

Acoustic Feedback Cancellation in Speech Reinforcement Systems for Vehicles

Alfonso Ortega, Eduardo Lleida, Enrique Masgrau, Luis Buera and Antonio Miguel

Communication Technologies Group (GTC)
 Aragon Institute of Engineering Research (I3A), University of Zaragoza, Spain
 {ortega,lleida,masgrau,lbuera,amiguel}@unizar.es

Abstract

Passengers communication inside a car can be improved by using a speech reinforcement system. This system picks up the speech of each passenger, amplifies it and plays it back into the cabin through the loudspeakers of the car. Due to the electro-acoustic coupling between loudspeakers and microphones, a closed-loop system is created. To avoid the risk of instability due to the acoustic feedback, acoustic echo cancellation must be performed. Using the Minimum Mean Square Error (MMSE) criterion to adapt the filter, what is very common in acoustic echo cancellation, leads to inaccurate estimates of the Loudspeaker-Enclosure-Microphone (LEM) path due to the closed-loop operation of the system. In this paper, the solution obtained with the MMSE criterion for a Finite-length Impulse Response (FIR) causal adaptive filter is derived, showing that the identification error depends on the amplification factor of the system, the delay of the loop and the spectral characteristics of the excitation signal. The use of whitening filters is proposed and justified to improve the acoustic echo cancellation in speech reinforcement systems for cars. Results obtained for a one-channel speech reinforcement system are presented.

1. Introduction

Inside a car, intelligibility can be degraded due to the high level of noise, the use of sound absorbing materials and the lack of visual contact between speakers. A speech reinforcement system for vehicles helps to improve passengers communications inside the car. It is composed of a set of microphones placed on the ceiling of the car, an amplification stage and a set of loudspeakers that delivers the voice of each speaker to the rest of the passengers in order to improve intelligibility [1]. As the distance between loudspeakers and microphones is relatively small, the signal radiated by the loudspeakers is picked up by the microphones creating a closed loop that can make the system becoming unstable and limits the maximum gain than can be used. To prevent this, acoustic echo cancellation is needed. An acoustic echo canceller uses an adaptive filter, parallel to the Loudspeaker-Enclosure-Microphone (LEM) path. This filter must identify the impulse response of this path in order to obtain a replica of the echo signal and subtract it from the microphone signal. In acoustic echo cancellation, the minimization of the Mean Square Error (MMSE) is widely used, defining the error signal as the difference between the microphone signal and the output of the adaptive filter. Nevertheless, due to the closed-loop operation of the speech reinforcement system, the input of the adaptive filter is a delayed and amplified version of the error signal. Because of this, the MMSE solution depends on

This work has been supported by the project TIC2002-04103-C03-01 from the Spanish MEC.

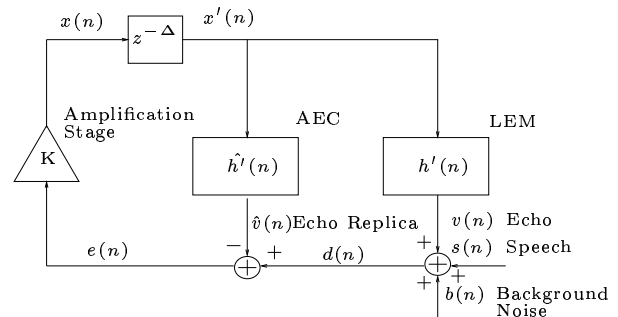


Figure 1: Diagram of a one-channel car cabin communication system

the characteristics of the input signal and in many cases it does not correspond to the desired solution. In this paper, the solution obtained with the MMSE criterion for a causal FIR filter is discussed. In order to improve the identification of the LEM path in a speech reinforcement system, the use of whitening filters is proposed and justified. The use of these decorrelation algorithms has the goal of avoiding the bias term in the identification of the LEM path since the use of the MMSE criterion without using whitening filters leads to a set of non-linear equations. The solution of these non linear equations may not be the identification of the LEM path and may be not unique.

This paper is organized as follows: a brief description of the system is given in Section 2. A study about the MMSE criterion in closed-loop systems is presented in Section 3. The proposed modification of the cost function is considered in Section 4. Some simulation results are presented in Section 5 and finally, the conclusions in Section 6.

2. One-Channel speech reinforcement system

In a one-channel speech-reinforcement system, the speech signal of the speaker $s(n)$ is picked up by the microphones along with the background noise $b(n)$ and the echo signal $v(n)$. The echo canceller filter $\hat{h}(n)$ must model the LEM path impulse response $h(n)$, and obtain an echo replica $\hat{v}(n)$ by filtering the output signal $x(n)$. This echo replica is subtracted from the microphone signal $d(n)$ creating the error signal $e(n)$. The error signal is amplified, multiplying it by a gain factor K , to obtain the output signal of the system, $x'(n)$.

Due to the propagation delay, the LEM path can be modeled as a delay block of Δ samples followed by a linear filter $h'(n)$. Fig. 1 shows the simplified block diagram of the speech reinforcement system with this decomposition, where $\hat{h}'(n)$ is the adaptive filter without the first Δ coefficients that are set to zero to compensate for the propagation delay.

3. MMSE Criterion in a Closed-Loop System

3.1. Time Domain Analysis of the MMSE Solution in a Closed-Loop System

The minimum mean square error criterion is widely used to perform system identification in echo cancellation [2]. Classical optimal filtering theory applied to open loop systems states that the necessary and sufficient condition for the cost function to attain its minimum value is, for the corresponding value of the error signal $e(n)$, to be orthogonal to each input sample, $x'(n)$, that enters to the linear filter

$$E[x'(n-k)e(n)] = 0, \quad k = 0, 1, 2, \dots, \quad (1)$$

where $E[\cdot]$ denotes the expectation of the quantity between the brackets.

Due to the closed-loop operation of the proposed system, the input signal to the adaptive filter $x'(n)$ is a delayed and amplified version of the error signal $e(n)$, so the orthogonality principle is no more the necessary and sufficient condition to attain the minimum value of the cost function.

According to Fig. 1, we obtain the gradient vector of the mean squared error with respect to the coefficients of the adaptive filter and set it equal to zero for the cost function to attain its minimum value.

In the proposed system, the input signal to the adaptive filter is not independent from its coefficients since it is a delayed and amplified version of the error signal. Thus, the condition for the cost function to attain its minimum in this closed-loop system can be expressed as [3]

$$\sum_{i=0}^{\infty} K^i \tilde{h}'^{(*i)}(n) * E[e(n)e(n - (i+1)\Delta - k)] = 0, \quad k = 0, 1, 2, \dots, \quad (2)$$

where $\tilde{h}'(n) = h'(n) - \hat{h}'(n)$ is the weight misadjustment, and

$$\tilde{h}'^{(*i)}(n) = \mathcal{F}^{-1} \left\{ \left(\tilde{H}'(e^{j\omega}) \right)^i \right\}. \quad (3)$$

That is, $\tilde{h}'(n)$ convolved with itself i times. Thus, $\tilde{h}'^{(*i)}(n)$ is equal to $\delta(n)$ when $i = 0$.

It can be shown that, [3], perfect identification of the LEM path is the solution of (2), if and only if the excitation signal satisfies

$$E[s(n)s(n-k-\Delta)] = 0, \quad k = 0, 1, 2, \dots, \quad (4)$$

To allow this, the length of the autocorrelation function of the input signal must be less than the delay Δ with $\Delta > 0$. This means that for a delay Δ equal to 1, the only signal that allows perfect identification of the LEM path is a white noise process.

3.2. Frequency Domain Analysis of the MMSE Solution in a Closed-Loop System

3.2.1. Unconstrained Adaptive Filter

We can express the optimization criterion for the adaptive filter in the frequency domain as

$$\hat{h}'_{opt} = \arg \min_{\hat{h}'} \frac{1}{2\pi} \int_0^{2\pi} S_e(e^{j\omega}) d\omega \quad (5)$$

where $S_e(e^{j\omega})$ is the Power Spectral Density (PSD) of $e(n)$.

According to Fig. 1 and assuming that no background noise is present, the error signal PSD can be expressed as

$$S_e(e^{j\omega}) = \left| \frac{1}{1 - K\tilde{H}'(e^{j\omega})e^{-j\omega\Delta}} \right|^2 S_s(e^{j\omega}) \quad (6)$$

where $S_s(e^{j\omega})$ is the PSD of the speech signal and $\tilde{H}'(e^{j\omega})$ is the difference between the LEM path transfer function, $H'(e^{j\omega})$, and the adaptive filter transfer function, $\hat{H}'(e^{j\omega})$.

In order to obtain a solution to (5), the excitation signal is modeled as a random process generated by applying a white noise process, $w(n)$, with variance σ_w^2 to a linear invertible and monic filter whose transfer function is $A(e^{j\omega})$. After modeling $s(n)$ as described before, it can be shown that the filter that minimizes the power of the error signal is [4]

$$\hat{H}'(e^{j\omega}) = H'(e^{j\omega}) + \frac{A(e^{j\omega}) - 1}{Ke^{-j\omega\Delta}}, \quad (7)$$

which depends on the LEM path transfer function and the model of the excitation signal. According to (7), the only solution that achieves perfect identification of the LEM path can be found when the excitation signal is a white noise process, that is, when $A(e^{j\omega}) = 1$.

3.2.2. Causal FIR Adaptive Filter

Depending on the LEM path transfer function and the input signal model, (7) can not be met when constraining the echo canceller to use a causal FIR filter. Thus, assuming that $\hat{h}'(n)$ is an FIR causal filter of length N , we will obtain the value for the transfer function that minimizes the power of the error signal under certain assumptions.

First of all, we consider that the LEM path transfer function is composed of two parts

$$H'(e^{j\omega}) = H'_1(e^{j\omega}) + H'_2(e^{j\omega}), \quad (8)$$

where $H'_1(e^{j\omega})$ refers to the first N coefficients of the LEM path impulse response and $H'_2(e^{j\omega})$ refers to the rest. The model of the excitation signal can be also divided into two parts

$$A(z) = \sum_{l=0}^{\infty} a_l z^{-l} = A_1(z)z^{-\Delta} + A_2(z), \quad (9)$$

defining $A_1(z)$ as

$$A_1(z) = \sum_{l=0}^{N-1} a_{(l+\Delta)} z^{-l}. \quad (10)$$

According to these definitions, the error signal in the frequency domain, can be expressed as

$$E(e^{j\omega}) = \left[F_1(e^{j\omega})e^{-j\omega\Delta} + F_2(e^{j\omega}) + 1 \right] \sigma_w, \quad (11)$$

where

$$F_1(e^{j\omega}) = \frac{A_1(e^{j\omega}) + K \left[H'_1(e^{j\omega}) - \hat{H}'(e^{j\omega}) \right]}{1 - Ke^{-j\omega\Delta} \tilde{H}'(e^{j\omega})}, \quad (12)$$

$$F_2(e^{j\omega}) = \frac{A_2(e^{j\omega}) - 1 + Ke^{-j\omega\Delta} H'_2(e^{j\omega})}{1 - Ke^{-j\omega\Delta} \tilde{H}'(e^{j\omega})}, \quad (13)$$

and the cost function to minimize

$$J = E[|e(n)|^2] = E[|e_A(n)|^2] + E[|e_B(n)|^2] + 2E[|e_A(n)e_B(n)|^2] + \sigma_w^2, \quad (14)$$

with

$$e_A(n) = \mathcal{F}^{-1} \left\{ F_1(e^{j\omega}) e^{-j\omega\Delta} \sigma_w^2 \right\} \quad (15)$$

$$e_B(n) = \mathcal{F}^{-1} \left\{ F_2(e^{j\omega}) \sigma_w^2 \right\}. \quad (16)$$

The exact value for the third term in (14) can be difficult to find, but under the assumption that the system is far from instability, $K \ll \frac{1}{\hat{H}'(e^{j\omega})}$, this term is almost zero, [3], and the value for the transfer function of the adaptive filter that minimizes the mean squared error is

$$\hat{H}'_{opt}(e^{j\omega}) = H'_1(e^{j\omega}) + \frac{A_1(e^{j\omega})}{K}, \quad (17)$$

which depends on the first N coefficients of the LEM path impulse response, the value of K and the part of the input signal model from coefficient Δ to coefficient $N + \Delta$.

4. Minimization of the Mean Squared Filtered Error

4.1. Minimum Mean Squared Filtered Error Condition in the Time Domain

Since the solution of the MMSE criterion in a closed-loop system does not achieve perfect identification of the LEM path, residual echo and distortion will be present in the error signal. To avoid this, the cost function must be modified, defining it as

$$J_f = E[|e_f(n)|^2] = E[|e(n) * p(n)|^2], \quad (18)$$

where $p(n)$ is the impulse response of a linear filter applied to the error signal. This filter must be chosen to force the identification of the LEM path to be the solution of the proposed minimization problem. Differentiating the cost function we can find the optimal filter as the solution to

$$E[e_f(n) \sum_{j=0}^{\infty} \tilde{h}'(j) \frac{\partial e_f(n - \Delta - j)}{\partial \tilde{h}'(k)}] - E[e_f(n) e_f(n - \Delta - k)] = 0, \quad k = 0, 1, 2, \dots, \quad (19)$$

According to (19), in order to allow $\tilde{h}'(n) = 0$ be a solution of the minimization problem, the filter $p(n)$ should be designed to achieve a correlation function of $e_f(n)$ equal to zero for a time lag greater than Δ .

4.2. Minimum Mean Squared Filtered Error Condition in the Frequency Domain

The criterion in (18) can be expressed in the frequency domain as

$$\hat{h}'_{opt}(n) = \arg \min_{\hat{h}'(n)} \frac{1}{2\pi} \int_0^{2\pi} S_{e_f}(e^{j\omega}) d\omega, \quad (20)$$

where $S_{e_f}(e^{j\omega})$ is the PSD of the filtered error, which according to Fig.1 and the definition of $e_f(n)$ is

$$S_{e_f}(e^{j\omega}) = \left| \frac{P(e^{j\omega})A(e^{j\omega})}{1 - K\hat{H}'(e^{j\omega})e^{-j\omega\Delta}} \right|^2 \sigma_w^2. \quad (21)$$

Using the definitions (8), (9) and (10), the value of the transfer function for a causal finite-duration impulse response filter that minimizes the mean squared filtered error is

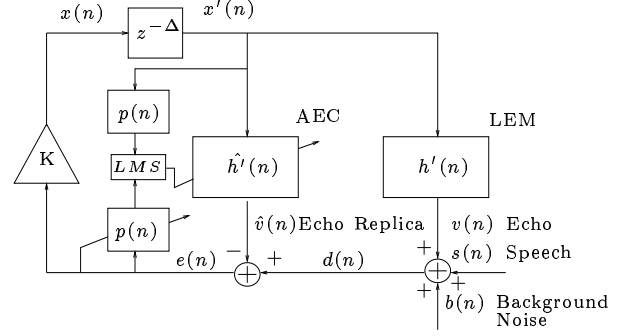


Figure 2: Diagram of a one-channel speech reinforcement system with adaptive linear prediction in the error path

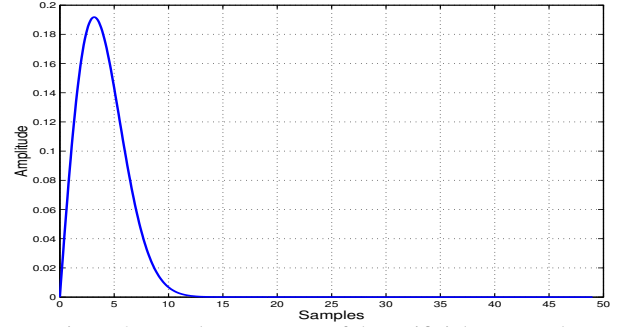


Figure 3: Impulse Response of the artificial LEM path

$$\hat{H}'(e^{j\omega}) = H'_1(e^{j\omega}) - \frac{1 - P(e^{j\omega})A(e^{j\omega})}{K} \quad (22)$$

where $P(e^{j\omega})$ must satisfy

$$P(e^{j\omega}) = A^{-1}(e^{j\omega}) \quad (23)$$

to allow $\tilde{H}'(e^{j\omega}) = 0$ be the solution of (20), with $A^{-1}(e^{j\omega})$ the inverse filter of $A(e^{j\omega})$.

4.3. Adaptive Linear Prediction and Echo Cancellation

As described before, the use of a whitening filter $p(n)$ is necessary to ensure that the minimization of the power of the filtered error leads to the identification of the LEM path.

In order to find a practical solution to the problem of acoustic echo cancellation in a speech reinforcement system for vehicles, adaptive linear prediction must be used since neither the error signal nor the speech signal are time-invariant.

The adaptive minimization of the mean squared filtered error leads to the well-known FX-LMS algorithm [5] if the instantaneous gradient is used as an approximation to the exact one. The resulting system is presented in Fig. 2.

5. Simulation Results

Several simulations have been carried out to illustrate the dependency on the identification error of the gain factor K , the loop delay Δ and the characteristics of the excitation signal, with a simple artificial LEM path, shown in Fig. 3. The excitation signal consists of a first order autoregressive stationary process with power spectral density

$$S(e^{j\omega}) = \frac{1}{(1 - ae^{-j\omega})(1 - ae^{j\omega})}. \quad (24)$$

The estimation error can be measured by using the normalized l_2 norm of the weight misadjustment vector defined as

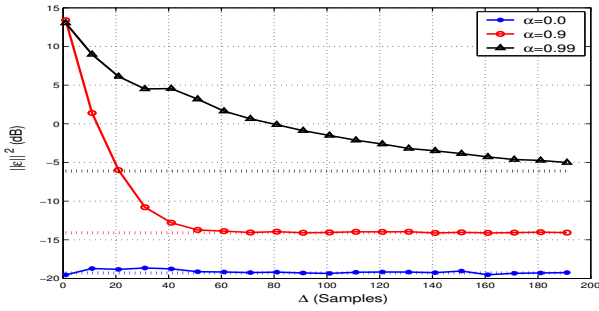


Figure 4: $\|\epsilon\|^2$ evolution with the loop delay Δ for different values of the parameter α .

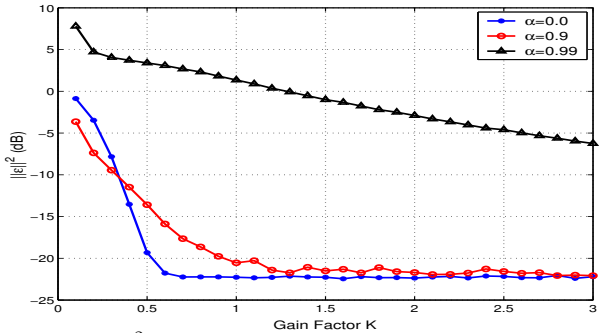


Figure 5: $\|\epsilon\|^2$ evolution with the Gain Factor K for different values of the parameter α .

$$\|\epsilon\|^2 = \frac{\sum_{k=0}^L |h'_k - \hat{h}'_k|^2}{\sum_{k=0}^L |h'_k|^2}. \quad (25)$$

Fig. 4 shows the evolution of $\|\epsilon\|^2$ with Δ for different values of the parameter α that controls the bandwidth of the excitation signal in a system with a gain factor $K = 0.5$. The value of $\|\epsilon\|^2$ that would be obtained, under the same circumstances, in an open loop system is presented in dotted line.

It can be seen that, the identification mismatch does not depend on Δ for a white noise excitation ($\alpha = 0$) as predicted (17), and for higher values of α , $\|\epsilon\|^2$ decreases as Δ increases. This evolution agrees with the theoretical result of section (3.2.2). Equation (17) states that the identification error depends only on the values of the input signal model from Δ to $N + \Delta$ and the autocorrelation function of the first order AR process decreases exponentially as the time lag increases.

The evolution of $\|\epsilon\|^2$ with K is depicted in Fig. 5 in a system with a delay loop Δ of 50 samples. According to (17), the identification error decreases as K increases for every value of α . For a white noise input, where the identification bias is always zero, the decrease of $\|\epsilon\|^2$ for values of $K < 0.6$ is due to the increase in the echo to excitation signal ratio that improves the identification accuracy.

Simulations with the proposed system were performed to compare the estimates of the LEM path impulse response, obtained with and without the adaptive linear prediction, using a real LEM path and speech as input signal.

A 600 coefficient LEM path impulse response measured in a medium size car was used. The length of the adaptive predictor was 8 coefficients and the length of the echo canceller adaptive filter was 350 coefficients with $\Delta = 50$ samples. Sev-

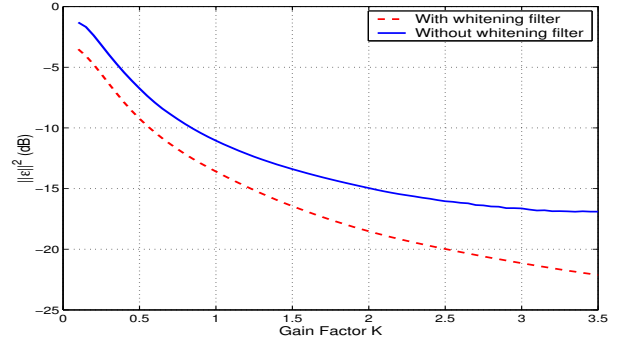


Figure 6: $\|\epsilon\|^2$ versus the gain factor K . With (dashed red) and without (solid blue) the adaptive linear predictor.

eral sentences were used with a sampling rate of 8 kHz. Real car noise, recorded while driving in a highway, was added to the noise free speech signals. The input SNR was around 10 dB.

Fig. 6 shows the evolution of the mean l_2 norm with (dashed red line) and without (solid blue line) the adaptive linear prediction. It can be seen that the error in the estimation process decreases while increasing K , as predicted in (22), and the estimate obtained using the whitening filter is always smaller than the one obtained without using the linear prediction filter.

6. Conclusions

In this paper, the application of the minimum mean square error criterion for the feedback cancellation problem in a speech reinforcement system for cars has been studied. The solution obtained for a causal FIR filter has been derived and compared to the solution for an unconstrained identification filter. The use of adaptive whitening filters has been proposed and justified for reducing the error in the identification of the LEM path. This identification mismatch can be reduced by modifying the cost function. This modification consists of the filtering of the error signal before minimizing its variance using a whitening filter. The use of the modified cost function leads to the use of the FX-LMS algorithm in its adaptive implementation. Simulation results confirm the theoretical study presented here. The comparison between the weight misadjustment obtained with and without adaptive linear prediction shows that the minimization of the mean squared whitened error produces more accurate estimates of the LEM path than the use of the classical LMS algorithm for speech reinforcement systems for vehicles.

7. References

- [1] Alfonso Ortega, Eduardo Lleida, and Enrique Masgrau, "Speech reinforcement system for car cabin communications," Accepted for publication in IEEE Transactions on Speech and Audio Processing, 2005.
- [2] Simon Haykin, *Adaptive Filter Theory*, Prentice-Hall International, 1996.
- [3] Alfonso Ortega, *Speech Reinforcement System for Vehicles. (In Spanish)*, Ph.D. thesis, University of Zaragoza, 2005.
- [4] Johan Hellgren and Urban Frossell, "Bias of feedback cancellation algorithms in hearing aids based on direct closed loop identification," *IEEE Trans. Speech Audio Processing*, vol. 9, no. 8, pp. 906–913, November 2001.
- [5] Sen M. Kuo and Dennis R. Morgan, *Active Noise Control Systems, Algorithms and DSP Implementations*, Wiley-Interscience, John Wiley & Sons, Inc., 1996.