in the loop due to the loop delay.

In [9], the authors present a feedback suppression method based on well known speech dereverberation techniques [27, 28]. Here, the feedback path is modeled with an statistical model. Based on this the feedback's PSD is estimated.

In [8], the model-based feedback suppression is tailored in a way that it can be used as a residual feedback suppression in combination with an adaptive feedback canceler. Therefore, three adapted statistical models have been proposed which can be used to model the feedback path taking an adaptive filter into account. Model-based approaches have already been used in adaptive echo cancellation systems [29, 30]. In [31], the authors also use a model-based approach as a postfilter for adaptive echo cancellation. The idea is to use adaptive approaches to model the residual echo power spectral density. However, in all of these approaches the adaptive filter is assumed to work perfectly and only the acoustic path, which is not covered by the filter is taken into account. In AEC applications, this might be sufficient as reasonable steady-state performance can be reached. However, this is not the case in adaptive feedback cancellation.

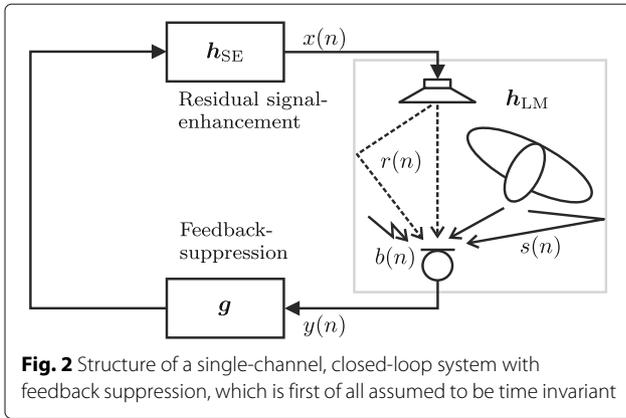
In the presented paper, the adapted models for residual feedback suppression are further investigated and a more detailed insight, as well as more simulations and results are given.

3 Model-based feedback suppression

In [9] it was shown that room dereverberation techniques as they were introduced in e.g. [27, 28] can be used to increase the stability of closed electro-acoustic loops as we face them in ICC systems. In this section, the model-based feedback suppression will be described before adapting it for a system with feedback cancellation. We will start with linear, time-invariant systems with coefficient index i . Of course, we can assume here only short-term stationarity. Therefore, we will introduce time-variance (by adding also a frame index k) after this generic view on the entire system.

A simple example of a time-domain system operating in a closed electro-acoustic loop can be seen in Fig. 2.

The signal $y(n)$ is the microphone signal at time index n and \mathbf{g} is a Wiener filter with coefficients based on the estimated feedback which is used to suppress the recorded feedback that is present in $y(n)$. \mathbf{h}_{SE} is the impulse response that belongs to the system of the individual application, where SE stands for *signal enhancement*. It differs with the individual application and may include noise suppression in case of an ICC system or an equalization filter in case of a public address (PA) system. After the signal enhancement stage, $x(n)$ is played back



using a loudspeaker resulting in a feedback $r(n)$. The latter is obtained by a convolution of $x(n)$ with the room impulse response h_{LM} . As mentioned above, the room impulse response is assumed to be constant for now. Thus, the time index n can be dropped and we obtain the feedback signal as

$$r(n) = \sum_{i=0}^{\infty} x(n-i) h_{LM,i}. \quad (1)$$

$r(n)$ is recorded again by the microphone together with additional background noise $b(n)$ and local speech $s(n)$, yielding the microphone signal

$$y(n) = s(n) + b(n) + r(n). \quad (2)$$

The transfer function of the entire electro-acoustic loop can be described as follows:

$$H_L(e^{j\Omega}) = \frac{H_{SE}(e^{j\Omega}) G(e^{j\Omega})}{1 - H_{SE}(e^{j\Omega}) G(e^{j\Omega}) H_{LM}(e^{j\Omega})}. \quad (3)$$

Here, $\Omega = 2\pi f/f_s \in [0, 2\pi)$ is the frequency f normalized with respect to the sampling frequency f_s . Furthermore,

$$C(e^{j\Omega}) = H_{SE}(e^{j\Omega}) G(e^{j\Omega}) H_{LM}(e^{j\Omega}) \quad (4)$$

is the so-called *open-loop gain*. For stability reasons

$$|C(e^{j\Omega})| < 1 \quad \forall \Omega \quad (5)$$

and

$$\angle C(e^{j\Omega}) \neq \eta 2\pi \quad \forall \Omega, \eta \in \mathbb{Z}_0^+ \quad (6)$$

should hold. $G(e^{j\Omega})$ can be used to fulfill this condition. One possibility would be to choose the transfer function based on the estimated feedback

$$G(e^{j\Omega}) = 1 - \frac{\hat{S}_{rr}(e^{j\Omega})}{\hat{S}_{yy}(e^{j\Omega})}. \quad (7)$$

Therefore, an estimate of the feedback PSD $\hat{S}_{rr}(e^{j\Omega})$ is required, which can be derived from Eq. (1).

To make this method more robust against small variation in h_{LM} and to get the ability to save computational complexity we define a model of h_{LM} based on its so-called *reverberation time* T_{60} and some other parameters, which will be explained next.

This finally results in an exponentially decaying model of the power envelope of the subband version of $H_{LM}(\mu)$

$$|H_{LM,mod,A}(\mu, k)|^2 \quad (8)$$

$$= \begin{cases} 0, & \text{for } k < P(\mu), \\ A(\mu) e^{-\gamma(\mu)(k-P(\mu))}, & \text{for } k \geq P(\mu), \end{cases} \quad (9)$$

where μ is the discrete subband index, k is the frame index and $A(\mu)$ are coupling factors that describe the coupling properties of the acoustic path. $P(\mu)$ is the delay of the acoustic path in frames and

$$\gamma(\mu) = \frac{2 \cdot 3 \ln(10)L}{T_{60}(\mu)f_s} \quad (10)$$

describes the decay behavior, where L denotes the frameshift in samples. Using Eq. (8) the estimated short-time PSD of the current feedback $\hat{S}_{rr,A}(\mu, k)$ can be calculated as the convolution of the short-time PSD of the loudspeaker signal $\hat{S}_{xx}(\mu, k)$ with the magnitude square of the modeled subband impulse response

$$\begin{aligned} \hat{S}_{rr,A}(\mu, k) &= \sum_{i=P(\mu)}^{\infty} \hat{S}_{xx}(\mu, k-i) A(\mu) e^{-\gamma(\mu)(i-P(\mu))} \\ &= A(\mu) \hat{S}_{xx}(\mu, k-P(\mu)) \\ &\quad + e^{-\gamma(\mu)} \hat{S}_{rr,A}(\mu, k-1). \end{aligned} \quad (11)$$

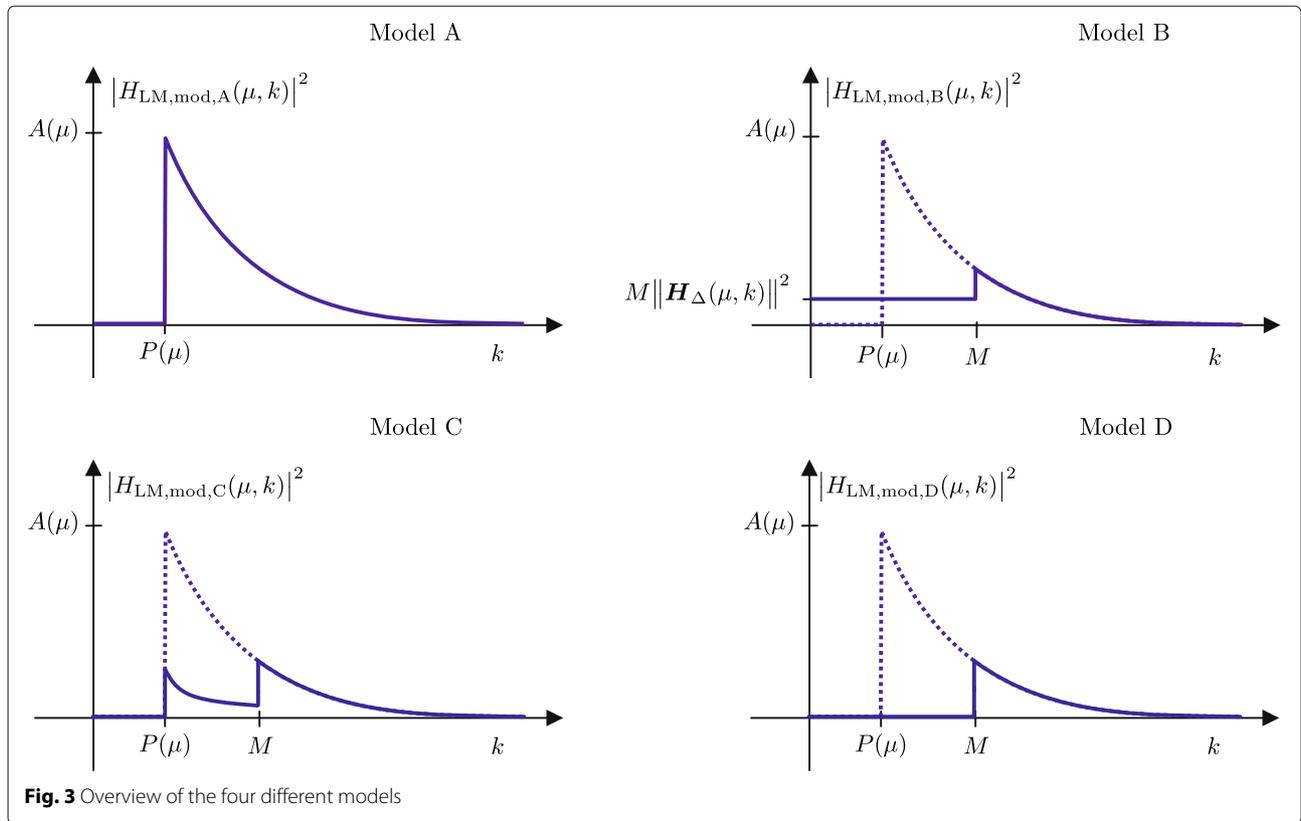
On the top left of Fig. 3, the energy envelope of the modeled subband impulse response for a single subband is depicted. This can now be used for a subband version

$$G_X(\mu, k) = \max \left\{ 0, 1 - \frac{\hat{S}_{rr,X}(\mu, k)}{\hat{S}_{yy}(\mu, k)} \right\} \quad (12)$$

of Eq. (7), with X indicating the individual model type, e.g., X=A.

4 Model-based feedback suppression as postfilter

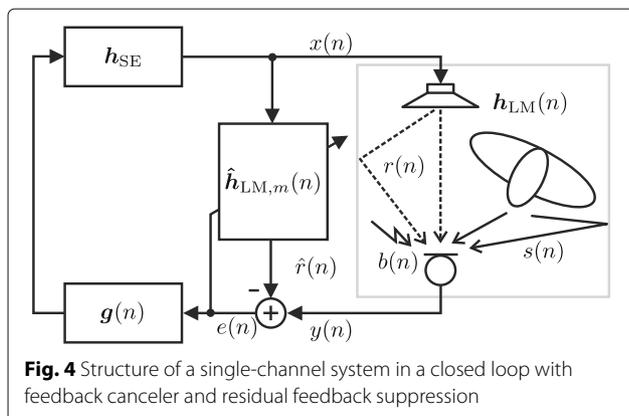
Due to stability reasons, FIR filters are commonly used in adaptive filter applications like echo- or feedback cancellation. If this kind of method is used in a closed electro-acoustic loop system, it is capable of subtracting parts of the feedback signal $r(n)$ from the microphone signal $y(n)$ depending on how good it is adapted to the true room impulse response. However, there are some limitations in the steady-state performance. One of them is that there will always be a residual system mismatch which is caused by non-optimal control or estimation errors, even if robust adaptive control schemes are used. If an efficient



implementation in the subband domain is chosen, the performance is also limited due to aliasing effects caused by the filter banks. The other limitation is due to the part of the true room impulse response which cannot be covered by the adaptive filter. This happens because the FIR filter needs to be implemented with a fixed length, which is often restricted by computational complexity.

Here, a postfilter is usually used to suppress the parts of the feedback which remain after a feedback cancellation approach as it is depicted in Fig. 4.

The idea is to use the method proposed in the previous section and adapt it, so it can be used as a postfilter.



Because of its recursive nature, the model will also cover rooms with a long reverberation time without significant impact on the complexity. However, it has to be adapted with respect to the presence of the adaptive filter. Therefore, the effective impulse response, which is a combination of the true impulse response and the one estimated by the adaptive filter, needs to be computed to derive a new model. The effective impulse response can be derived using the signal $e(n)$ from Fig. 4. For this signal holds:

$$\begin{aligned}
 e(n) &= s(n) + b(n) + \sum_{i=0}^{\infty} x(n-i) h_{LM,i}(n) - \hat{r}(n) \\
 &= s(n) + b(n) + \sum_{i=0}^{\infty} x(n-i) h_{LM,i}(n) - \mathbf{x}_m^T(n) \hat{\mathbf{h}}_{LM}(n) \\
 &= s(n) + b(n) + \sum_{i=0}^{\infty} x(n-i) h_{LM,Res,i}(n).
 \end{aligned} \tag{13}$$

Here,

$$\hat{\mathbf{h}}_{LM}(n) = [\hat{h}_{LM,0}(n), \hat{h}_{LM,1}(n), \dots, \hat{h}_{LM,m-1}(n)]^T \tag{14}$$

is a vector containing the adaptive filter coefficients and

$$\mathbf{x}_m(n) = [x(n), x(n-1), \dots, x(n-m+1)]^T \tag{15}$$

contains the m latest samples of $x(n)$, where m is the number of filter coefficients in the adaptive filter. With this the effective impulse response can be derived as:

$$\begin{aligned} h_{\text{LM,Res},i}(n) &= \begin{cases} h_{\text{LM},i}(n) - \hat{h}_{\text{LM},i}(n), & \text{for } 0 \leq i < m, \\ h_{\text{LM},i}(n), & \text{else,} \end{cases} \\ &= \begin{cases} h_{\text{LM},\Delta,i}(n), & \text{for } 0 \leq i < m, \\ h_{\text{LM},i}(n), & \text{else.} \end{cases} \end{aligned} \quad (16)$$

As one may expect, knowledge about the actual system mismatch for each subband is necessary as it has an influence on the amount of residual feedback. Another difference compared to model A is that there is a direct connection from the loudspeaker signal to the error signal now which is caused by the misalignment in the adaptive filter.

However, this is unknown and has to be estimated. Furthermore, assumptions about the shape of the system misalignment over filter taps has to be made.

During the time when adaptive filters were being studied very extensively—decades ago—two different ideas of the progress of the filter coefficients during an adaptation period co-existed—and still do so today.

4.1 Model B

One idea is that adaptive algorithms spread the error more or less equally over all coefficients. As a consequence, the system mismatch vector can be modeled as a white process with zero mean and a time-variant, but lag-independent variance. Our investigations showed that this seems to be correct if the filter is well converged. So the first approach is to model a constant system mismatch for all filter taps, yielding a power envelope of the residual subband impulse response:

$$\begin{aligned} &|H_{\text{LM,mod,B}}(\mu, k)|^2 \\ &= \begin{cases} M \|\mathbf{H}_{\Delta}(\mu, k)\|^2, & \text{for } 0 \leq k \leq M-1, \\ A(\mu) e^{-\gamma(\mu)(k-P(\mu))}, & \text{for } M \leq k. \end{cases} \end{aligned} \quad (17)$$

In this work, this approach is named model B. M is the filter length in frames. In case of subband processing, $\mathbf{H}_{\Delta}(k)$ is a vector containing the system mismatch vector in every subband.

$$\|\mathbf{H}_{\Delta}(\mu, k)\|^2 = \|\mathbf{H}(\mu, k) - \hat{\mathbf{H}}(\mu, k)\|^2 \quad (18)$$

$$= \sum_{j=0}^{M-1} |H(\mu, k-j) - \hat{H}(\mu, k-j)|^2 \quad (19)$$

is a vector containing the norms of the system mismatch vectors. This is often estimated within adaptive control schemes. An overview about several estimation procedures can be found in [32]. The estimated residual

feedback $\hat{S}_{rr,B}(\mu, k)$ can be obtained by convolving this modeled subband impulse response with the PSD of the loudspeaker signal yielding a solution consisting of two parts:

$$\begin{aligned} \hat{S}_{rr,\text{Res},B}(\mu, k) &= \sum_{i=0}^{M-1} \hat{S}_{xx}(\mu, k-i) M \|\mathbf{H}_{\Delta}(\mu, k)\|^2 \\ &\quad + \sum_{i=M}^{\infty} \hat{S}_{xx}(\mu, k-i) A(\mu) e^{-\gamma(\mu)(i-P(\mu))} \\ &= \sum_{i=0}^{M-1} \hat{S}_{xx}(\mu, k-i) M \|\mathbf{H}_{\Delta}(\mu, k)\|^2 \\ &\quad + \sum_{i=0}^{\infty} \hat{S}_{xx}(\mu, k-i-M) A(\mu) e^{-\gamma(\mu)(i-P(\mu)+M)} \\ &= \hat{S}_{xx,\text{rec}}(\mu, k) M \|\mathbf{H}_{\Delta}(\mu, k)\|^2 + \hat{S}_{mm}(\mu, k). \end{aligned} \quad (20)$$

Each of these two parts can be calculated recursively, leading to a very low computational complexity:

$$\begin{aligned} \hat{S}_{xx,\text{rec}}(\mu, k) &= \hat{S}_{xx,\text{rec}}(\mu, k-1) \\ &\quad + \hat{S}_{xx}(\mu, k) - \hat{S}_{xx}(\mu, k-M) \end{aligned} \quad (21)$$

$$\begin{aligned} \hat{S}_{mm}(\mu, k) &= e^{-\gamma(\mu)} \hat{S}_{mm}(\mu, k-1) \\ &\quad + \hat{S}_{xx}(\mu, k-M) A(\mu) e^{-\gamma(\mu)(M-P(\mu))}. \end{aligned} \quad (22)$$

In the top right of Fig. 3, the energy envelope of the subband-impulse response for a single subband of model B compared to model A (introduced in the previous section) can be seen.

4.2 Model C

The second idea is that the system mismatch of the individual coefficients is more or less proportional to the magnitude of the room impulse response of the system that should be identified. This behavior is also observable, but mainly at the beginning of adaptation processes or—in general—whenever the filter is not well adapted. Since feedback cancellation approaches for ICC systems face generally hard conditions e.g. permanent double-talk and high background noise levels, this model would be an option here.

To model this we assume the first interval for the direct part of the residual impulse response to be zero. This is usually the case only when the adaptive filter is initialized. After some iterations the coefficients will differ from zero. However, the system mismatch in this interval will always stay small compared to the interval between the largest coupling and the rest of the adaptive filter. Here, the system mismatch is modeled as exponentially decaying with the same T_{60} as it is used in all other approaches. Furthermore, we introduce $Q(\mu)$ which is the power of the

maximum value in the system mismatch vector. This leads to the following power envelope of the model

$$\begin{aligned} & |H_{LM,mod,C}(\mu, k)|^2 \\ &= \begin{cases} 0, & \text{for } k < P(\mu), \\ Q(\mu, k) e^{-\gamma(\mu)(k-P(\mu))}, & \text{for } P(\mu) \leq k < M, \\ A(\mu) e^{-\gamma(\mu)(k-P(\mu))}, & \text{for } M \leq k. \end{cases} \end{aligned} \quad (23)$$

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

$$\begin{aligned} & \hat{S}_{rr,Res,C}(\mu, k) \\ &= \sum_{i=P(\mu)}^{M-1} \hat{S}_{xx}(\mu, k-i) Q(\mu) e^{-\gamma(\mu)(i-P(\mu))} \\ &+ \sum_{i=M}^{\infty} \hat{S}_{xx}(\mu, k-i) A(\mu) e^{-\gamma(\mu)(i-P(\mu))} \quad (24) \\ &= e^{-\gamma(\mu)} \hat{S}_{rr,Res,C}(\mu, k-1) \\ &+ \hat{S}_{xx}(\mu, k-P(\mu)) Q(\mu, k) \\ &+ \hat{S}_{xx}(\mu, k-M) (A(\mu) - Q(\mu)) e^{-\gamma(\mu)(M-P(\mu))}. \end{aligned} \quad (25)$$

4.3 Model D

The drawback of the two proposed models is that the performance will depend on how accurate the estimation of the system distance is. An easier method compared to models B and C is to assume that the adaptive filter operates perfectly well and there is consequently only the length limitation which has to be covered by the postfilter. This would be advantageous, but it is not very realistic in practical approaches. Models B and C are better in this regard. In this case, $\|\mathbf{H}_{\Delta}(\mu, k)\|^2$ can be set to zero and the estimated PSD of the feedback signal simplifies to

$$\begin{aligned} & \hat{S}_{rr,Res,D}(\mu, k) \\ &= e^{-\gamma(\mu)(M-P(\mu))} A(\mu) \hat{S}_{xx}(\mu, k-M) \\ &+ e^{-\gamma(\mu)} \hat{S}_{rr,Res,D}(\mu, k-1), \end{aligned} \quad (26)$$

which corresponds to $\hat{S}_{mm}(\mu, k)$ in Eq. (22).

5 Model parameters

To use the proposed model-based approach, a priori knowledge about the room is needed. In this work, the room is assumed to be *power stationary*, meaning that the power envelope of the room impulse response does not vary much over time. Depending on the particular control mechanism used for the adaptive filter, a correction of the model parameters based on the adaptive filter is also possible. The parameters can be extracted using a measured impulse response like it was proposed in [33].

To follow this approach, the time-domain impulse response has to be transformed into the subband domain

$$H_{LM}(\mu, k) = \sum_{\kappa=0}^{N-1} h_{ana,\kappa} h_{LM,kL+\kappa} e^{-j2\pi \frac{\mu\kappa}{N}}, \quad (27)$$

where N is the window length and, thus, the length of the DFT, and \mathbf{h}_{ana} is the window function used in the filter bank. The absolute value of the subband impulse response is smoothed along the frequency axis in both positive (Eq. (28)) and negative (Eq. (29)) direction for every frame with the smoothing constant ζ :

$$\begin{aligned} & \tilde{H}_{LM}(\mu, k) \\ &= (1 - \zeta) \tilde{H}_{LM}(\mu - 1, k) + \zeta |H_{LM}(\mu, k)| \end{aligned} \quad (28)$$

$$\begin{aligned} & \bar{H}_{LM}(\mu, k) \\ &= (1 - \zeta) \bar{H}_{LM}(\mu + 1, k) + \zeta \tilde{H}_{LM}(\mu, k). \end{aligned} \quad (29)$$

This way, a zero-phase low-pass filter is realized to reduce the variance along the frequency axis. As a first step, the delay can be determined for each subband by finding the index of the first maximum value of the smoothed magnitude subband impulse response in each subband

$$P(\mu) = \underset{k \in \{0, 1, \dots, L_{LM}-1\}}{\operatorname{argmax}} \left\{ \bar{H}_{LM}(\mu, k) \lambda^k \right\}, \quad (30)$$

with L_{LM} representing the considered length of the impulse response in frames and λ^k representing an exponentially decaying series with $\lambda \in (0, 1)$, which can be used to avoid choosing late constructive interferences as maxima.

Next, a vector of length M_{LM} is defined for every subband. It contains the logarithmic impulse responses, starting at the delay which was found before:

$$\begin{aligned} & \check{\mathbf{h}}_{LM}(\mu) = \left[\ln(\bar{H}_{LM}(\mu, P(\mu))), \dots, \right. \\ & \left. \ln(\bar{H}_{LM}(\mu, M_{LM} + P(\mu))) \right]^T. \end{aligned} \quad (31)$$

This can be modeled linearly for each subband

$$\begin{aligned} & \check{\mathbf{h}}_{LM}(\mu, k') = \tilde{A}(\mu) + \tilde{\gamma}(\mu) k' \\ & k' \in \{0, 1, \dots, M_{LM} - P(\mu) - 1\}. \end{aligned} \quad (32)$$

Written in matrix notation, the logarithmic subband impulse responses can be combined to

$$\check{\mathbf{h}}_{LM} = \mathbf{\Gamma} \Theta, \quad (33)$$

where $\mathbf{\Gamma}$ is the observation matrix and Θ represents the parameter vector

$$\mathbf{\Gamma} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ \vdots & \vdots \\ M_{LM} - P(\mu) - 1 & 1 \end{bmatrix}, \quad \Theta = \begin{bmatrix} \tilde{\gamma}(\mu) \\ \tilde{A}(\mu) \end{bmatrix}. \quad (34)$$

To find the unknown parameters in Θ , Eq. (33) can be modified according to

$$\Theta = \left[\Gamma^T \Gamma \right]^{-1} \Gamma^T \check{h}_{LM}, \quad (35)$$

with $\left[\Gamma^T \Gamma \right]^{-1} \Gamma^T = \Gamma^\dagger$ being the pseudo inverse of Γ . The last step is to convert the logarithmic values to linear values

$$\gamma(\mu) = e^{\tilde{\gamma}(\mu)}, \quad (36)$$

$$A(\mu) = e^{\tilde{A}(\mu)}. \quad (37)$$

A simplification, which we also used for our simulations, would be to assume the model parameters except the coupling factors to be identical for all frequencies. In this case, the delay and the decay constants can be computed directly from the energy decay curve (EDC) defined as

$$\text{EDC}(i) = \frac{\sum_{i'=i}^{N_{LM}-1} |h_{LM,i'}|^2}{\sum_{i'=0}^{N_{LM}-1} |h_{LM,i'}|^2}. \quad (38)$$

The EDC describes the remaining energy in the system at time instance i normalized to 0 dB. An example of an EDC measured in a car cabin can be seen in Fig. 5. The delay T_D is the time instant where the EDC begins to drop. Using this value, the delay in frames can be derived as

$$P(\mu) \approx P = \left\lfloor T_D \cdot \frac{f_S}{L} \right\rfloor \quad \forall \mu, \quad (39)$$

with [...] denoting rounding towards the next smaller integer. The reverberation time T_{60} is the time instant when the energy that remains in the tail of the room impulse response reaches -60 dB. Often, this cannot be found in the EDC, because there is measurement noise dominating when using a measured impulse response. In this case, the EDC has to be extrapolated linearly, as it was done in the

example shown in Fig. 5. Using this value the decay instant can be derived:

$$\gamma(\mu) \approx \gamma = \frac{2 \cdot 3 \ln(10) L}{T_{60} f_S} \quad \forall \mu. \quad (40)$$

The remaining parameters which need to be estimated are the coupling factors. These are the absolute squared values of the maximum of the subband impulse response, which can be found at frame index $k = P$ according to Eq.(30). Therefore, the smoothed version of Eq.(29) should be used to reduce the variance along the frequency axis to yield

$$A(\mu) = |\bar{H}_{LM}(\mu, P)|^2. \quad (41)$$

In Fig. 6, a time-frequency analysis of the measured impulse response and the corresponding model can be seen.

6 Results and discussion

The proposed schemes were tested in an ICC application. For this, an impulse response measured in a van was used. It is the same one as shown in Fig. 5. The complete setup is shown in Fig. 7.

For our simulation we used clean speech signals from different male and female speakers sampled at 44.1 kHz. The DFT order was set to $N = 512$ and we used a frameshift of $L = 256$. The NLMS-based adaptive filter

$$\hat{H}(\mu, k) = \hat{H}(\mu, k-1) + \alpha \frac{E^*(\mu, k-1)X(\mu, k-1)}{\|X(\mu, k-1)\|^2} \quad (42)$$

with a fixed step-size α and $E(\mu, k)$ and $X(\mu, k)$ being the error signal or the excitation signal vector, respectively, was adapted until a specific system distance for all subbands was obtained. Afterwards the adaption was stopped by setting α to zero. As the system distance for this particular simulation was known, we used this value also as a parameter in model B and C to avoid the influence of

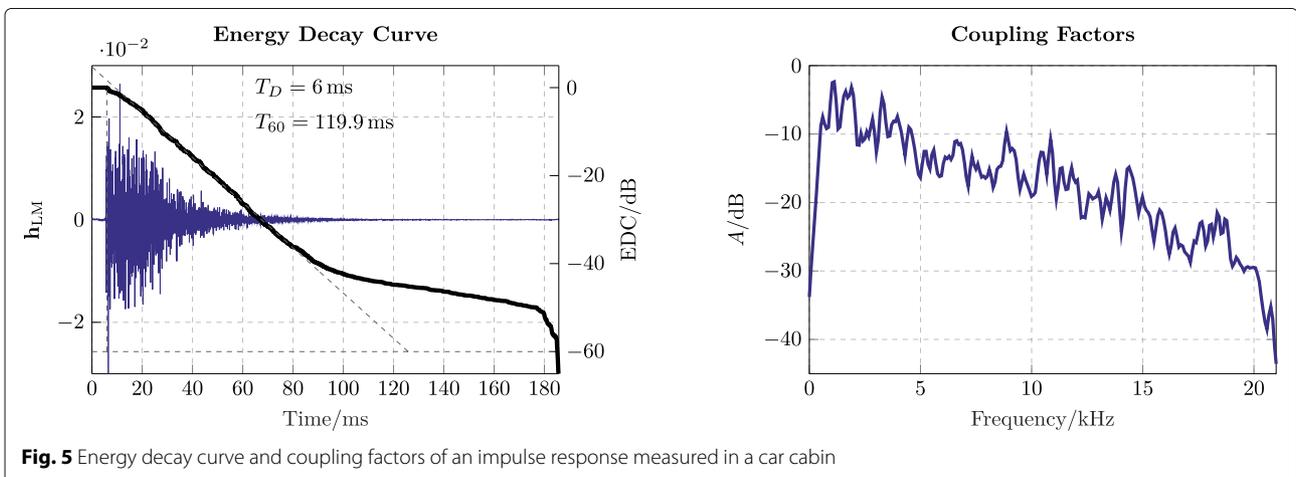


Fig. 5 Energy decay curve and coupling factors of an impulse response measured in a car cabin

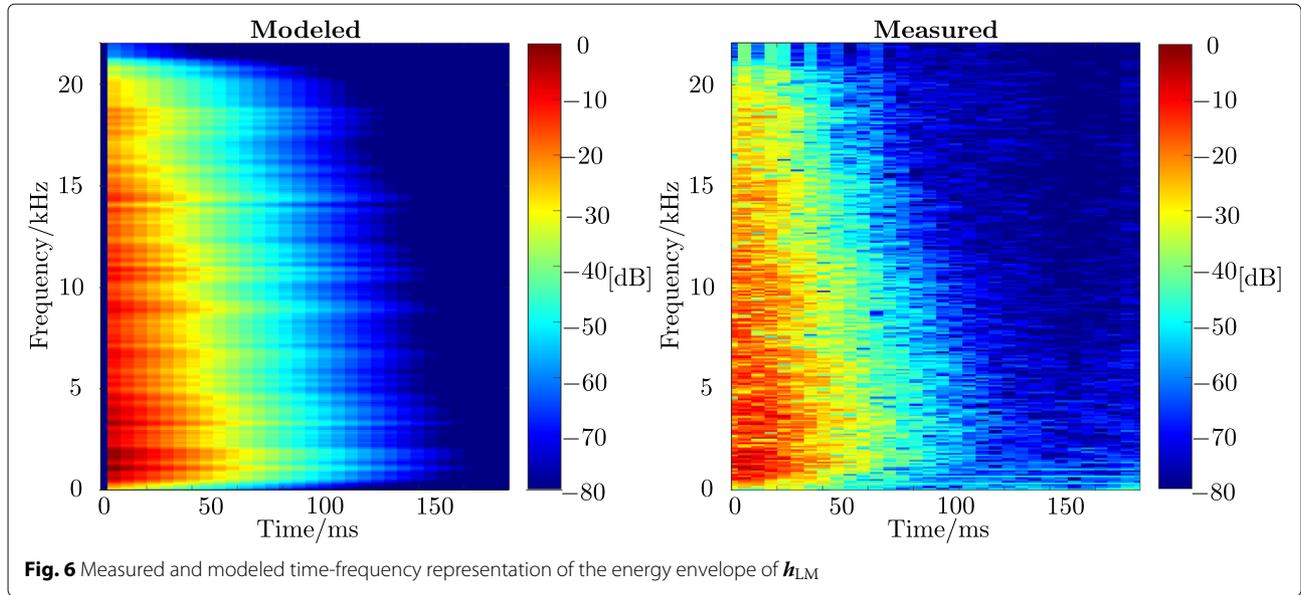


Fig. 6 Measured and modeled time-frequency representation of the energy envelope of h_{LM}

estimation errors. M was set to four frames, corresponding to a filter length of 46.4 ms. Afterwards the step-size was set to zero for the simulation. This was done since the models would affect the convergence behavior of the adaptive filter, so the results would also be affected. The aim of the postfilter is to reduce the residual feedback in the microphone signal as much as possible without affecting the desired speech signal. To prove this, the ratio of the mean logarithmic speech power and the mean weighted logarithmic speech power was calculated as

$$P_s = \frac{\sum_{k=0}^{N_{\text{Frames}}-1} \sum_{\mu=\mu_{\text{Start}}}^{\mu_{\text{Stop}}} \tilde{P}_s(\mu, k)}{\sum_{k=0}^{N_{\text{Frames}}-1} \sum_{\mu=\mu_{\text{Start}}}^{\mu_{\text{Stop}}} \Phi_s(\mu, k)} \quad (43)$$

with

$$\begin{aligned} \tilde{P}_s(\mu, k) &= 10 \log_{10} \left(\frac{S_{ss}(\mu, k) \Phi_s(\mu, k)}{S_{ss}(\mu, k) \Phi_s(\mu, k) G_X^2(\mu, k)} \right) \\ &= 10 \log_{10} \left(\frac{1}{G_X^2(\mu, k)} \right) \Big|_{\Phi_s(\mu, k) \neq 0} \end{aligned} \quad (44)$$

The same can be done for the *reverberation parts* of the signal

$$P_r = \frac{\sum_{k=0}^{N_{\text{Frames}}-1} \sum_{\mu=\mu_{\text{Start}}}^{\mu_{\text{Stop}}} \tilde{P}_r(\mu, k)}{\sum_{k=0}^{N_{\text{Frames}}-1} \sum_{\mu=\mu_{\text{Start}}}^{\mu_{\text{Stop}}} \Phi_r(\mu, k)} \quad (45)$$

with

$$\begin{aligned} \tilde{P}_r(\mu, k) &= 10 \log_{10} \left(\frac{S_{ss}(\mu, k) \Phi_r(\mu, k)}{S_{ss}(\mu, k) \Phi_r(\mu, k) G_X^2(\mu, k)} \right) \\ &= 10 \log_{10} \left(\frac{1}{G_X^2(\mu, k)} \right) \Big|_{\Phi_r(\mu, k) \neq 0} \end{aligned} \quad (46)$$

where N_{Frames} is the total number of processed frames in this simulation and $\mu_z \in [\mu_{\text{Start}}, \mu_{\text{Start}+1}, \dots, \mu_{\text{Stop}}]$ are the investigated subbands representing frequencies between about 90 Hz and 8000 Hz, where most of the speech power is located. $S_{ss}(\mu, k)$ is the short-term power spectrum of the input frame of the clean speech signal. $\Phi_s(\mu, k)$ is a binary mask based on a subband voice activity detection which is one for subbands where voice is detected and zero for those without voice. $\Phi_r(\mu, k)$ is the equivalent for the reverberation signal with the additional condition that all time-frequency bins where $\Phi_s(\mu, k) = 1$ are set to zero. An example of the described masks can be seen in Fig. 8, where the speech signal of the investigated

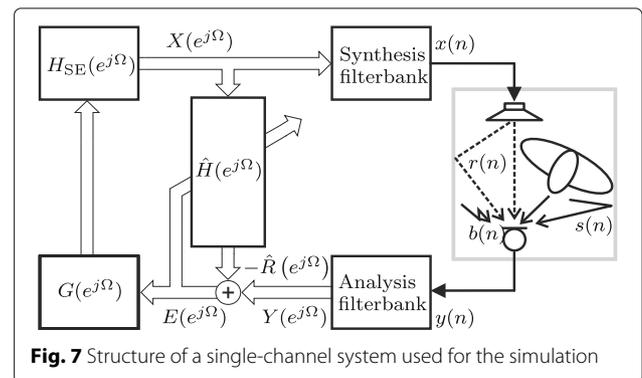
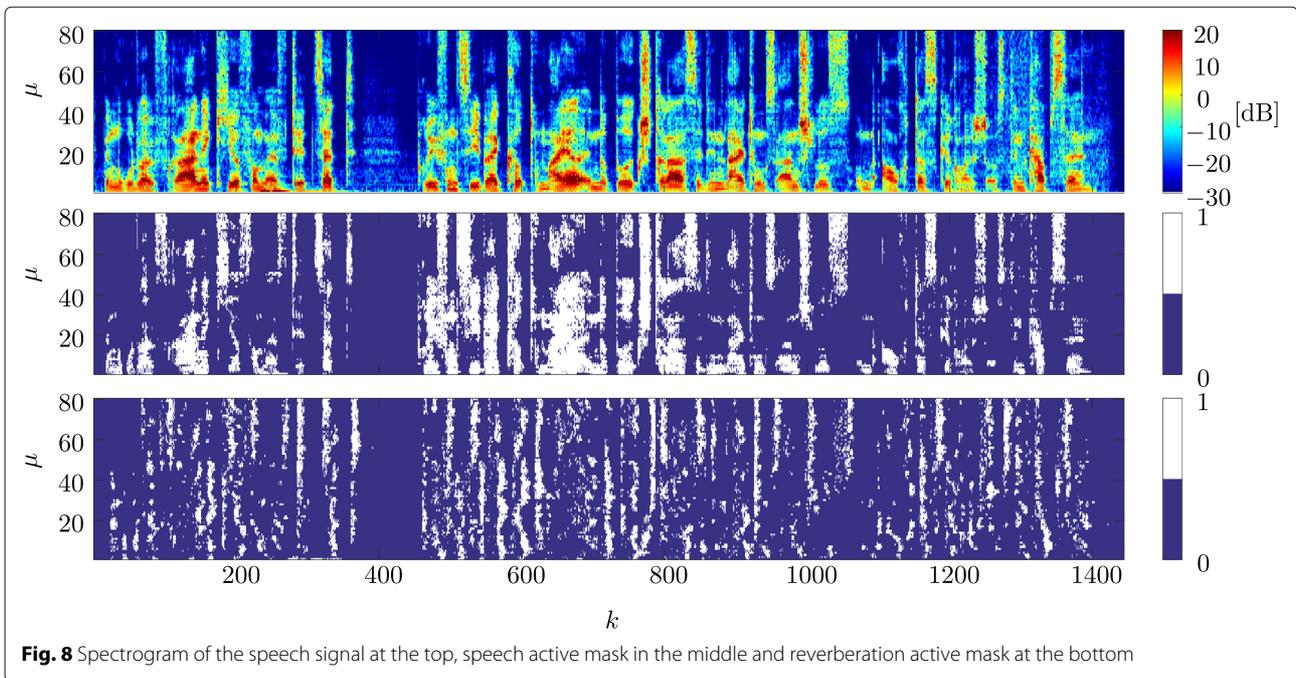


Fig. 7 Structure of a single-channel system used for the simulation



subbands is at the top, the mask for the clean speech is in the middle, and the mask for the reverberation is at the bottom. Using these masks, the impairment is only evaluated where the respective signals are present. The reason for using this method is that we wanted to observe the ratio between speech impairment and feedback suppression as well as both quantities alone. This is important because a large speech impairment could lead to audible artifacts. More established methods like segmental SNR would only show the feedback reduction, which is not sufficient for our purpose.

Furthermore, we used a short-time objective intelligibility (STOI) measure, which was proposed in [34]. This measurement shows good performance in evaluating degradation caused by time-frequency based algorithms e.g. noise reduction. As a reference signal we used the clean speech $s(n)$. The loudspeaker signal $x(n)$ was the signal to be evaluated.

In Table 1, the simulation results for different system distances are shown. P_s is the unwanted impairment of the clean speech signal, P_r is the equivalent for the reverberation signal. $P_r - P_s$ is the distance between both. Since one impairment is wanted and the other one is not, it describes the attenuation of the unwanted signal. Consequently P_s has to be treated as an offset and needs to be compensated.

When the system distance is at -40 dB, the results in terms of the different approaches (B–D) for $P_r - P_s$ as well as STOI are similar. However, it can also be seen that the impairment of the clean speech is significantly higher in case of model A, which is the original unadapted model.

Table 1 Simulation results for the four different models at different system distances

System-distance/dB	Model	P_s /dB	P_r /dB	$(P_r - P_s)$ /dB	STOI
- 40	A	10.72	18.22	7.50	0.86
	B	5.29	15.24	9.95	0.94
	C	5.24	15.24	10.00	0.94
	D	5.24	15.25	10.01	0.94
	-	0	0	0	0.99
- 30	A	10.72	18.46	7.74	0.86
	B	5.64	15.67	10.03	0.94
	C	5.25	15.27	10.02	0.94
	D	5.24	15.27	10.03	0.94
	-	0	0	0	0.99
- 20	A	10.69	18.03	7.34	0.86
	B	7.97	16.66	8.69	0.91
	C	5.35	14.26	8.91	0.94
	D	5.25	13.96	8.71	0.94
	-	0	0	0	0.99
- 10	A	10.77	16.98	6.21	0.84
	B	13.68	18.34	4.66	0.77
	C	6.19	13.05	6.86	0.91
	D	5.14	11.73	6.59	0.92
	-	0	0	0	0.95

Model B, which is assumed to be correct for this simulation because the filter is well adapted and has a constant system mismatch, is very similar to models C and D with respect to P_s and P_r . This happens due to the fact that in case of a small system distance models (B–D) are nearly the same. When the system distance is increased, the impairment of the clean speech caused by models A and D are nearly constant whereas it increases in case of models B and C. This is exactly what one would expect because of the short-time stationary nature of speech leading to a large correlation when reducing the lag between an input frame and the models response. The system distance is used as a weight. By increasing it while decreasing the response time the correlation between the wanted and the unwanted signal increases.

In case of model B, there is an immediate response to an input signal. Model C also produces a response in the region of early feedback, resulting in a higher value of P_s when increasing the system distance. However, $P_r - P_s$ is slightly better than in case of all other models. The best compromise regarding the impairment of the clean speech is reached with models C and D. Evaluating with STOI shows similar results. It can be seen that the results improve when the system distance decreases. For large system distances the scores are low even if there is no postfilter applied (-). This is due to the existing feedback in the processed signal. For a system distance of -10 dB it can also be seen that the score of model D is slightly higher, although $P_r - P_s$ of model C is slightly larger. This is due to the fact that model D causes less speech impairment in this particular setup. Furthermore, a relation between STOI and P_s can be observed. A large value for P_s leads to a low STOI rating, which is due to the fact that STOI evaluates the degradation of the speech signal, which is mainly influenced by P_s . The best results in terms of STOI are reached when there is no postfilter at all. However, this does not mean that it makes no sense to use a postfilter at all, because one of its main purpose is to increase stability while saving computing power.

The impulse responses used for the model-based approach were measured under certain conditions. For example, this could be an empty vehicle at a certain temperature. In reality, however, these are subject to permanent fluctuations due to room changes. For example, a car could be fully loaded and fully occupied or empty. In addition, objects directly in front of sound sources or microphones could cause large attenuation.

Even changes in the distance between loudspeaker and microphone are conceivable. In the following, we will investigate such a situation where the model parameters are determined based on an impulse response of an empty van, but in fact there is a fully occupied interior. This variation of the acoustic path results in a reduced

$T_{60} = 80.3$ ms compared to the original 119.9 ms as well as different coupling factors, which can be seen in Fig. 9.

The results (see Table 2) show that both STOI and ($P_r - P_s$) are only slightly worse than before, but still very good.

In order to have a more robust evaluation, we simulated a second acoustic path. This time it is one that was recorded in a lecture hall and has a significantly longer decay time T_{60} of 777.8 ms. The delay T_D is 18.7 ms. The results can be seen in Table 3. The higher reverberation time results in a slightly higher influence on the desired signal than in the simulation before. However, a clear attenuation of the feedback can still be seen. The previously discussed effects of the different models apply here without restriction.

In order to evaluate the subjective impairment of the desired signal we conducted a listening test with 26 untrained participants aged between 21 and 46 years. We used the same setup as shown before with a system distance of -30 dB.

Overall the procedure was a degradation category rating (DCR) according to ITU-T Rec. P.800 [35], which was modified for our purpose. We always played the unprocessed clean speech signal as reference and then the simulated versions with either one of the models (A–D) or the cancellation only (-) in random order. We used three female and two male speakers saying German sentences according to ITU-T Rec. P.501 [36]. In sum, every participant had to rate 25 signals. One of the female speakers was used for a trial run which we did not take into account to give the participants the opportunity to get used to the test procedure. The signals are provided on a web page [37].

The rating was defined as following:

- 5. Excellent – Speech sounds like the unprocessed signal

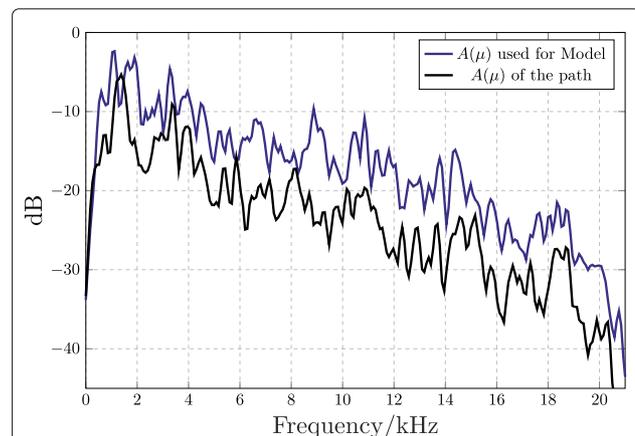


Fig. 9 Coupling factors of the empty car used as model parameters for simulation and true coupling factors of the fully loaded car

Table 2 Simulation results for the four different models at different system distances with incorrect system parameters after a room change

System-distance/dB	Model	P_s /dB	P_r /dB	$(P_r - P_s)$ /dB	STOI
- 40	A	10.72	18.08	7.36	0.86
	B	5.29	14.26	8.97	0.94
	C	5.28	14.22	8.94	0.94
	D	5.28	14.23	8.95	0.94
	-	0	0	0	1.00
- 30	A	10.72	17.90	7.18	0.86
	B	5.41	13.90	8.49	0.94
	C	5.28	13.75	8.47	0.94
	D	5.28	13.73	8.45	0.94
	-	0	0	0	1.00
- 20	A	10.74	17.14	6.40	0.86
	B	6.60	14.00	7.40	0.93
	C	5.30	12.42	7.12	0.94
	D	5.26	12.29	7.03	0.94
	-	0	0	0	1.00
- 10	A	10.73	16.41	5.68	0.86
	B	11.47	16.60	5.19	0.83
	C	5.63	11.24	5.61	0.93
	D	5.20	10.41	5.21	0.93
	-	0	0	0	0.99

- 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

The results in terms of a mean opinion score (MOS) can be seen in Fig. 10.

The unadapted model and the version without any postfilter were rated with a mean opinion score below 2.5. This shows that artifacts caused by the postfilter are as bad as the residual feedback when there is no postfilter at all. All of the adapted models are rated with a MOS between 3.2 and 3.6 which means that the impairment is less compared to the unadapted model. Here model C shows the best results compared to models B and D. However, a Tukey honest significant difference (HSD) test with $\alpha = 0.05$ as suggested in [35] shows that there is no significant difference between models B, C, and D as well as between model A and no postfilter at all. However, the approaches can be grouped in model (B, C, and D) and (A and -).

Table 3 Simulation results for the four different models at different system distances with impulse response of a lecture room

System-distance/dB	Model	P_s /dB	P_r /dB	$(P_r - P_s)$ /dB	STOI
- 40	A	12.94	19.33	6.39	0.80
	B	6.84	17.24	10.40	0.90
	C	6.84	17.24	10.40	0.90
	D	6.84	17.24	10.40	0.90
	-	0	0	0	0.99
- 30	A	12.94	19.25	6.31	0.80
	B	6.90	17.25	10.35	0.90
	C	6.84	17.24	10.40	0.90
	D	6.84	17.24	10.40	0.90
	-	0	0	0	0.99
- 20	A	12.94	18.82	5.88	0.80
	B	7.41	17.07	9.66	0.90
	C	6.88	16.59	9.71	0.90
	D	6.87	16.58	9.71	0.90
	-	0	0	0	0.99
- 10	A	12.94	17.72	4.78	0.80
	B	10.50	17.39	6.89	0.87
	C	6.96	14.80	7.84	0.89
	D	6.92	14.76	7.84	0.89
	-	0	0	0	0.98

The results in Table 4 are consistent with the objective results, with the exception of STOI. In the hearing test, the subjects rated a larger feedback as disturbing as a strong degradation of the speech signal. In contrast, the results according to STOI must be interpreted in such a way that a stronger feedback has less influence than the degradation of speech.

In a next step, we want to evaluate the model influence on the convergence behavior of an adaptive filter. For this, we used an NLMS-based adaptive filter based on pseudo-optimal step-size

$$\alpha_{\text{opt}}(\mu, k) \approx \frac{E\{|E_u(\mu, k)|^2\}}{E\{|E(\mu, k)|^2\}} \quad (47)$$

according to [26], where the expected value $E\{\cdot\}$ was approximated by first order IIR smoothing. The so-called undisturbed error $E_u(\mu, k)$, which is the error signal $E(\mu, k)$ without local signals, must be estimated as well. For this, it is replaced by

$$E_u(\mu, k) = E(\mu, k) - S(\mu, k) - B(\mu, k) \quad (48)$$

$$= \mathbf{H}_{\Delta}^H(\mu, k)\mathbf{X}(\mu, k), \quad (49)$$

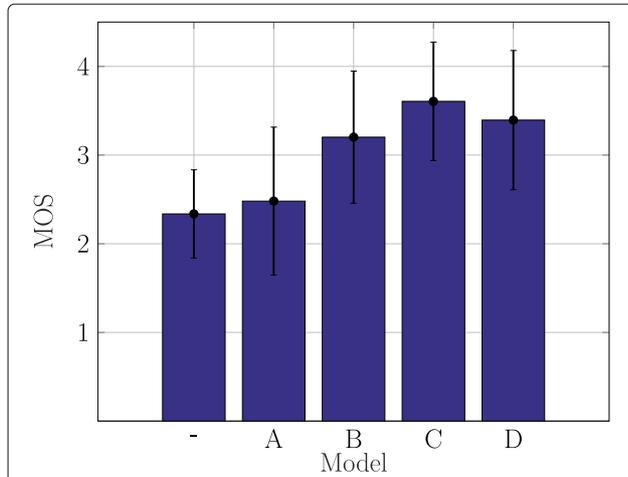


Fig. 10 Results (mean opinion score and variance) of subjective evaluation for the adaptive filter only (–), the “standard” model (A), the model with const. system mismatch (B), the model with an exponentially decaying shape of the system mismatch (C), and the model based on perfect adaptive filter (D)

with $\mathbf{H}_\Delta(\mu, k)$ being still not known. First, we replace it with an estimate yielding

$$\left| \bar{\hat{E}}_u(\mu, k) \right|^2 = |\bar{X}(\mu, k)|^2 \cdot \beta_x^2(\mu, k). \quad (50)$$

In acoustic echo cancellation, $\beta_x^2(\mu, k)$ could be estimated by minimum tracking the power of the noise-reduced error signal, which is then divided by the smoothed power of $X(\mu, k)$. However, this is not possible in feedback cancellation due to the permanent presence of local speech. Here, an approach [38] is to split $\beta_x^2(\mu, k)$ into

$$\beta_x^2(\mu, k) = \beta_{\text{LEM}}^2(\mu, k) \cdot \frac{|E(\mu, k)|^2 - \hat{S}_{bb}(\mu, k)}{|Y(\mu, k)|^2 - \hat{S}_{bb}(\mu, k)} \quad (51)$$

$$= \beta_{\text{LEM}}^2(\mu, k) \cdot \beta_y^2(\mu, k), \quad (52)$$

with $\beta_{\text{LEM}}^2(\mu, k)$ being a pre-measured quantity based on $A(\mu)$ and $\beta_y^2(\mu, k)$ being the smoothed power ratio of the microphone and the error signal.

For the simulation, we used speech signals recorded in a car at 100 km/h. Now we replaced the fixed values of $\|\mathbf{H}_\Delta(\mu, k)\|^2$ with its estimate which we get from the step-size control $\beta_x^2(\mu, k)$. $Q(\mu, k)$ in model C was replaced with $MA(\mu)\beta_x^2(\mu, k)$. The loop gain initially was at 5 dB and was increased with 0.8 dB/second until it reached the final value of 28 dB. The results in terms of the system distance over time with the same adaptive filter and the different models for the postfilter are shown in Fig. 11.

It can be seen that the best performance is reached when there is no postfilter at all. This is due to the fact that the achievable system distance at a fixed step-size depends only on the power ratio between feedback signal and local signal [26]. Even with an adaptive step-size control, as in this case, it does not always work well enough to compensate for this. As mentioned before, this is mainly due to P_s . This value attenuates the desired signal, which reduces the power of the loudspeaker signal by the same amount. To adjust the achievable filter performance, the filtered signal must be amplified by the value P_s , which is shown in Fig. 12.

Here, we adjusted the gain by an offset of 5 dB in case of models C and D and 10 dB in case of model A. For these three models this nearly matches the individual values of P_s . However, this did not work for model B, where we had to add 17 dB to the loop gain. The difference is that P_s of models A, C, and D has only a small or no dependency on

Table 4 Results of Tukey’s HSD test with $\alpha = 0.05$

Model 1	Model 2	Mean difference	Adjusted p value	Reject
(–)	A	0.2596	0.1635	False
(–)	B	1.0481	0.001	True
(–)	C	1.3077	0.001	True
(–)	D	1.125	0.001	True
A	B	0.7885	0.001	True
A	C	1.0481	0.001	True
A	D	0.8654	0.001	True
B	C	0.2596	0.1635	False
B	D	0.0769	0.9	False
C	D	–0.1827	0.5075	False

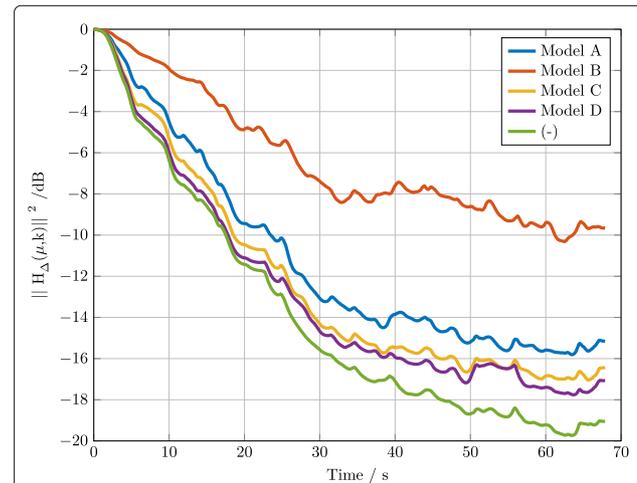
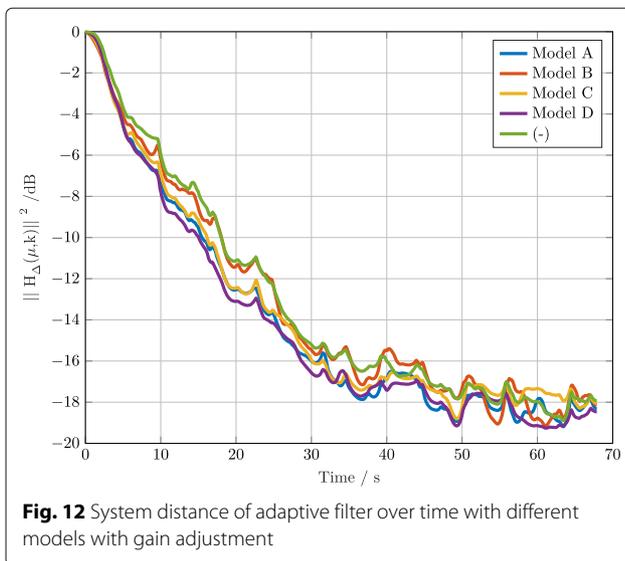


Fig. 11 System distance of adaptive filter over time with different models



the system distance, whereas it increases with decreasing system difference in case of model B.

7 Conclusion

In this work, we investigated existing and proposed slightly extended postfilter schemes which are capable of suppressing residual feedback in closed-loop systems where adaptive feedback cancelers are used. We showed that there are different ways to adapt the reverberation model with respect to the feedback canceler. We were able to show by means of subjective and objective evaluation that all of our adapted models provide a better performance than using the *standard* reverberation model (model A) as a postfilter in a system with acoustic feedback canceler. However, there is a drawback. In the model, it was assumed that knowledge about the current system distance is available. This is, however, a quantity which is not available in real systems. But there are several step-size control methods available where a robust estimation of this quantity is included. In this case, it can also be used for the model-based postfilter. In all other cases, when no estimation of the system distance exists we propose to use the other adapted model, which assumes the system distance to be zero (model D). Beside models A to D several other models could be thought of and some of them were also tested during this research work, but at the end, we decided to continue only with these four approaches to keep this publication at a reasonable length.

Abbreviations

STO: Short-time objective intelligibility; DCR: Degradation category rating; PA: Public address; HSD: Honest significant difference; ICC: In-car communication; MOS: Mean-opinion score; RIR: Room-impulse response; EDC: Energy-decay curve; FIR: Finite impulse response; NLMS: Normalized least mean square; PSD: Power spectral density

Acknowledgements

The authors would like to thank all participants of the hearing test.

Authors' contributions

MG and PB have conducted the research and analyzed the data. MG, PB, and GS authored the paper. All authors read and approved the final manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Availability of data and materials

The sample files which we used for the subjective evaluation can be found online [37].

Declarations

Competing interests

The authors declare that they have no competing interests.

Author details

¹Digital Signal Processing and System Theory, Kiel University, Kaiserstr. 2, 24143 Kiel, Germany. ²Cerence, Soeflinger Strasse 100, 89077 Ulm, Germany.

Received: 18 May 2020 Accepted: 6 April 2021

Published online: 28 May 2021

References

1. F. Strasser, H. Puder, Adaptive feedback cancellation for realistic hearing aid applications. *IEEE/ACM Trans. Audio Speech Lang. Process.* **23**(12), 2322–2333 (2015). <https://doi.org/10.1109/TASLP.2015.2479038>
2. A. Spriet, S. Doclo, M. Moonen, J. Wouters, *Feedback Control in Hearing Aids*. (J. Benesty, M. M. Sondhi, Y. A. Huang, eds.) (Springer, Berlin, Heidelberg, 2008), pp. 979–1000. https://doi.org/10.1007/978-3-540-49127-9_48
3. B. C. Bispo, D. Freitas, in *E-Business and Telecommunications. ICETE 2014. Communications in Computer and Information Science*, ed. by M. Obaidat, A. Holzinger, and J. Filipe. Performance evaluation of acoustic feedback cancellation methods in single-microphone and multiple-loudspeakers public address systems, vol. 554 (Springer, Cham, 2015). https://doi.org/10.1007/978-3-319-25915-4_25
4. G. Rombouts, T. van Waterschoot, K. Struyve, M. Moonen, Acoustic feedback cancellation for long acoustic paths using a nonstationary source model. *IEEE Trans. Signal Process.* **54**(9), 3426–3434 (2006). <https://doi.org/10.1109/TSP.2006.879251>
5. G. Schmidt, T. Haulick, in *Topics in Acoustic Echo and Noise Control*, ed. by E. Hänsler, G. Schmidt. Signal processing for in-car communication systems (Springer, Berlin, 2006), pp. 437–493. Chap. 14
6. C. Lüke, G. Schmidt, A. Theiß, J. Withopf, *In-Car Communication*. (G. Schmidt, H. Abut, K. Takeda, J. H. L. Hansen, eds.) (Springer, New York, 2014), pp. 97–118. https://doi.org/10.1007/978-1-4614-9120-0_7
7. S. Haykin, *Adaptive Filter Theory*, 5edn. (Prentice-Hall, Inc., London, 2013)
8. M. Gimm, P. Bulling, G. Schmidt, in *Konferenz Elektronische Sprachsignalverarbeitung (ESSV)*. Energy decay based postfilter for ICC systems with feedback compensation, (Ulm, 2018)
9. A. Wolf, B. Iser, in *5th Biennial Workshop on DSP for In-Vehicle Systems*. Energy decay d feedback suppression: Theory and application, (Kiel, 2011)
10. T. V. Waterschoot, M. Moonen, in *Proceedings of the IEEE*. Fifty years of acoustic feedback control: State of the art and future challenges, vol. 99, (2011), pp. 288–327
11. E. Lleida, E. Masgrau, A. Ortega, in *7th European Conference on Speech Communication and Technology (EUROSPEECH)*. Acoustic echo control and noise reduction for cabin car communication, (Aalborg, 2001), pp. 1585–1588
12. A. Ortega, E. Lleida, E. Masgrau, F. Gallego, in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. Cabin car communication system to improve communications inside a car, vol. 2, (Orlando, 2002). <https://doi.org/10.1109/ICASSP.2002.5745493>
13. A. Ortega, E. Lleida, E. Masgrau, Speech reinforcement system for car cabin communications. *IEEE Trans. Speech Audio Process.* **13**(5), 917–929 (2005). <https://doi.org/10.1109/TSA.2005.853006>

14. J. Hellgren, F. Urban, Bias of feedback cancellation algorithms in hearing aids based on direct closed loop identification. *IEEE Trans. Speech Audio Process.* **9**(8), 906–913 (2001). <https://doi.org/10.1109/89.966094>
15. J. Withopf, G. Schmidt, in *14th International Workshop on Acoustic Signal Enhancement (IWAENC)*. Estimation of time-variant acoustic feedback paths in in-car communication systems, (Antibes, 2014). <https://doi.org/10.1109/IWAENC.2014.6953347>
16. J. Withopf, S. Rhode, G. Schmidt, in *11th ITG Conference on Speech Communication*. Application of frequency shifting in in-car communication systems, (Erlangen, 2014)
17. M. Guo, S. H. Jensen, J. Jensen, S. L. Grant, in *20th European Signal Processing Conference (EUSIPCO)*. On the use of a phase modulation method for decorrelation in acoustic feedback cancellation, (Bukarest, 2012), pp. 2000–2004
18. G. Rombouts, T. V. Watershoot, M. Moonen, Robust and efficient implementation of the PEM-AFROW algorithm for acoustic feedback cancellation. *J. Audio Eng. Soc.* **55**(11), 955–966 (2007)
19. F. Strasser, H. Puder, Correlation detection for adaptive feedback cancellation in hearing aids. *IEEE Signal Process. Letters.* **23**(7), 979–983 (2016). <https://doi.org/10.1109/LSP.2016.2575447>
20. P. Bulling, K. Linhard, A. Wolf, G. Schmidt, in *12th ITG Conference on Speech Communication*. Acoustic feedback compensation with reverb-based stepsize control for in-car communication systems, (Paderborn, 2016)
21. P. Bulling, K. Linhard, A. Wolf, G. Schmidt, in *Conference of the International Speech Communication Association (INTERSPEECH)*. Stepsize control for acoustic feedback cancellation based on the detection of reverberant signal periods and the estimated system distance, (Stockholm, 2017)
22. C. Beaugeant, V. Turbin, P. Scalart, A. Gilloire, New optimal filtering approaches for hands-free telecommunication terminals. *Signal Process.* **64**(1), 33–47 (1998). [https://doi.org/10.1016/S0165-1684\(97\)00174-6](https://doi.org/10.1016/S0165-1684(97)00174-6)
23. W. L. B. Jeannes, P. Scalart, G. Faucon, C. Beaugeant, Combined noise and echo reduction in hands-free systems: a survey. *IEEE Trans. Speech Audio Process.* **9**(8), 808–820 (2001). <https://doi.org/10.1109/89.966084>
24. G. Enzner, R. Martin, P. Vary, in *Proceedings of International Workshop on Acoustic Echo and Noise Control (IWAENC)*. On spectral estimation of residual echo in hands-free telephony, (Darmstadt, 2001)
25. V. Turbin, A. Gilloire, P. Scalart, in *1997 IEEE International Conference on Acoustics, Speech, and Signal Processing*. Comparison of three post-filtering algorithms for residual acoustic echo reduction, vol. 1, (1997), pp. 307–310. <https://doi.org/10.1109/ICASSP.1997.599633>
26. E. Hänsler, G. Schmidt, *Acoustic Echo and Noise Control - A Practical Approach*. (John Wiley & Sons, Inc., Hoboken, 2004)
27. E. A. P. Habets, S. Gannot, I. Cohen, Late reverberant spectral variance estimation based on a statistical model. *IEEE Signal Process. Letters.* **16**(9), 770–773 (2009). <https://doi.org/10.1109/LSP.2009.2024791>
28. K. Lebart, J. M. Boucher, P. Denbigh, A new method based on spectral subtraction for speech dereverberation. *Acta Acustica United Acustica.* **87**, 359–366 (2001)
29. A. Favrot, C. Faller, F. Kuech, in *IWAENC 2012; International Workshop on Acoustic Signal Enhancement*. Modeling late reverberation in acoustic echo suppression, (2012), pp. 1–4
30. M. L. Valero, E. Mabande, E. A. P. Habets, in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Signal-based late residual echo spectral variance estimation, (2014), pp. 5914–5918. <https://doi.org/10.1109/ICASSP.2014.6854738>
31. N. K. Desiraju, S. Doclo, M. Buck, T. Wolff, Online estimation of reverberation parameters for late residual echo suppression. *IEEE/ACM Trans. Audio Speech Lang. Process.* **28**, 77–91 (2020). <https://doi.org/10.1109/TASLP.2019.2948765>
32. A. Mader, H. Puder, G. Schmidt, Step-size control for acoustic echo cancellation filters - an overview. *Signal Process.* **80**(9), 1697–1719 (2000). [https://doi.org/10.1016/S0165-1684\(00\)00082-7](https://doi.org/10.1016/S0165-1684(00)00082-7)
33. J. Withopf, Signalverarbeitungsverfahren zur Verbesserung der Sprachkommunikation im Fahrzeug. Dissertation. Christian-Albrechts-Universität zu Kiel (2017)
34. C. H. Taal, R. C. Hendriks, R. Heusdens, J. Jensen, An algorithm for intelligibility prediction of time–frequency weighted noisy speech. *IEEE Trans. Audio Speech Lang. Process.* **19**(7), 2125–2136 (2011). <https://doi.org/10.1109/TASL.2011.2114881>
35. Methods for subjective determination of transmission quality. International Telecommunication Union. **1996** (1996). <https://www.itu.int/rec/T-REC-P.800-199608-I>
36. Test signals for use in telephony and other speech-based applications. International Telecommunication Union. **2018** (2018). <https://www.itu.int/rec/T-REC-P.501-201806-SlAmd1/en>
37. G. Schmidt, Residual Feedback Suppression with Extended Model-based Postfilters (2021). <https://www.dss.tf.uni-kiel.de/index.php/research/publications/publications-add-material/residual-feedback-suppression>
38. M. Gimm, A. Namenas, G. Schmidt, *11 Combination of Hands-Free and ICC Systems*. (De Gruyter, Berlin, 2020), pp. 165–182. <https://doi.org/10.1515/9783110669787-011>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)