DSP TO IMPROVE ORAL COMMUNICATIONS INSIDE VEHICLES

Alfonso Ortega, Eduardo Lleida, Enrique Masgrau.
Department of Electronics Engineering and Communications.
University of Zaragoza, Spain
ortega@posta.unizar.es

ABSTRACT

A Cabin Car Communication System (CCCS) has the goal of improving the communication among passengers inside vehicles [1]. The lack of visual contact between speakers, the high level of noise and many other factors degrade the communications inside vehicles.The CCCS makes use of a set of microphones placed on the overhead to pick up the speech of each passenger, those signals are amplified and played back into the cabin through the car audio loudspeaker system. This system has to deal with two main problems, electro-acoustic coupling and noise amplification. To overcome these problems, CCCS makes use of echo cancellation and noise reduction techniques. In this work a discussion about the echo cancellation implementation with simulation results about the performance of the proposed echo canceller and noise reduction stage are shown.

1 INTRODUCTION

Communication inside vehicles can be difficult due to the high noise level inside the car, the distance among passengers, and many other factors. As a result of that, passengers must rise their voices, move out of their normal seating positions and the driver must look away from the road. The goal of the Cabin Car Communication System (CCCS) is to improve the communication among passengers. With a set of microphones mounted on the overhead, the CCCS picks up the speech of each passenger, amplifies it and plays it back into the cabin. This solution presents two main problems: electro-acoustic coupling and noise amplification.

An Acoustic Echo Canceller (AEC) and an Echo Suppression Filter (ESF) are used to reduce the electro-acoustic coupling problem.The AEC tries to remove the undesired echo that feeds back from the loudspeakers by modeling the echo path impulse response with an adaptive filter in parallel with the echo path. As microphone signals are always composed of the acoustic echo, the noise present in the cabin and the speech signal, the adaptive filter can't converge to a good estimate of the echo path impulse response due to the disturbance caused by the speech, this effect is usually known in telephony as double talk. The classical solution to deal with double-talk is to detect it and freeze the coefficients. However, this can not be the solution for the CCCS because the filter would be always frozen as the system is always in a double talk situation. As a result of this misestimation, acoustic echo will pass through to the amplification stage.The Echo Suppression Filter placed after the Acoustic Echo Canceller performs a further echo attenuation of the echo signal .

This work has been supported by the grants TIC98-0423-C06-04, AMB99-1095-C02, FEDER 2FD97-1070 and the European Technological Center of LEAR Automotive (EED), Valls, Spain

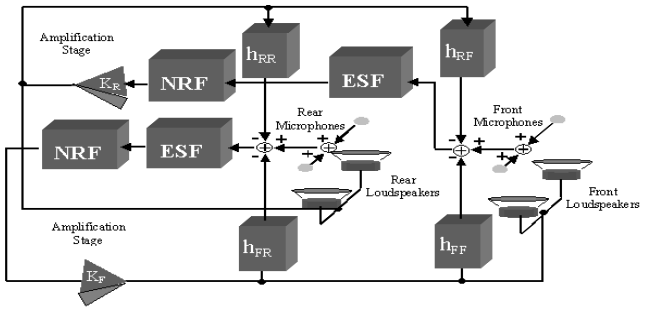

Figure 1: Block Diagram of a Two-Channel CCCS

To avoid increasing the noise present in the cabin a Noise Reduction Filter (NRF) based on the Wiener filter is placed after the Echo Suppression Filter.

Another important aspect of the system is that to maintain the intelligibility, the maximum delay must be less than 20 ms in order to achieve full integration of the direct sound and the CCCS output [2].

In the next section a description of the CCCS will be presented, the Acoustic Echo Canceller will be studied in section 3, the Echo Suppression Filter and the Noise Reduction Filter will be discussed in section 4 and finally, in section 5 some simulation results will be shown.

2 SYSTEM OVERVIEW

A minimum CCCS is composed of at least two channels, one must take the speech of the rear passengers to the front part of the car, and the other one must take the

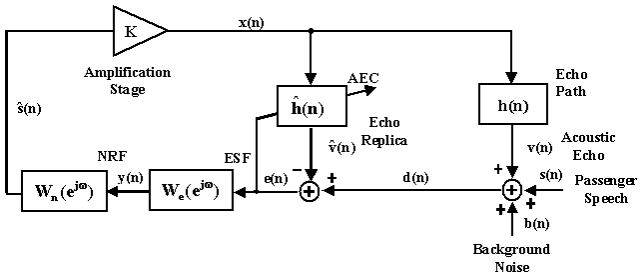

Figure 2: Block Diagram of a One-Channel CCCS

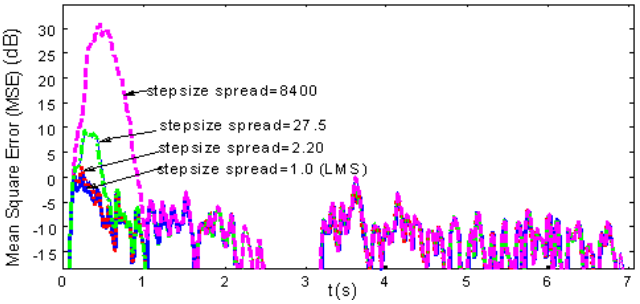

Figure 3: Evolution of the Mean Square Error for different values of the step-size dynamic range.

front passengers speech to the rear seats, as can be seen in the block diagram in figure 1. In a two-channel CCCS (CCCS 2x2), for each channel, there must be two echo cancellers, an Echo Suppression Filter, a Noise Reduction Filter and an amplification stage. For the sake of simplicity, a one-channel system will be described here, the extrapolation to a two-channel system is straightforward.

A block diagram of a one-channel CCCS is shown in figure 2. In this diagram, h(n) represents the impulse response of the echo path, v(n) is the acoustic echo, s(n) the speech signal and b(n) the background noise.

The transfer function of the system in figure 2 between the input signal $s(n) + b(n)$ and the output signal $x(n)$ is

$$P(e^{j\omega}) = \frac{K \cdot W_e(e^{j\omega}) \cdot W_n(e^{j\omega})}{1 - K \cdot W_e(e^{j\omega}) \cdot W_n(e^{j\omega}) \cdot \tilde{H}(e^{j\omega})} \qquad (1)$$

Where $\tilde{H}(e^{j\omega})$ known as misadjustment, is the difference between the transfer function of the echo path $H(e^{j\omega})$ and the transfer function of the adaptive filter $\hat{H}(e^{j\omega})$. Due to the permanent double talk situation, this difference can be significant, and depending on the value of the gain factor K, the denominator in equation 1 can approach to zero, leading the system to instability.

The optimal solution for the ESF $W_e(e^{j\omega})$ that ensures stability and avoids howling without disturbing the Noise Reduction Filter operation can be found making $P(e^{j\omega}) = K \cdot W_n(e^{j\omega})$ which gives

$$W_e(e^{j\omega}) = \frac{1}{1 + K \cdot W_n(e^{j\omega}) \cdot \tilde{H}(e^{j\omega})} \qquad (2)$$

The estimation of the misadjustment function will be discussed in section 4 along with the design of the NRF based on the optimal Wiener solution.

3 ACOUSTIC ECHO CANCELLER

The Acoustic Echo Canceller tries to model the echo path impulse response with a Finite Impulse Response (FIR) filter in parallel with it and subtract an echo replica from the microphone signal. Many algorithms can be considered to perform this task [3]. The choice of one method can depend on the convergence behavior or the computation complexity among other factors. In our case, real time operation must be considered and low computation complexity algorithms as LMS or NLMS must be used. In this section the convenience of using the Least Mean Square (LMS) [4] algorithm versus the Normalized Least Mean Square (NLMS) is discussed.

A very important aspect in the performance of the adaptive algorithm is the choice of the step-size. Because of the feedback in the CCCS a small step-size leads to instability as the adaptive filter is not fast enough to model the echo impulse response before the system starts howling. On the other hand, a step-size too high makes the canceller to perform a bad estimation of the echo path impulse response and thus echo cancellation could be deficient because the misadjustment can be too high. Using the NLMS criterion, that uses a step-size dependent on the inverse of the input signal power is not the best option in this system as the power of the adaptive filter input signal $x(n)$ increases when the power of the error signal $e(n)$ increases according to the diagram in figure 2. As a result of that, when the identification of the echo path impulse response is not accurate enough, the error will be far away from its minimum and the step-size will decrease, making the convergence too slow to track the variations in the echo path resulting in howling or even instability. On the other hand, when we are close to the optimum, the error will be small and the step-size will increase resulting in high misadjustment.

To overcome this problem the best alternative is to reduce the dynamic range of the step-size. In figure 3 the evolution over time of the Mean Square Error is shown for different values of the step-size dynamic range. The curves where obtained with the same steady state performance in terms of distortion and Echo Return Loss Enhancement (ERLE) computed according to equation 8. During the simulations the ESF and the NRF were deactivated. The step-size spread is the ratio of the maximum to the minimum value of the step-size , the distortion was evaluated by means of the Itakura distance between the input speech signal $s(n)$ and the error signal $e(n)$ with an LP model of 10 coefficients. To achieve the same steady state performance the mean step-size, measured over voiced segments, must be used. The value of this mean step-size is 0.0025 with a speech

signal dynamic range of ±1 and an adaptive filter of 350 coefficients. As can be seen in figure 3 the LMS with constant step-size is the best option.

4 ECHO SUPPRESSION AND NOISE REDUCTION FILTERS

4.1 Echo Suppression Filter

As discussed in section 2, the optimal solution for the ESF according to equation 2 was dependent on the misadjustment function $\tilde{H}(e^{j\omega})$. To estimate it, our approach is to design a Wiener filter able to estimate the residual echo existing after the echo canceller. Assuming stationarity on short periods of time, the optimal Wiener solution for the k-th segment is

$$H_r(e^{j\omega};k) = \frac{S_{re}(e^{j\omega};k)}{S_e(e^{j\omega};k)} \tag{3}$$

where $S_e(e^{j\omega};k)$ is the power spectral density (PSD) of the error signal and $S_{re}(e^{j\omega};k)$ is the cross-power spectral density of the residual echo and the error signal. Assuming that the residual echo and the CCCS input signal $(s(n)+b(n))$ are almost uncorrelated signals because of the delay of the overall system, then $S_{re}(e^{j\omega};k) = S_r(e^{j\omega};k)$.

The PSD of the residual echo $r(n)$ can be expressed as follows:

$$S_r(e^{j\omega};k) = S_x(e^{j\omega};k)\left|\tilde{H}(e^{j\omega};k)\right|^2 \tag{4}$$

According to the structure of the system in figure 2

$$S_x(e^{j\omega};k) = K^2 S_e(e^{j\omega};k)\left|W_e(e^{j\omega};k)W_n(e^{j\omega};k)\right|^2 \tag{5}$$

and substituting the misadjustment in equation 2 the optimal ESF results

$$W_e(e^{j\omega};k) = 1 - \sqrt{H_r(e^{j\omega};k)} \tag{6}$$

4.2 Noise Reduction Filter

The CCCS must avoid increasing the noise inside the car reducing the noise picked up by the microphones. This is performed by means of a Wiener filter placed after the ESF, $W_n(e^{j\omega};k)$.

Assuming that the speech signal is uncorrelated with the background noise and proper operation of the ESF [1] the Noise Reduction Filter for the k-th segment can be expressed as follows:

$$W_n(e^{j\omega};k) = 1 - \frac{S_b(e^{j\omega};k)}{S_y(e^{j\omega};k)} \tag{7}$$

where $S_y(e^{j\omega};k)$ is the PSD of the output signal of the ESF and $S_b(e^{j\omega};k)$, the PSD of the noise.

4.3 Power Spectral Densities Estimations

As shown above, it is necessary to know the PSD of the background noise, the residual echo or the error signal to compute the ESF and the NRF and only the error signal $e(n)$ is directly accessible. To obtain an estimation of the PSD of the error signal, $\hat{S}_e(e^{j\omega};k)$, periodogram estimations are used over 16 ms frames. To reduce musical noise effects, a Mel scale frequency smoothing is used. The estimation of the rest of the PSD needs a more elaborated procedure as they are not directly accessible. The PSD of the residual echo, $\hat{S}_r(e^{j\omega};k)$, and the PSD of the background noise, $\hat{S}_b(e^{j\omega};k)$, are recursively computed by means of a biased estimator based on previously estimated Wiener filters. A more detailed description can be found in [1].

5 SIMULATION RESULTS

In order to shown how CCCS improve oral communications inside vehicles some simulation results are shown here. The performance measures were obtained during double-talk periods using 30 ms long blocks. The echo path used for the simulations was a real impulse response measured in a car of 75 ms length. The length of the adaptive filter $\hat{h}(n)$ is 50 ms, corresponding to a filter of 350 coefficients and bulk delay of 50 samples at a sampling rate of 8KHz. This bulk delay is used to compensate for the propagation delay between the loudspeaker and the microphone. Every PSD estimation is updated every 4 ms using a window size of 16 ms. The indexes used to measure the performance of the CCCS were:

1. Echo Return Loss Enhancement (ERLE)

$$ERLE = 10 \cdot log_{10}\left(\frac{1}{N}\sum_{n=0}^{N-1}\frac{E\left[v^2(n)\right]}{E\left[\tilde{r}^2(n)\right]}\right) \tag{8}$$

where $\tilde{r}(n)$ is the residual echo after the NRF and the ESF.This index tells about how much echo is reducing the echo control stage (composed of AEC and ESF).

2. Open-Loop Echo Gain (OLEG)

$$OLEG = 10 \cdot log_{10}\left(\frac{E\left[K^2\tilde{r}^2(n)\right]}{E\left[v^2(n)\right]}\right) \tag{9}$$

Which is the gain of the system when the loop is open but considering only the echo signal. It helps us to know how far is the system from becoming unstable because there exists a limit for this value, so that, higher values of OLEG lead the system to instability. The maximum value of the $OLEG$ for the echo path impulse response used during simulation is $OLEG_{max} = 20 \cdot log_{10}(0.4) = -7.9588\,dB$.

3. Stability Margin (SM)

$$SM = OLEG_{max} - OLEG(K) \tag{10}$$

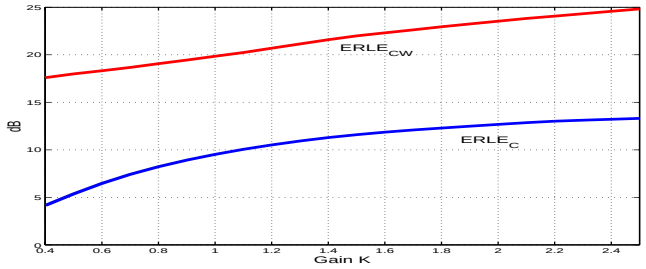

Figure 4: Evolution of the ERLE over different values of K with and without the ESF.

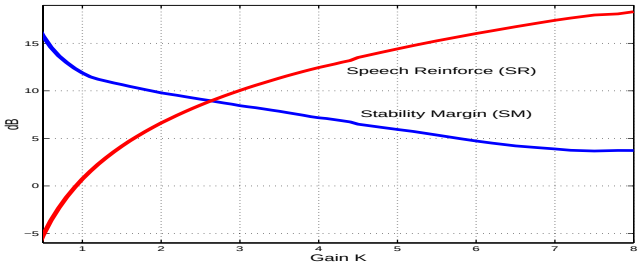

Figure 5: Speech Reinforce and Stability Margin vs. Gain K

The nearer this value is to zero the closer the system is to become unstable as $OLEG$ never can be greater than $OLEG_{max}$.

4. Speech Reinforce (SR)

$$SR = 10 \cdot log_{10} \left(\frac{\frac{E[x^2(n)]}{E[s^2(n)]}}{0.4} \right) \qquad (11)$$

Defined as the ratio, in dB, of the speech signal power gain to the maximum reinforce achievable without the acoustic echo canceller, the Echo Suppression Filter and the Noise Reduction Filter. This maximum value is 0.4 because there is no echo attenuation without AEC, ESF and NRF, therefore, without echo control, the echo gain is equal to K and equal to the speech signal gain.

The need of using the ESF can be clearly seen in figure 4 where the evolution of the ERLE for different values of the gain K is shown with and without the ESF. An ERLE 10dB higher is achieved using the ESF along with the Acoustic Echo Canceller. The maximum value of K without the ESF is 2.5, higher values of K leads the system to instability.

The evolution of the Speech Reinforce and the Stability Margin can be seen in figure 5. The Stability Margin decreases as the gain factor K increases and the maximum value of K that ensures stability is around 8 which gives a Speech Reinforce near to 20 dB. Nevertheless with this values of the gain factor K the distortion is noticeable and lower values must be used to achieve good speech quality. Values around 6 ensures acceptable speech quality avoiding howling and ensuring stability. With this value of the gain factor K the Stability Margin is 4.75 dB and the Speech Reinforce is greater than 16 dB. To evaluate the performance of the Noise Reduction Filter the segmental signal to noise ratio improvement was used defined as:

$$\Delta SNR = 10 \cdot log_{10} \left(\frac{1}{N} \sum_{k=0}^{N-1} \frac{SNR_o(k)}{SNR_i(k)} \right) \qquad (12)$$

where $SNR_o(k)$ and $SNR_i(k)$ are the output and input, respectively, signal to noise ratio for the k-th 30 ms segment. This signal to noise ratio improvement doesn't have significant variations over different values of the input SNR, in fact ΔSNR ranges from 8 dB to 10 dB for input SNR ranging from 0 to 30 dB. The measures were obtained over voiced segments using real car noise. The SNR improvement is around 25 dB over silence frames.

6 CONCLUSIONS

In this paper a discussion about the echo control performed by the CCCS has been presented. This is a challenging task due to the feedback nature of the system that can make CCCS to become unstable. Regarding the Acoustic Echo Canceller, the convenience of using LMS criterion versus NLMS has been discussed. Using a fixed step-size achieves better convergence behavior with the same steady state performance than using a step-size dependent on the inverse of the input signal power.The way to obtain a Echo Suppression Filter placed after the AEC able to achieve further echo attenuation is also presented. Simulation results show that the presence of this ESF allows higher levels of Speech Reinforce maintaining stability and speech quality. A full Two-Channel CCCS has been developed and tested in a medium-size car giving high speech reinforces with guaranteed stability and good speech quality even in adverse situations like doors openings or passengers movements.

References

[1] E. Lleida, E. Masgrau, and A. Ortega, "Acoustic echo and noise reduction for cabin car communication," in *Proceeding of Eurospeech 2001*, vol. 3, pp. 1585–1588, September 2001.

[2] H. Hass, "The influence of a single echo on the audibility of speech," *Acustica*, vol. 1, no. 2, 1951.

[3] C. Breining, P. Dreiseitel, E. Hänsler, A. Mader, B. Nitsch, H. Puder, T. Schertler, G. Schmidt, and J. Tilp, "Acoustic echo control. an application of very-high-order adaptive filters," *IEEE Signal Processing Magazine*, vol. 16, pp. 42–69, July 1999.

[4] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Prentice-Hall, Inc, 1985.