Approximation of the Optimum Stepsize for Acoustic Feedback Cancellation Based on the Detection of Reverberant Signal Periods

Philipp Bulling[1], Klaus Linhard[1], Arthur Wolf[1], Gerhard Schmidt[2]

[1] *Daimler AG, D-89081 Ulm, Germany, Email: {philipp.bulling, klaus.linhard, arthur.wolf}@daimler.com*

[2] *Christian-Albrechts-Universität zu Kiel, D-24143 Kiel, Germany, Email: gus@tf.uni-kiel.de*

Abstract

A major challenge in acoustic feedback cancellation is the strong correlation between the excitation signal and the error signal, caused by the closed electro-acoustic loop. Due to this correlation, the convergence rate of adaptive algorithms, such as the NLMS, is limited. It was shown in a recent publication that the convergence can be improved to a large extent by using a reverb-based stepsize control. This stepsize control aims at increasing the stepsize during reverberant signal periods and decreasing the stepsize during local speech activity. Caused by reverberation of the acoustic system, there is still energy in the system immediately after local speech periods. This reverberation can be exploited to adapt the filter, since the signals are not correlated here. In this paper, the reverb-based stepsize control is further improved. Therefore, the gain in the forward path of the closed-loop system is controlled with the system distance. It is shown that in this case, the reverb-based stepsize can be interpreted as an approximation of the theoretically optimal stepsize. The proposed method uses a frequency domain NLMS algorithm for feedback cancellation. The target application is an in-car communication system.

Introduction

An in-car communication (ICC) system can improve the communication between the passengers inside a car. Therefore, the voice of the front-seat passengers is captured by means of microphones and played back via loudspeakers close to rear-seat passengers. This is especially useful in the presence of high background noise levels, e. g. while driving at high velocities.

Acoustic feedback occurs, if an audio signal, captured by a microphone, is played back via a loudspeaker close to the microphone. If there is no or little damping between loudspeaker and microphone, the loudspeaker signal is fed back to the microphone. This means that the system operates in a closed electro-acoustic loop. The feedback can cause a howling sound, if the gain of the microphone signal exceeds a certain limit. The closed electro-acoustic loop of an in-car communication system is shown in Fig. 1. In the figure, the transfer function of the forward path is $H_{\text{icc}}(f)$, where f denotes the continuous frequency. This transfer function also includes the system gain. The acoustic coupling between loudspeaker and microphone is $H(f)$. The transfer function of the resulting closed-loop system is

$$H_{\text{res}}(f) = \frac{H_{\text{icc}}(f)}{1 - H_{\text{icc}}(f) \cdot H(f)}. \tag{1}$$

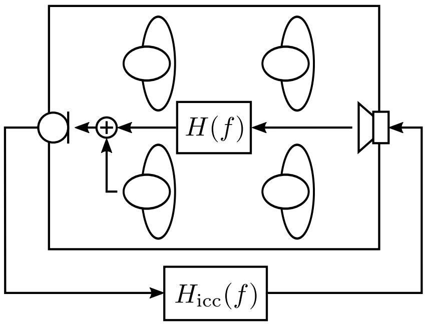

Figure 1: Block diagram of an ICC system, operating in a closed electro-acoustic loop.

The system is stable, if the so-called open loop gain is smaller than unity

$$|H_{\text{icc}}(f) \cdot H(f)| < 1. \tag{2}$$

If high gains are required, the magnitude of $H_{\text{icc}}(f)$ is large. Thus the maximum stable gain (MSG) is given by Eq. (2). To meet condition (2) even at high gains, methods to control the feedback are required. One approach will be presented in this work.

Related Work

Acoustic feedback control is of interest in various research areas. Besides ICC systems, acoustic feedback occurs for example in hearing aids or public address systems. A large collection of general approaches, addressing the problem of acoustic feedback, can be found in [1]. Existing approaches can be divided into three groups. *Spatial filters* aim at reducing the feedback by means of microphone or loudspeaker arrays. *Feedback suppression* methods try to suppress howling by reducing the gain in critical frequency bands. This can be done for example in the frequency domain with spectral subtraction [2] or in the time domain with adaptive notch filters [3]. The third group is *adaptive feedback cancellation.* Here, the acoustic coupling between loudspeaker and microphone is estimated by means of an adaptive filter. The principle is similar to adaptive echo cancellation, used for example in hands-free systems. However, in feedback cancellation, the challenge is a strong correlation between the loudspeaker signal and the local speech, caused by the electro-acoustic loop. As a result, the convergence rate of the filter is slow. By applying decorrelation methods such as frequency shift or prewhitening with linear prediction, convergence can be improved [4, 5, 6]. In a recent publication, an adaptive feedback canceler with a

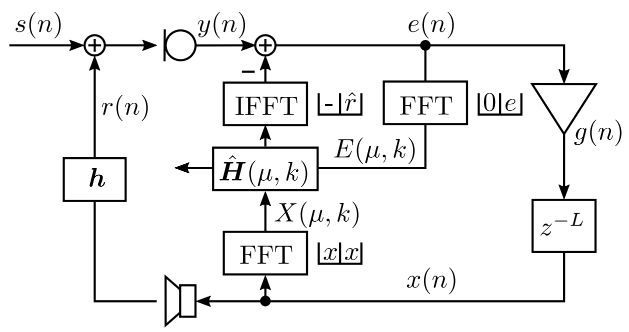

Figure 2: Block diagram of the adaptive feedback canceler.

stepsize controlled filter update was presented. The approach is capable of improving convergence without further decorrelation methods [7]. The stepsize is controlled reverb-based. In this paper, the stepsize control will be further improved.

Acoustic Feedback Cancellation

The acoustic coupling between loudspeaker and microphone is described by the impulse response $\boldsymbol{h}(n)$, where n denotes the discrete time index. For the derivation, it is assumed here that the impulse response does not vary over time, i. e.

$$\boldsymbol{h}(n) = \boldsymbol{h}(n+1) = \boldsymbol{h}. \tag{3}$$

In order to cancel the feedback, the impulse response $\boldsymbol{h}$ is estimated by means of a normalized least mean squares (NLMS) algorithm. To avoid time-consuming convolutions, the NLMS is calculated in the frequency domain. Therefore an overlap-save filterbank with FFT-length N and frameshift $L = N/2$ is used, as for example described in [8].
The structure of the algorithm is shown in Fig. 2. The microphone signal $y(n)$ consists of local speech $s(n)$ and feedback $r(n)$. The error signal $e(n)$ is obtained by subtracting the estimated feedback from the microphone signal. In the forward path, $e(n)$ is amplified by gain $g(n)$ and delayed by L samples, resulting in the loudspeaker signal

$$x(n) = g(n) \cdot e(n-L). \tag{4}$$

Gain $g(n)$ is the desired system gain. The delay is caused by the block processing. The FFT/IFFT blocks symbolize the filterbanks. The spectra of loudspeaker- and error signal at block $k = n/L$ are obtained by

$$\boldsymbol{X}(k) = \text{FFT}\left\{[x(n-N+1), \ldots, x(n-1), x(n)]^T\right\} \tag{5}$$

and

$$\boldsymbol{E}(k) = \text{FFT}\left\{[\boldsymbol{0}_L, e(n-L+1), \ldots, e(n-1), e(n)]^T\right\}. \tag{6}$$

$\boldsymbol{0}_L$ is a null vector of length L. Zero-padding is necessary to avoid errors caused by circular convolution. $\boldsymbol{X}(k)$ is a vector containing the frequency samples

$$\boldsymbol{X}(k) = [X(\mu_0, k), X(\mu_1, k), \ldots, X(\mu_{N-1}, k)]^T, \tag{7}$$

where $\mu = \mu_0, \ldots \mu_{N-1}$ denote the discrete frequency bins. The same applies to $\boldsymbol{E}(k)$. $\hat{\boldsymbol{H}}(\mu, k)$ is the estimated subband impulse response of $\boldsymbol{h}$. It contains M filter taps at time instant k and frequency bin μ

$$\hat{\boldsymbol{H}}(\mu, k) = [H_0(\mu, k), H_1(\mu, k), \ldots, H_{M-1}(\mu, k)]^T. \tag{8}$$

The M previous taps of the subband loudspeaker signal are summarized in vector $\boldsymbol{X}(\mu, k)$

$$\boldsymbol{X}(\mu, k) = [X(\mu, k), X(\mu, k-1), \ldots, X(\mu, k-M+1)]^T. \tag{9}$$

The filter update of the adaptive filter can then be written as

$$\hat{\boldsymbol{H}}(\mu, k+1) = \hat{\boldsymbol{H}}(\mu, k) + \alpha(\mu, k) \cdot \frac{E(\mu, k) \cdot \boldsymbol{X}^*(\mu, k)}{\|\boldsymbol{X}(\mu, k)\|^2}, \tag{10}$$

where $\|\cdot\|$ denotes the Euclidean norm of a vector and $(\cdot)^*$ means complex conjugate. The gradient of the filter update must again be constrained with zeros. Therefore it is transformed to the time domain. After setting the right half to zero, it is transformed back to the frequency domain. The stepsize of the filter update is $\alpha(\mu, k)$. The control mechanisms explained in the following sections are applied directly to $\alpha(\mu, k)$.

Pseudooptimal Stepsize

The convergence characteristic of an adaptive filter can be described by the system distance

$$\|\boldsymbol{H}_\Delta(\mu, k)\|^2 = \|\boldsymbol{H}(\mu) - \hat{\boldsymbol{H}}(\mu, k)\|^2, \tag{11}$$

where $\boldsymbol{H}(\mu) \bullet\!\!-\!\!\circ\, \boldsymbol{h}$ is the real subband impulse response. Since the system distance becomes smaller, if the estimation $\hat{\boldsymbol{H}}(\mu, k)$ converges towards $\boldsymbol{H}(\mu)$, the optimization criterion for the optimal stepsize $\alpha_{\text{opt}}(\mu, k)$ is minimizing Eq. (11).
For convenience, the derivation of the optimal stepsize is not explained here. It was found in [9, 10] that the optimal convergence is achieved, if

$$\alpha_{\text{opt}}(\mu, k) = \frac{\text{E}\left\{|E_{\text{u}}(\mu, k)|^2\right\}}{\text{E}\left\{|E(\mu, k)|^2\right\}}. \tag{12}$$

$\text{E}\{\cdot\}$ denotes the expected value. $E_{\text{u}}(\mu, k)$ is the undisturbed error signal. Undisturbed means in absence of local speech, i. e. $s(n) = 0$. In this case, $E_{\text{u}}(\mu, k)$ can be calculated as follows

$$\begin{aligned} E_{\text{u}}(\mu, k) &= \boldsymbol{X}(\mu, k)^T \boldsymbol{H}(\mu) - \boldsymbol{X}(\mu, k)^T \hat{\boldsymbol{H}}(\mu, k) \\ &= \boldsymbol{X}(\mu, k)^T \cdot \boldsymbol{H}_\Delta(\mu, k). \end{aligned} \tag{13}$$

Inserting (13) into (12) leads to

$$\alpha_{\text{opt}}(\mu, k) = \frac{\text{E}\left\{\left|\boldsymbol{X}(\mu, k)^T \cdot \boldsymbol{H}_\Delta(\mu, k)\right|^2\right\}}{\text{E}\left\{|E(\mu, k)|^2\right\}}. \tag{14}$$

If the assumption is made that loudspeaker signal and system distance are not correlated, Eq. (14) can be written as

$$\alpha_{\text{opt}}(\mu, k) \approx \frac{\text{E}\left\{|X(\mu, k)|^2\right\}}{\text{E}\left\{|E(\mu, k)|^2\right\}} \cdot \text{E}\left\{\|\boldsymbol{H}_\Delta(\mu, k)\|^2\right\}. \tag{15}$$

It must be noted that in case of the closed-loop system shown in Fig. 2, $e(n)$ and $x(n)$ are decorrelated only by the delay z^{-L}. Since the delay time is in an order where speech is stationary, this decorrelation is not sufficient. Thus, the assumption is violated. However, it will be shown in the following sections that with an appropriate gain control it is possible to control the stepsize with Eq. (15) also in a closed-loop system.

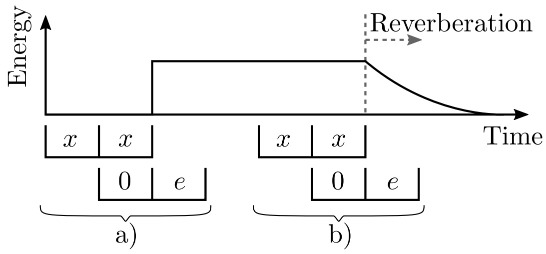

Figure 3: Reverb detection based on delayed blocks. a) Beginning of speech activity. b) Reverberation.

Reverb-based Stepsize

The stepsize control described in [7] is based on the detection of reverberant signal periods. The reason for reverberation is that the closed-loop system operates in an acoustic environment. As soon as local speech stops, the reverberation energy decays exponentially. These short periods of time are similar to periods of remote single talk, known from adaptive echo cancellation. Since there is only excitation of the loudspeaker signal but no local speech, signals are not correlated here. As a consequence, the stepsize of the adaptive filter can be large. To detect reverberation and increase the stepsize accordingly, two facts of the overlap-save filterbank are exploited:

- The first half of error signal is set to zero, before it is transformed to the frequency domain.
- The loudspeaker is delayed by a half block (or L samples) compared to the error signal.

The principle is shown schematically in Fig. 3. The blocks below the graph symbolize the time domain samples that are transformed to the frequency domain according to Eq. (5) and (6) at two different time instants a) and b). At the beginning of speech activity, the energy in the forward path rises abruptly. When speech activity stops, it decays exponentially. Due to the zeros and the delay, the expected power spectral density of the error block follows signal changes faster and earlier than the loudspeaker signal. This can be used as reverb detection, if the relation

$$\alpha_{\mathrm{rev}}(\mu,k) = \frac{\mathrm{E}\left\{|X(\mu,k)|^2\right\}}{g(k)^2 \cdot \mathrm{E}\left\{|E(\mu,k)|^2\right\}} \tag{16}$$

is regarded. Since the error signal is multiplied by $g(n)$ (Eq. (4)), the power spectral density must be multiplied by $g(k)^2$ to ensure comparability with the loudspeaker signal. At the beginning of speech activity the reverb detection becomes

$$\alpha_{\mathrm{rev}}(\mu,k_{\mathrm{a})}) = \frac{\mathrm{E}\left\{|X(\mu,k_{\mathrm{a})})|^2\right\}}{g(k_{\mathrm{a})})^2 \cdot \mathrm{E}\left\{|E(\mu,k_{\mathrm{a})})|^2\right\}} < 1 \tag{17}$$

and during reverberation

$$\alpha_{\mathrm{rev}}(\mu,k_{\mathrm{b})}) = \frac{\mathrm{E}\left\{|X(\mu,k_{\mathrm{b})})|^2\right\}}{g(k_{\mathrm{b})})^2 \cdot \mathrm{E}\left\{|E(\mu,k_{\mathrm{b})})|^2\right\}} > 1. \tag{18}$$

If Eq. (16) is used to control the stepsize, the filter adaption is fast during reverberation. At the beginning of speech activity, the adaption process is slowed down. During periods of constant speech or in absence of speech, $\alpha_{\mathrm{rev}}(\mu,k)$ is approximately 1. To avoid that the stepsize becomes too large or negative, $\alpha_{\mathrm{rev}}(\mu,k)$ is limited to minimum and maximum values $\alpha_{\min}$ and $\alpha_{\max}$.

Gain Control

By comparing Eq. (15) and Eq. (16), it follows that the optimal stepsize and the reverb-based stepsize are identical if

$$\frac{1}{g(k)^2} = \mathrm{E}\left\{\|\boldsymbol{H}_\Delta(\mu,k)\|^2\right\}. \tag{19}$$

This means that the reverb-based stepsize becomes optimal, if the gain in the forward path of the system is controlled by the inverse system distance. Since with the reverb-based stepsize control adaption mainly happens during reverberation, meaning when signals are not correlated, the assumption made in Eq. (15) is valid.
The gain $g(n)$ is applied in the time domain while $\|\boldsymbol{H}_\Delta(\mu,k)\|^2$ is frequency selective. After Parseval's theorem, its equivalent whether the system distance is calculated in the time or in the frequency domain. Thus, the gain can be calculated as follows

$$g(n) = \frac{1}{\sqrt{\mathrm{E}\left\{\|\boldsymbol{h}_\Delta(n)\|^2\right\}}}, \tag{20}$$

where

$$\|\boldsymbol{h}_\Delta(n)\|^2 = \|\boldsymbol{h} - \hat{\boldsymbol{h}}(n)\|^2. \tag{21}$$

$\hat{\boldsymbol{h}}(n)$ is the estimated impulse response, reconstructed from the estimated subband impulse responses $\hat{\boldsymbol{H}}(\mu,k)$.

Estimation of the Unknown Parameters

To process the adaptive filter, the system distance and the expected values must be known. The expected values can be approximated by first order IIR-smoothing of the squared magnitude [9]. Estimating the system distance is a more challenging task. Different approaches already exist, most of them used in the context of adaptive echo cancellation. However, it is beyond the scope of this work to derive an estimator for the adaptive feedback canceler. Since during simulations the real system distance is known, in the next section it is assumed that a perfect estimator is available.

Simulation Results

A male speech signal, normalized to unity, is used for simulations. The sampling frequency is $f_s = 32\,\mathrm{kHz}$. To simulate the feedback, the loudspeaker signal is convolved with the impulse response from rear seat loudspeaker to driver microphone, measured in a real sedan car. The reverberation time T_{60} is approx. 55 ms. To cover the whole length of the impulse response $\boldsymbol{h}$, the adaptive filter $\hat{\boldsymbol{H}}(\mu,k)$ has $M = 8$ coefficients. The FFT-length is $N = 512$ samples, thus the delay in the forward path is $L = 256$ samples or 8 ms. To obtain the stepsize of the adaptive filter update (Eq. (10)), $\alpha_{\mathrm{rev}}(\mu,k)$ is limited to $\alpha_{\min} = 0$ and $\alpha_{\max} = 1.2$. Without feedback cancellation, the MSG is reached at 0 dB gain. The gain is controlled with Eq. (20). With decreasing system distance the gain would increase steadily. Since this is not useful in a real application, it is limited to a maximum value $g_{\max} = 50\,\mathrm{dB}$. However, this value is arbitrary, meaning that the system remains stable also at higher gains. Fig. 4 shows gain and system distance. At

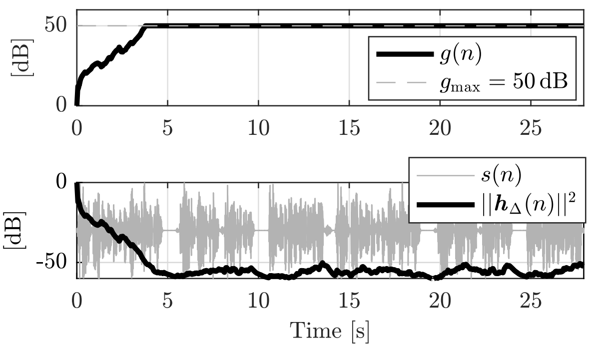

Figure 4: Gain and system distance.

the beginning, the gain rises inversely to the system distance. As soon as the desired gain is reached, it cannot rise further. From that moment on, the system distance fluctuates at around 5 dB below $-g_{\max}$. One can observe that the system distance always drops, immediately after local speech stops (e. g. 5 s, 10 s, 14 s, 20 s and 24 s). These are time instants with pure reverberation. One of these speech pauses is shown in Fig. 5 in detail. For bet-

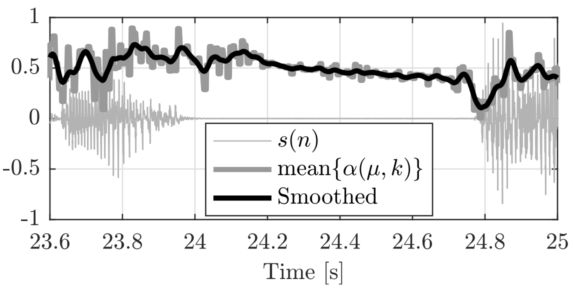

Figure 5: Mean stepsize during speech pause.

ter presentation, the stepsize is averaged over the speech spectrum (from 125 Hz to 8000 Hz) and smoothed. At the end of speech activity, the stepsize is large. As a consequence, the system distance can be reduced, since signals are not correlated during reverberation. Once the speaker starts talking again (24.8 s), the stepsize is small, preventing divergence of the filter.

The possible MSGs are summarized in Tab. 1. A feedback canceler without stepsize control is stable up to 7 dB gain. This shows that the proposed stepsize control improves stability to a large extend. Similar results are obtained for different speakers and also in presence of background noise.

Table 1: Comparison of the possible MSGs.

Without feedback cancellation	0 dB
Fixed stepsize $\alpha(\mu, k) = 0.02$	$\approx 7\,\mathrm{dB}$
Reverb-based stepsize w/o gain control [7]	$\approx 20\,\mathrm{dB}$
Reverb-based stepsize with gain control	$\gg 50\,\mathrm{dB}$

Conclusion and Outlook

A stepsize control for an adaptive feedback canceler was presented. The stepsize control exploits reverberant signal periods to improve convergence. It was shown that the reverb-based stepsize can be interpreted as optimal stepsize, if the gain of the system is controlled with the inverse system distance. The advantage of the stepsize control is its stability even at high gains. The drawback of the method is that the system distance must be known. However, it will be shown in a subsequent publication that a suitable method to estimate the system distance can be developed.

References

[1] T. v. Watershoot and M. Moonen, "Fifty years of acoustic feedback control: State of the art and future challenges," *Proceedings of the IEEE*, vol. 99, pp. 288–327, Feb. 2011.

[2] A. Wolf and B. Iser, "Energy decay model based feedback suppression: Theory and application," in *5th Biennial Workshop on DSP for In-Vehicle Systems*, (Kiel), 2011.

[3] G. Schmidt and T. Haulick, "Signal processing for in-car communication systems," in *Topics in Acoustic Echo and Noise Control* (E. Hänsler and G. Schmidt, eds.), ch. 14, pp. 437–493, Berlin: Springer, 2006.

[4] J. Withopf 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), Sept. 2014.

[5] J. Withopf, S. Rhode, and G. Schmidt, "Application of frequency shifting in in-car communication systems," in *11th ITG Conference on Speech Communication*, (Erlangen), Sept. 2014.

[6] G. Rombouts, T. v. Watershoot, and M. Moonen, "Robust and efficient implementation of the pem-afrow algorithm for acoustic feedback cancellation," *Journal of the Audio Engineering Society*, vol. 55, pp. 955–966, Nov. 2007.

[7] P. Bulling, K. Linhard, A. Wolf, and G. Schmidt, "Acoustic feedback compensation with reverb-based stepsize control for in-car communication systems," in *12th ITG Conference on Speech Communication*, (Paderborn), Oct. 2016.

[8] J.-S. Soo and K. K. Pang, "Multidelay block frequency domain adaptive filter," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, pp. 373–376, Feb. 1990.

[9] E. Hänsler and G. Schmidt, *Acoustic Echo and Noise Control - A Practical Approach.* John Wiley & Sons, Inc., 2004.

[10] S. Haykin, *Adaptive Filter Theory.* Prentice-Hall, Inc., 4th ed., 2002.