Acoustic Echo Control and Noise Reduction for Cabin Car Communication

Eduardo Lleida, Enrique Masgrau, Alfonso Ortega

Department of Electronics Engineering and Communications
University of Zaragoza, Zaragoza, Spain
lleida@posta.unizar.es

Abstract

A Cabin Car Communication System (CCCS) has the goal of improving the communication among passengers inside the car. Wind, road and engine noise, the distance between passengers and other factors make difficult the communication inside vehicles. The driver must often look away from the road and passengers move out of normal seating positions. The CCCS makes use of a set of microphones to pick up the speech and the car-audio loudspeakers to reinforce the sound level. This scenario presents a great challenger for acoustic echo control and noise reduction. Acoustic echo control must prevent the overall system from howling and becoming unstable with the additional problem that the system must always work with double talk. The noise reduction must clean the microphone signal to avoid the reinforce of the noise inside the car. In this paper, we describe a combined acoustic echo control and noise reduction algorithm suitable for cabin car communication systems. We present experimental results in terms of echo return loss enhancement, stability and maximum reinforce without howling.

1. Introduction

Cabin Car Communication (CCCS) is a great challenge for acoustic echo control and noise reduction. Several microphones mounted in front of each passenger pick up the speech signal plus some noise from the engine, road and so on. This signal is amplified and return to the cabin through the car-audio loudspeaker system. This scenario creates two main problems. First, as a result of the electro-acoustic coupling between loudspeakers and microphones, the overall system may become unstable with the annoying effect of howling. Second, as the microphones pick up speech and noise, the overall noise level inside the cabin will increase. So, an acoustic echo canceller is needed to prevent the overall system from howling and a noise reduction is needed to avoid the reinforce of the overall noise inside the cabin.

Acoustic echo cancellation is performed by means of an adaptive filter parallel to the loudspeaker-cabin-microphone system (LEM path) [1]. In the CCCS, the input signal is speech and additive noise which lead to a permanent disturbance of the adaptive filter with continuous double-talk and noise in the error signal. Acoustic echo is produced by the near-end speech, so the acoustic echo canceller must always deal with echo and near-end speech.

A classical solution to this problem proposes to freeze the echo canceller when double talk is detected [2]. However, due to the continuous changes on the LEM path, the system could become unstable and start howling while the passenger is talking. Some solutions proposes to add a low-level white noise suitable to identify the LEM path when no near-end signal is present. However, this low-level white noise could be annoying if the loudspeakers are close to the passengers. The CCCS proposed in this paper doesn't make use of either freezing the echo canceller or low-level white noise.

Another important aspect of the CCCS is the delay introduced by the overall system. As the passengers receive the direct speech and the speech reinforced by the CCCS, the delay between both sounds must be less than 20 ms to full integration of the sounds, maintaining the intelligibility [3].

In this paper, we present the performance study of a one channel CCCS. The CCCS consists on an adaptive acoustic echo canceller, a Wiener echo suppression filter and a Wiener noise reduction filter. In section 2, we present the one channel CCCS putting the accent on the aim of each component of the system. The adaptive acoustic echo canceller is studied in section 3. The Wiener echo suppression and noise reduction filter are presented in section 4. In section 5, the experimental setup is described and the experimental results are presented.

2. One Channel Cabin Car Communication System

A full communication between passengers requires at least a two channel CCCS. However, this system has four acoustic echo paths which makes difficult to study the performance of the overall system. In this paper, a one channel CCCS is used to study the problems associated with the cabin car communication. The extrapolation to a two channel CCCS is straightforward from the one channel. The one channel system could be used to reinforce the communication between the front passengers and the rear passengers. Two microphones, one in front of the driver, and another in front of the co-driver passenger pick up the speech signal. The output of the CCCS is applied to the rear loudspeakers. Figure 1 shows the block diagram of the CCCS. In this diagram, h(n) represents the impulse response of the LEM path, s(n) is the near-end speech, b(n) is the background noise and K is the amplification gain.

In absence of background noise and with the echo canceller and noise reduction system deactivated, the transfer function between the input signal d(n) and the output signal o(n) is,

$$P(z) = \frac{O(z)}{S(z)} = \frac{K}{1 - KH(z)} \qquad (1)$$

Because of the phase behavior of the loop transfer function, if $|KH(e^{jw})|>1$, the closed-loop system is likely to become unstable and the undesirable howling effect appears. So, the echo canceller, $\hat{h}(n)$, attempts to identify the acoustic path

This work has been partially fund by the grants TIC98-0423-C06-04, AMB99-1095-C02 and FEDER 2FD97-1070.

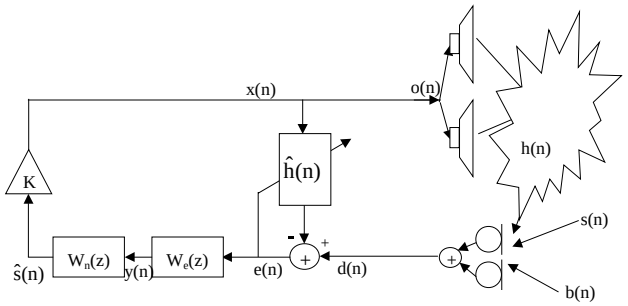

Figure 1. One channel CCCS Block diagram. s(n): speech, b(n): background sound (noise + music).

h(n) to cancel the feedback. Including the echo canceller, the transfer function P(z) is,

$$P(z) = \frac{K}{1 - K\left(H(z) - \hat{H}(z)\right)} \tag{2}$$

where, if $\hat{H}(z) = H(z)$ the desired compensation of the feedback is obtained. However, this condition is impossible to accomplish for several reasons. First of all, we don't know the exact length of the impulse response of the acoustic path and more over, h(n) is time variant so there will always be a misadjustment between h(n) and the estimated $\hat{h}(n)$. That means that it is necessary to cancel the residual echo to prevent the system from howling. To this purpose, we introduce the filter $W_e(z)$, also called echo suppression filter. Adding the new filter, the transfer function P(z) becomes,

$$P(z) = \frac{KW_e(z)}{1 - KW_e(z)M(z)} \tag{3}$$

where, M(z) is de misadjustment, and it is equal to,

$$M(z) = H(z) - \hat{H}(z) \tag{4}$$

So, to avoid howling, the following condition must be true

$$\left|W_e(e^{j\omega})\right|\left|M(e^{j\omega})\right| < 1/K \tag{5}$$

A trivial solution to avoid howling and distortion is to make the echo suppression filter equal to

$$W_e(z) = \frac{1}{1 + KM(z)} \tag{6}$$

which makes P(z)=K. However, the misadjustment function M(z) is unknown, and it is difficult to get a good estimate of it due to the presence of near-end signal n(n)=s(n)+b(n). So, in this work, the optimal Wiener solution is adopted to design the echo suppression filter. Assuming stationarity on short period of time (10 – 20 ms. with speech signals), the optimal Wiener solution for the k-th stationary segment will be

$$W_e(z;k) = \frac{S_{en}(z;k)}{S_e(z;k)} \tag{7}$$

where $S_e(z;k)$ is the complex spectral density of the error signal and $S_{er}(z;k)$ is the complex cross-spectral density of the error signal and the near-end signal. The error signal consists of the sum of the near-end signal and the residual echo signal, e(n)=n(n)+r(n). Due to the delay of the overall system, we can assume that the residual echo, r(n), and the near-end signal, n(n), are almost uncorrelated, then

$$\begin{aligned} S_{en}(z;k) &= S_n(z;k) \\ S_e(z;k) &= S_r(z;k) + S_n(z;k) \end{aligned} \tag{8}$$

The frequency response of the optimal Wiener solution for the echo suppression filter can be expressed as follows:

$$\begin{aligned} W_e(e^{j\omega};k) &= \frac{S_n(e^{j\omega};k)}{S_r(e^{j\omega};k) + S_n(e^{j\omega};k)} \\ &= 1 - \frac{S_r(e^{j\omega};k)}{S_r(e^{j\omega};k) + S_n(e^{j\omega};k)} \\ &= 1 - \frac{S_r(e^{j\omega};k)}{S_e(e^{j\omega};k)} \end{aligned} \tag{9}$$

The power spectral density of the error signal $S_e(e^{j\omega};k)$ is directly estimated from the error signal e(n), but the estimation of power spectral density of the residual echo $S_r(e^{j\omega};k)$ or the power spectral density of the near-end signal $S_n(e^{j\omega};k)$ requires a more elaborated procedure because there is no direct access to any of both signals.

Finally, the CCCS has to reduce the noise signal present at the microphone to avoid increasing the overall noise in the cabin. The noise reduction is performed with a Wiener filter, $W_n(z)$, after the echo suppression filter. Assuming the same assumptions used to develop the echo suppression filter, the noise reduction filter can be expressed as follows:

$$W_n(e^{j\omega};k) = \frac{S_{ys}(e^{j\omega};k)}{S_y(e^{j\omega};k)} = \frac{S_s(e^{j\omega};k)}{S_y(e^{j\omega};k)} \tag{10}$$

where $S_s(e^{j\omega};k)$ is the power spectral density of the near-end speech signal s(n) and $S_y(e^{j\omega};k)$ is the power spectral density of the echo suppression filter output $y(n) \approx (s(n) + b(n))$, assuming the proper work of the echo suppression filter. The noise reduction filter could be defined in terms of the power spectral density of the background noise $S_b(e^{j\omega};k)$ as follows:

$$W_n(e^{j\omega};k) = 1 - \frac{S_b(e^{j\omega};k)}{S_y(e^{j\omega};k)} \tag{11}$$

where $S_y(e^{j\omega};k)$ could be estimated directly from the signal y(n), but the estimation of $S_b(e^{j\omega};k)$ requires a more elaborated procedure which will be discussed in the section 4 jointly with the estimation of $S_r(e^{j\omega};k)$ in eq. 9.

3. Adaptive Acoustic Echo Canceller

As shown in the previous section, the acoustic echo canceller plays an important role in the stability of the CCCS. Due to the continuos movement of the passengers and the variation on the temperature, the LEM path impulse response is continuously changing, so it is necessary to continuously identify the LEM path. A great difference between the CCCS and other echo cancellation systems is the feedback loop. The feedback loop plays the role of amplifying the signal. The adaptive algorithm is used to update the acoustic echo cancellation filter is the NLMS (normalized least mean square) given by

$$\hat{\underline{\mathbf{h}}}(n+1) = \hat{\underline{\mathbf{h}}}(n) + \mu(n)e(n)\underline{\mathbf{x}}(n) \tag{12}$$

with

$$\mu(n) = \frac{\alpha}{L\hat{\sigma}_x^2(n)} \tag{13}$$

where L is the echo-canceller length, α is the step-size parameter that controls the stability and convergence rate and $\hat{\sigma}_x^2(n)$ is the estimation of the power of the input signal x(n). Due to the delays in the acoustic path, A/D and D/A conversors, a bulk delay of D samples is used. As the acoustic path delay decrease with the temperature, D is chosen for the worst case, assuming the higher temperature in the cabin.

The choice of the step-size μ(n) is critical for the good performance of the CCCS. A small value makes the system unstable due to the low convergence time, the system starts howling before the convergence of the filter is achieve. However, a large value provide a faster convergence with a good tracking capabilities but with a higher excess mean-squared error. That means a bad estimation of the LEM path which implies an increase in the echo, distortion and decrease in the stability margin of the CCCS. The NLMS algorithm has been selected because of its computational simplicity, which is needed for a real-time implementation of the two channel CCCS with four adaptive echo-canceller.

4. Wiener Echo Suppression and Noise Reduction filters

In section 2, the need of the Wiener echo suppression and noise reduction filters has been argued. In this section, the design of these filters is presented.

4.1. Wiener Noise Reduction Filter

Equation 11 defines the way to compute the Wiener noise reduction filter. First, an estimation of the signal plus noise power spectral density $S_y(e^{j\omega};k)$ is computed directly from the input signal y(n) using a periodogram estimation. To reduce the musical noise associated to this kind of noise reduction filters, a frequency smoothing is performed using a smoothing window distributed in a mel scale, given an estimated $\hat{S}_y(e^{j\omega};k)$.

The background power spectral density estimation is performed as follows:

- Assuming stationary background noise
- The Wiener solution, eq. 11, is written as

$$W_n(e^{j\omega};k) = 1 - \hat{H}_n(e^{j\omega};k) \quad (14)$$

with

$$\hat{H}_n(e^{j\omega};k) = \frac{\hat{S}_b(e^{j\omega};k-1)}{\hat{S}_y(e^{j\omega};k)} \quad (15)$$

that can be interpreted as a Wiener filter to estimate the background noise signal b(n). A exponential smoothing is performed over the time evolution of $\hat{H}_n(e^{j\omega};k)$ to reduce the musical noise. Then, assuming we know this filter, an estimation of the background power spectral density for the k-th signal segment can be found as,

$$\tilde{S}_b(e^{j\omega};k) = \left[\lambda_n + (1-\lambda_n)\hat{H}_n(e^{j\omega};k)\right]^2 \hat{S}_y(e^{j\omega};k) \quad (16)$$

where $0<\lambda_n<1$, is a bias term to avoid clipping of any frequency of $\hat{H}_n(e^{j\omega};k)$ to 0.

- Finally, the background power spectral density is computed averaging over time with a forgetting factor ∂_n as

$$\hat{S}_b(e^{j\omega};k) = \partial_n \hat{S}_b(e^{j\omega};k-1) + (1-\partial_n)\tilde{S}_b(e^{j\omega};k) \quad (17)$$

where $0<\partial_n<1$. In this case, with the assumption of stationary background noise and non-stationary speech signal, ∂_n is very close to 1 (0.995) to avoid the use of a voice activity detector.

4.2. Wiener Echo Suppression Filter

As it is difficult to isolate the residual echo signal from the speech and noise, an approach similar to the previous one is used to estimate the Wiener echo suppression filter. The main difference is that now the filter has to track a signal that is not stationary, so the assumption of stationarity could not be used at all. However, the bulk delay could be used to separate the near-end signal from the echo signal. The echo signal is the previous near-end signal with higher correlation due to the LEM path filtering with a delay of D samples with regard to the actual near-end signal. That means that it is possible to estimate the residual echo using a Wiener approach as in the noise reduction filter.

- The Wiener solution, eq. 9, could be written as

$$W_e(e^{j\omega};k) = 1 - \hat{H}_e(e^{j\omega};k) \quad (18)$$

with

$$\hat{H}_e(e^{j\omega};k) = \frac{\hat{S}_r(e^{j\omega};k-1)}{\hat{S}_e(e^{j\omega};k)} \quad (19)$$

- The power spectral density $\hat{S}_e(e^{j\omega};k)$ is computed from the signal e(n) as done for y(n) for the noise reduction filter.
- The estimation of $\hat{S}_r(e^{j\omega};k)$ is performed using the same approach that for the noise reduction filter with the difference of the last step

$$\tilde{S}_r(e^{j\omega};k) = \left[\lambda_e + (1-\lambda_e)\hat{H}_e(e^{j\omega};k)\right]^2 \hat{S}_e(e^{j\omega};k) \quad (20)$$

$$\hat{S}_r(e^{j\omega};k) = \partial_e \hat{S}_r(e^{j\omega};k-1) + (1-\partial_e)\tilde{S}_r(e^{j\omega};k) \quad (21)$$

where $0<\partial_e<1$. As the echo residual is a non stationary signal, the value of ∂_e must give a short time memory factor. However, as the echo residual signal has a longer time correlation than the near-end signal due to the LEM path, it is possible to improve the estimation with values of ∂_e equivalent to time constants between 10 and 30 ms.

5. Simulation Results

Figure 2 shows the simulation setup used to measure the performance of the one channel CCCS. The performance is studied in terms of:

- Echo Return Loss Enhancement (ERLE) defined as

$$ERLE = 10\log_{10}\left(\frac{\frac{1}{N}\sum_{n=0}^{N-1}\frac{E\{p^2(n)\}}{E\{\tilde{r}^2(n)\}}}{}\right) \quad (22)$$

where p(n) is the echo signal in the microphone and $\tilde{r}(n)$ is the final echo residual, N is the number of signal blocks used to average the estimation of the ERLE. Each estimation block is 30 ms long and only double talk segments have been used to compute ERLE using a simple energy VAD.

- Speech reinforce defined as the ratio, in dB, of the speech signal power gain and the maximum reinforce obtained without the echo and noise cancellers.
- Open-Loop Echo Gain (OLEG) defined as the gain in open-loop for the echo path

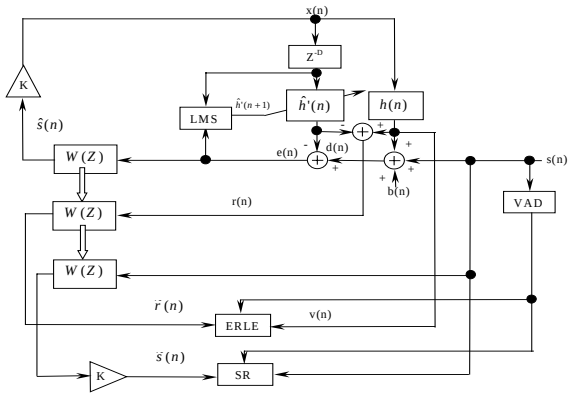

Figure 2. Simulation setup for performance measure

$$\mathrm{OLEG} = 10\log_{10}\left(\frac{E\{K^2\tilde{r}^2(n)\}}{E\{v^2(n)\}}\right) \qquad (23)$$

where and $K\tilde{r}(n)$ is the final residual echo at the output and v(n) is echo signal at the microphone. So, the open-loop echo gain could be computed as a function of ERLE and the gain K as follows

$$\mathrm{OLEG} = 20\log_{10}(K) - \mathrm{ERLE} \qquad (24)$$

Figure 3 shows the LEM path impulse response, h(n), used on the simulations. The sampling frequency is 8 kHz. The length of the FIR adaptive filter is 350 (43,7 ms) with a bulk delay of 50 samples (6,3 ms) and the Wiener echo suppression and noise reduction filters are computed using a 128 points FFT with a time overlap of 96 samples to keep the overall delay of the system (including A/D and D/A) lower than 20 ms. The step-size μ(n) has been selected to maintain a fast convergence time with a negligible distortion in the speech signal (the distortion was quantified by means of the symmetric Itakura distance between the input and output speech).

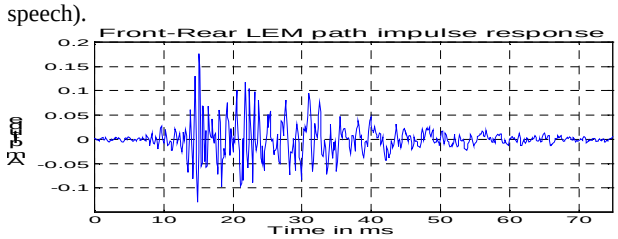

Figure 3. LEM path impulse response from the front microphone to the rear loudspeakers.

Without echo canceller and Wiener filters, ERLE=0 dB, the maximum value of the gain K with a guarantee of stability is around 0.4, which gives a $\mathrm{OLEG}_{max}=20\log_{10}(0.4)=-7.9588$. Figure 4 shows the ERLE evolution when using only the echo canceller and when using the combined system with the echo canceller and Wiener filters as a function of the gain K. Note that a gain of more than 10 dB is obtained in the combined system, which supposes an improvement in the margin of stability. Also, the ERLE increases with the increase of K as a result of a lower misadjustment due to the relative increase of the echo signal with regard to the near-end signal.

Figure 5 shows the evolution of the Speech Reinforce (SR) and the Stability Margin (SM) defined as

$$\mathrm{SM} = \mathrm{OLEG}_{max} - \mathrm{OLEG}$$

values of SM negative or close to 0 means that the system may become unstable and start howling. The maximum value of the gain K without howling with the combined system is around 7.5 where SM is 1.5 dB and the SR is 19 dB. However, for this value of K the distortion in the speech signal is noticeable. The maximum value of K for a negligible distortion is 5, which gives a SR of 15 dB and a SM of 4 dB.

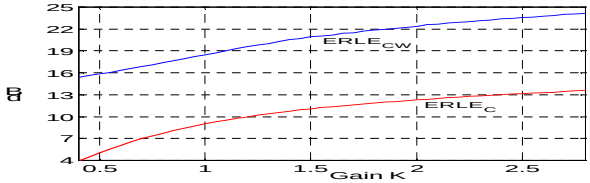

Figure 4. ERLE_{CW}: ERLE combined system, ERLE_{C}: ERLE echo canceller alone.

The maximum theoretical value of K with the echo canceller alone and no misadjustment (perfect identification of the first 400 coefficients of h(n)) is around 3 (can be found by computing the roots of denominator of eq. 2). However, if a misadjustment is allowed, the maximum value of K could be increased up to 6 which means that the echo canceller tries to maintain the stability of the system at the prize of worst identification of the LEM path.

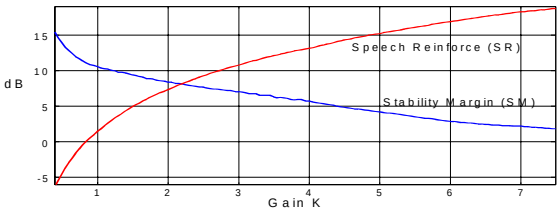

Figure 5. Speech Reinforze and Stability Margin vs. gain K

6. Conclusions

In this paper we have presented a cabin car communication system to improve the communication among passengers inside a car. A one channel CCCS has been studied and defined. The extrapolation to a two-channel CCCS is straightforward but it is out of the length of this paper. The proposed system is able to work with double-talk maintaining the stability of the system up to speech reinforces of 20 dB. The main conclusion of this work is the need of the use of an echo canceller jointly with an echo suppression filter which allow to get an acceptable speech reinforce. We are currently testing the two-channel CCCS in real conditions in a mini-van using four echo cancellers and two echo suppression and noise reduction filter in a DSP. First test results show a good performance of the system, given an acceptable speech reinforce with low distortion and no howling.

7. References

[1] Breining, C., Dreiseitel, P., Hänsler, E., Mader, A., Nitsch B., Puder H., Schertler T., Schmidt G., Tilp J."Acoustic Echo Control", *IEEE Signal Processing Magazine*, Vol. 16, No. 4, pp. 42-69, 1999.

[2] B. M. Finn, "Acoustic Echo Cancellation in an Integrated Audio and Telecommunication System", *U.S. Patent 5,706,344*, Jan. 6, 1998.

[3] Haas H. "The influence of a Single Echo on the Audibility of Speech", *Acustica*, vol. 1, no. 2, 1951, in German. (English translation by K. Ehrenberg in *Journal of the Audio Engineering Society*, vol 20, n 2, pp 146-159, March 1972).