Frequency Domain Step-Size Control in Non-Stationary Environments

Julia Vogel
Swiss Federal Institute of Technology
Institute of Scientific Computing
8092 Zurich, Switzerland
vogel@inf.ethz.ch

Martin Heckmann Kristian Kroschel
University of Karlsruhe
Institute of Communications
76128 Karlsruhe, Germany
{heckmann,kroschel}@int.uni-karlsruhe.de

Abstract

The presented algorithm detects changes of the loudspeaker enclosure microphone (LEM) system in an acoustic echo cancellation setup under non-stationary noise conditions. Changes of the LEM system entail a mismatch of the cancellation filter, consequently making the beforehand cancelled echos audible again. The new algorithm avoids the reduction of the adaptation speed after these changes. System changes are detected by monitoring the different behavior of two coherences: These changes have only a small impact on the coherence between the microphone and the loudspeaker signal, whereas they significantly influence the coherence between the microphone signal and the output of the cancellation filter.

1. Introduction

In many countries, the use of a mobile phone while driving a car is already prohibited by law. Therefore, hands-free communication will soon be a feature of modern telephone systems that is taken for granted by the customer. Though, the limitations of hands-free telephone systems in a car environment are very challenging due to the non-stationarity of the environment, the high noise level and, because of the long distance between loudspeaker and microphone, the appearance of acoustic echos. For that reason, the combination of noise and echo cancellation systems is still the topic of much active research [1][2].

The most popular concept for echo cancellation is based on adaptive filtering. Here, the loudspeaker enclosure microphone (LEM) system is estimated via the correlation between the loudspeaker signal and the microphone signal. Due to the fact that noise cancellation systems usually work in the frequency domain, a frequency domain echo cancellation algorithm such as the one based on the fast least-mean-square (FLMS) algorithm is preferable for combined echo and noise cancellation (Fig. 1). Unfortunately, additive noise such as environmental noise and double-talk severely affect the convergence of the adaptive filter. Double-talk is defined as the situation where both the remote and the local speaker are active. Therefore, the step-size of the adaptation algorithm has to be adjusted depending on the noise level. Several step-size control algorithms have been proposed in the past [3]- [7].

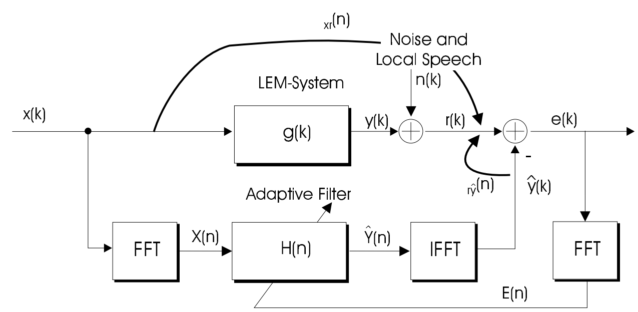

Figure 1. Basic frequency domain echo cancellation system

Changes of the LEM system are particularly difficult to handle for most of these algorithms. This results from the fact that both the change of the LEM system and a sudden increase of the noise level lead to an increase of the error signal. The increasing error signal then entails a reduction of the step-size. Hence, instead of quickly adapting to the changed LEM system and thereby compensating the echo again, the adaptation slows down and the echos reappear. To circumvent this behavior, the detection of system changes is indispensable.

We propose a method for controlling the step-size of an FLMS echo canceller that incorporates the detection of LEM system changes. The step-size is set according to the coherence between the microphone signal and the output of the compensation filter and the coherence between the microphone and the loudspeaker signal.

The next section introduces briefly the basic concept of frequency-selective step-size control based on the coherence function. In Section 3, the concept is extended by the special treatment of the beginning of the adaptation and the detection of LEM system changes. Section 4 presents the simulations and discusses the results. A conclusion is given in Section 5.

Definitions The power ratio κ between the echo signal $y(k)$ and the local speech signal is defined similar to the signal-to-noise ratio (SNR):

$$
\begin{aligned}
SNR &= 10 \log \frac{\sigma_y^2}{\sigma_{noise}^2} \qquad (1) \\
\kappa &= 10 \log \frac{\sigma_y^2}{\sigma_{local}^2}
\end{aligned}
$$

The definition of the system distance $D(k)$ between the LEM system $\mathbf{g}(k)$ and the adaptive filter $\mathbf{h}(k)$ is:

$$
D(k) = \frac{|\mathbf{g}(k) - \mathbf{h}(k)|^2}{|\mathbf{g}(k)|^2} \qquad (2)
$$

2. Frequency-selective step-size control

In echo cancellation systems, the coefficients of the adaptive filter $h(k)$ are adjusted according to the cross-correlation between the input signal $x(k)$ and the error signal $e(k)$ [8]. Additive noise that is not correlated with the input signal does not disturb the adaptation. Due to the limited length of the observation interval in real-time applications, this assumption does generally not hold and a correlation of the additive noise with the input signal is observed. Hence, additive noise can lead to misadjustment and even divergence of the adaptation algorithm. As a result, the step-size $\alpha(k)$ needs to be controlled depending on the power of the additive local signals (compare Fig. 1).

As the noise signal $n(k)$ is not available separately, we evaluate the coherence $\gamma_{xr}^2(n)$ between the microphone signal $r(k)$

$$
r(k) = g(k) * x(k) + n(k) = y(k) + n(k) \qquad (3)
$$

and the loudspeaker signal $x(k)$:

$$
\begin{aligned}
\gamma_{xr}^2(n) &= \frac{|S_{xr}(n)|^2}{S_{xx}(n) S_{rr}(n)} \qquad (4) \\
0 &\leq \gamma_{xy}^2(n) \leq 1
\end{aligned}
$$

to estimate the noise level in each frequency bin [7]. "$*$" denotes the convolution. Previously, the coherence function has been used to detect double-talk situations [9].

The coherence between two signals $x(k)$ and $r(k)$ is defined as the normalized cross-spectral density between $x(k)$ and $r(k)$ [10]. Therefore, it depends on both the cross-correlation between the loudspeaker signal and the noise and the noise power. Thus, the coherence function can be used to control the step-size of the adaptation algorithm depending on the level of the additive noise [7]. For low coherence values, the power of the noise signal $n(k)$ is high and the step-size $\alpha(k)$ needs to be set to a low value. The same is true vice versa.

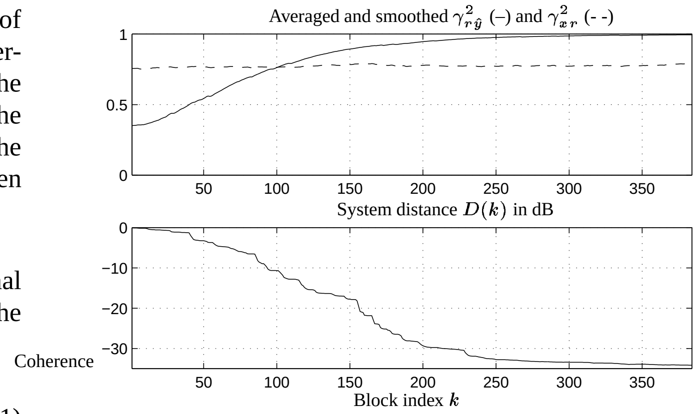

Figure 2. Averaged and smoothed coherences in comparison to the improving system distance (infinite SNR)

Due to the short observation intervals and the comparatively long impulse response of the LEM system, the coherence between the loudspeaker signal $x(k)$ and the microphone signal $r(k)$, although fully correlated in a noise-free environment, shows a significant influence of the LEM system and does not reach the theoretical limit of $\gamma_{xr}^2 = 1$.

In order to reduce the effect of the LEM system, in our case, the coherence is evaluated between the microphone signal $r(k)$ and the output $\hat{y}(k)$ of the compensation filter $h(k)$:

$$
\hat{y}(k) = h(k) * x(k)\,. \qquad (5)
$$

This implies the assumption that $h(k)$ and $g(k)$ are similar, which is true after a few adaptation steps. Assuming a system distance of at least $-10\,db$, the signals are affected by very similar filters $g(k)$ and $h(k)$ resulting in a negligible influence of these filters on the coherence between these two signals. This behavior of the two coherences can be observed in Fig. 2. Note that the block index k corresponds to the blocks processed by the FLMS algorithm. The coherences $\gamma_{xr}^2(n)$ and $\gamma_{r\hat{y}}^2(n)$, in fact frequency-domain functions, have been averaged over all frequency bins n to obtain one coherence value per block k. γ_{xr}^2 and $\gamma_{r\hat{y}}^2$ have also been smoothed over some blocks.

By applying the coherence function, the noise level can be evaluated in independent frequency bins. Thus, the coherence function enables us to control the step-size independently for each frequency bin. As both the loudspeaker and the noise signal, especially in the case of double-talk, are not spectrally flat, frequency-selective step-size control improves not only the speed of convergence but also the quality of the processed speech: in frequency bins where adaptation fails because of a high noise level, the echo is covered by the noise. In every bin with a low noise level the adaptation can still take place.

3. Step-size control in non-stationary environments

The step-size control based on the coherence between $r(k)$ and $\hat{y}(k)$ has two major drawbacks. The first occurs at the beginning of the adaptation. Here, the coefficients of the compensation filter $h(k)$ are equal to or only slightly greater than zero. Thus, the assumption that $r(k)$ and $\hat{y}(k)$ are affected by similar filters does not hold anymore. The large difference between the two signals results in a very low coherence even in the absence of additive noise. As a consequence, the step-size $\alpha(k)$ is low and the speed of convergence is very poor which is especially counteractive at the beginning of the adaptation.

A similar problem arises when the environment is time-variant and the LEM system changes (e.g. by opening the car window). This results in a significant increase of the residual echo signal. To reduce the residual echo again to a tolerable level, the compensation filter has to adapt quickly to the new LEM system. Unfortunately, the assumption that both signals $r(k)$ and $\hat{y}(k)$ have passed equal linear systems does not hold in this situation either. The coherence will be low and will lead to a low step-size and thus slow convergence. Therefore, a system for the detection of LEM changes is required.

Previously proposed algorithms to detect changes of the LEM system are time domain algorithms [5][11]. For that reason, they are not suitable for use with the FLMS algorithm. We exploit the effects of linear systems on the coherence to detect changes of the LEM system. As stated in Section 3, for limited observation intervals, linear systems have a significant impact on the coherence. To control the adaptation, we assumed that the compensation filter is adjusted to the LEM system and hence the coherence between the microphone signal $r(k)$ and the compensation filter output signal $\hat{y}(k)$ is only affected by additive noise $n(k)$. Consequently, the coherence $\gamma^2_{r\hat{y}}(n)$ will be larger than the coherence $\gamma^2_{xr}(n)$. Deviations from this expected behavior indicate a misadjustment of the compensation filter.

Controlling the beginning of the adaptation At the beginning of the adaptation, the assumption that $g(k)$ and $h(k)$ are similar obviously cannot be made. Therefore, we estimate both the coherence $\gamma^2_{r\hat{y}}(n)$ between the microphone signal $r(k)$ and the compensation filter output $\hat{y}(k)$ and the coherence $\gamma^2_{xr}(n)$ between the input signal $x(k)$ and the microphone signal $r(k)$. As long as the compensation filter is not adapted to the LEM system, $\gamma^2_{xr}(n)$ will be larger than $\gamma^2_{r\hat{y}}(n)$. Thus, the step-size will be controlled by $\gamma^2_{xr}(n)$ until $\gamma^2_{r\hat{y}}(n)$ is as large as $\gamma^2_{xr}(n)$ showing that the adaptation has proceeded significantly and the previous assumption is now applicable. The behavior of the two coherences depending on the decreasing system distance at the beginning of the adaptation can be seen in Fig. 2. As mentioned before, the coherence values were highly smoothed in order to reduce the variance of the estimation and that only blocks with voice activity in the loudspeaker signal $x(k)$ were counted. The coherence $\gamma^2_{xr}(n)$ is independent of the system distance and remains constant whereas $\gamma^2_{r\hat{y}}(n)$ is at a very low level at the beginning of the adaptation and increases to the theoretical limit for infinite SNR $\gamma^2_{r\hat{y}}(n) = 1$.

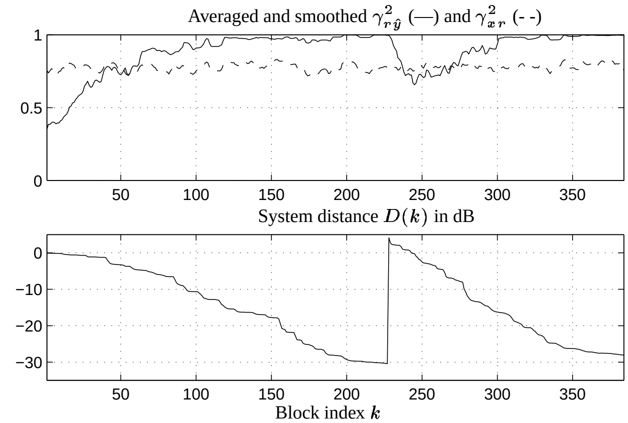

Figure 3. Coherence and system distance with a system change in block no. 230. The simulation was carried out with infinite SNR.

Detection of system changes In a similar manner, system changes can be detected. The coherence $\gamma^2_{xr}(n)$ is only slightly sensitive to changes of the LEM system. However, the coherence $\gamma^2_{r\hat{y}}(n)$ is highly influenced by system changes because they lead to a violation of the assumption that the filters $g(k)$ and $h(k)$ are identical. As a consequence, the coherence $\gamma^2_{r\hat{y}}(n)$ drops. It might even drop below the value of the coherence $\gamma^2_{xr}(n)$ depending on the extent of the system change. This decrease of the coherence $\gamma^2_{r\hat{y}}(n)$ can be exploited to detect changes of the LEM system.

Fig. 3 shows the two coherences when the LEM system changes. In block no. 230 of the simulation, the filter $g(k)$ used to simulate the LEM system was shifted by two taps. The impact of this shift on the coherence $\gamma^2_{r\hat{y}}(n)$ is significant while the coherence $\gamma^2_{xr}(n)$ remains unchanged. The corresponding plot for the system distance $D(k)$ shows that $D(k)$ rises abruptly after the change and then quickly drops back to its previous value. A change of the LEM system is detected as soon as the coherence $\gamma^2_{r\hat{y}}(n)$ drops below a certain threshold and, at the same time, the coherence $\gamma^2_{xr}(n)$ remains constant. In order to reduce the variance of the coherence estimates, a smoothing over consecutive blocks is necessary. This also results in a delay in the detection of system changes. Therefore, the degree of smoothing is a trade-off between the reduction of the variance and the inertia of the algorithm.

An important issue is the differentiation of system changes from simple increases of the noise level as for example during double-talk: In our system setup (see Fig. 1), a rise of the noise signal $n(k)$ not only affects the coherence $\gamma^2_{r\hat{y}}(n)$ between the compensation filter output signal $\hat{y}(k)$ and the microphone signal $r(k)$ but also the coherence $\gamma^2_{xr}(n)$ between the input signal $x(k)$ and the microphone signal $r(k)$. Thus, it is possible to differentiate between changes of the LEM system and increases of the noise signal by comparing the behavior of the two coherences. This is illustrated in Fig. 4. Here, in block no. 230 a local speech signal at a power ratio of $\kappa = 0\,dB$ was added. As can be seen, both coherences are affected in the same way by the additive speech signal. This is in contrast to the behavior of the coherences after a LEM system change (Fig. 3), where the coherence $\gamma^2_{xr}(n)$ remains constant.

Controlling the step-size after a system change The purpose of the detection of system changes is to avoid the reduction of the convergence speed due to a low step-size. In order to achieve this, the difference $\Delta(n)$ between the coherences $\gamma^2_{r\hat{y}}(n)$ and $\gamma^2_{xr}(n)$ is monitored in each frequency bin. As soon as a system change is detected, the mean value $\bar{\Delta}(n)$ over some blocks k before the system change is computed. This difference is added to the coherence values $\gamma^2_{xr}(n)$ which were not affected by the system change. This adjusted coherence $\gamma^2_{xr}(n) + \bar{\Delta}(n)$ controls the step-size after the system change. As a result, the adaptation speed will remain on a constant level despite of the system change. With proceeding adaptation, the coherence $\gamma^2_{r\hat{y}}(n)$ will rise to its previous value. From that point on, the step-size will again only be determined by $\gamma^2_{r\hat{y}}(n)$.

4. Simulation results

For the simulation, a recorded male voice signal was used as input signal $x(k)$. Both the LEM system $g(k)$ and the additive noise signal have been measured in a car. The noise level was set to $SNR = 10\,dB$. In block no. 230, the filter $g(k)$ was shifted by two taps to simulate a LEM system change. Fig. 5 compares the behavior of the system distance for three different step-size control strategies: A) manually set fixed step-size, B) frequency-selective step-size control without the detection of system changes, C) frequency-selective step-size control with the detection of system changes.

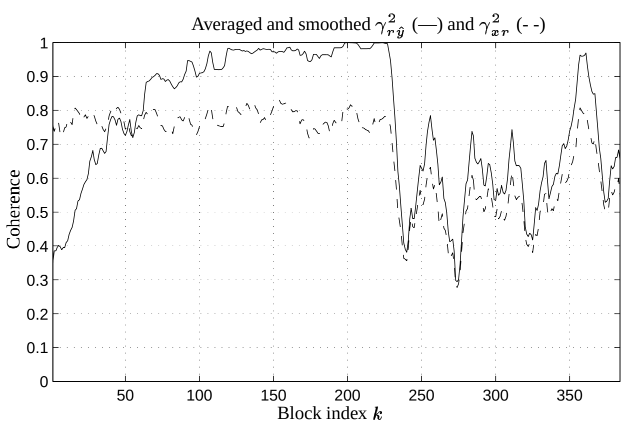

Figure 4. Coherences with sudden start of a local speaker at a power ratio $\kappa = 0\,dB$ in block no. 230. No additional local noise was present.

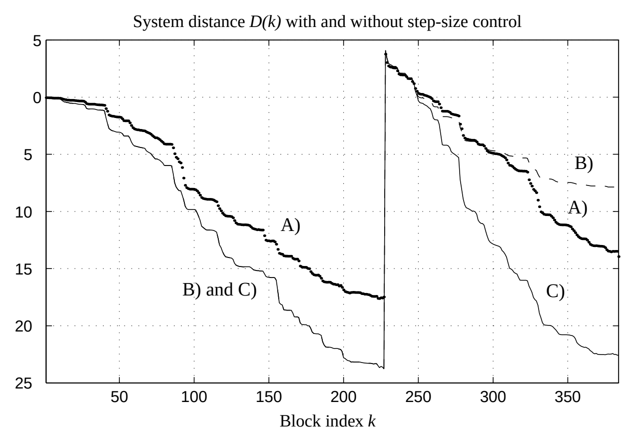

Figure 5. Comparison of the system distance in the cases: A) no step-size control, B) step-size control, but no detection of system changes, C) both step-size control and detection of system changes

When no step-size control is used, the performance is the worst, because the fixed step-size has to be set to a constant small value in order to prevent the algorithm from divergence (curve A)).

Before the system changes, there is no difference in performance with and without the use of the detection algorithm (curves B) and C)), although the performance with the frequency-selective step-size control is clearly better than without any step-size control. After the system change, the adaptation with fixed step-size (curve A)) outperforms the adaptation with frequency-selective step-size control but without the detection of system changes (curve B)). The reason is the automatic setting of a low step-size as mentioned in Section 4. This behavior is overcome in case C): When system changes are detected, the algorithm converges much faster than in cases A) and B).

5. Conclusion

We presented an algorithm that improves the performance of a frequency-domain echo cancellation system. The coherences between input/microphone signal and microphone signal/output of the cancellation filter are estimated and monitored. By this, the detection of LEM system changes in a non-stationary environment can be combined with the step-size control of the echo cancellation algorithm. Both the detection of system changes and the step-size control are performed in the frequency domain. The advantage is that the echo cancellation algorithm can easily be merged with a frequency-domain noise cancellation system. Also, the frequency-selectivity of the step-size control leads to a faster convergence of the adaptation and an improved hearing impression. In this new algorithm, a separate detection of double-talk situations is not necessary. Double-talk is dealt with by the algorithm as "normal" additive noise and thus needs not to be treated separately.

References

[1] M. Ihle and K. Kroschel, "Integration of noise reduction and echo attenuation for handset-free communication", in *Intl. Workshop on Acoustic Echo and Noise Control*, London, Great Britain, 1997, pp. 69–72.

[2] D. Gustafsson and P. Jax, "Combined residual echo and noise reduction: A novel psychoacoustically motivated algorithm", in *European Signal Processing Conference, EUSIPCO*, Rhodes, Greece, 1998, pp. 961–964.

[3] U. Schultheiß, *Über die Adaption eines Kompensators für akustische Echos*, Number 90 in Reihe 10. VDI Fortschrittberichte, Düsseldorf, 1988.

[4] Y. Haneda, S. Makino, J. Kojima, and S. Shimauchi, "Implementation and evaluation of an acoustic echo canceller using duo-filter control system", in *European Signal Processing Conference, EUSIPCO*, Trieste, Italy, 1996, pp. 1115–1118.

[5] C. Breining, *Steuerung eines Kommunikationsterminals mit Freisprecheinrichtung*, Number 570 in Reihe 10. VDI Fortschrittberichte, Düsseldorf, 1999.

[6] C. Antweiler, J. Grunwald, and H. Quack, "Approximation for optimal step-size control for acoustic echo cancellation", in *IEEE Intl. Conf. Acoustics, Speech and Signal Processing, ICASSP*, Munich, Germany, 1997, pp. 295–298.

[7] M. Heckmann, J. Vogel, and K. Kroschel, "Frequency selective step-size control for acoustic echo compensation", in *European Signal Processing Conference, EUSIPCO*, Tampere, Finland, September 2000.

[8] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, Englewood Cliffs, NJ, 3. edition, 1996.

[9] T. Gänsler, M. Hansson, C.-J. Ivarsson, and G. Salomonsson, "A double-talk detector based on coherence", *IEEE Trans. Communications*, vol. 44, no. 11, pp. 1421–1427, November 1996.

[10] J.S. Bendat and A.G. Piersol, *Measurement and Analysis of Random Data*, John Wiley & Sons, New York, 1966.

[11] R. Frenzel, *Freisprechen in gestörter Umgebung*, Number 228 in Reihe 10. VDI Fortschrittberichte, Düsseldorf, 1992.