

Chapter 15

ACOUSTIC ECHO REDUCTION IN A TWO-CHANNEL SPEECH REINFORCEMENT SYSTEM FOR VEHICLES

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

Communication Technologies Group (GTC). University of Zaragoza, Spain.

Abstract: This chapter presents a two-channel speech reinforcement system which has the goal of improving speech intelligibility inside cars. As microphones pick up not only the voice of the speaker but also the reinforced speech coming from the loudspeakers, feedback paths appear. These feedback paths can make the system become unstable and acoustic echo cancellation is needed in order to avoid it. In a two-channel system, two system identifications must be performed for each channel, one of them is an open-loop identification and the other one is closed-loop. Several methods have been proposed for echo suppression in open-loop systems like hands-free systems. We propose here the use of echo suppression filters specially designed for closed-loop subsystems along with echo suppression filters for open-loop subsystems based on the optimal filtering theory. The derivation of the optimal echo suppression filter needed in the closed-loop subsystem is also presented.

Key words: Speech Reinforcement; echo cancellation; acoustic feedback reduction.

1. INTRODUCTION

A speech reinforcement system can be used in medium and large size motor vehicles to improve the communications among passengers [1],[2]. Inside a car, speech intelligibility can be degraded due to the lack of visual contact between speaker and listener, the noise and the use of sound absorbing materials among other factors. Using a set of microphones placed on the ceiling of the car, this system picks up the speech of each passenger. After that, it is amplified and played back into the cabin using the loudspeakers of the audio system of the car.

Acoustic echo appears because the signal radiated by the loudspeakers is picked up again by the microphones. Due to the amplification stage between the microphones and the loudspeakers, the system can become unstable.

Along with the speech signal, the noise is also picked up by the microphones and amplified by the system increasing the overall noise level present inside the car. To prevent this, a noise reduction stage must be used.

Echo Cancellers (AEC) are widely used to overcome electro-acoustic coupling between loudspeakers and microphones [3]. In a two-channel system, each channel must have two echo cancellers, one corresponding to an open-loop subsystem and the other one corresponding to a closed-loop subsystem. Nevertheless, to achieve enough echo attenuation the use of Echo Suppression Filters (ESF) is needed. Several techniques have been proposed for further echo attenuation using residual echo reduction filters [4],[5]. These techniques can be used for open-loop systems but in a speech reinforcement system for vehicles, the ESF must also ensure stability in the closed-loop subsystems. The study for a one-channel system can be found in [6] and the optimal ESF transfer function for the closed-loop subsystems in a two-channel speech reinforcement system is derived here.

Another important aspect of this system is that the overall delay must be short enough to achieve full integration of the sound coming from the direct path and the reinforced speech coming from the loudspeakers.

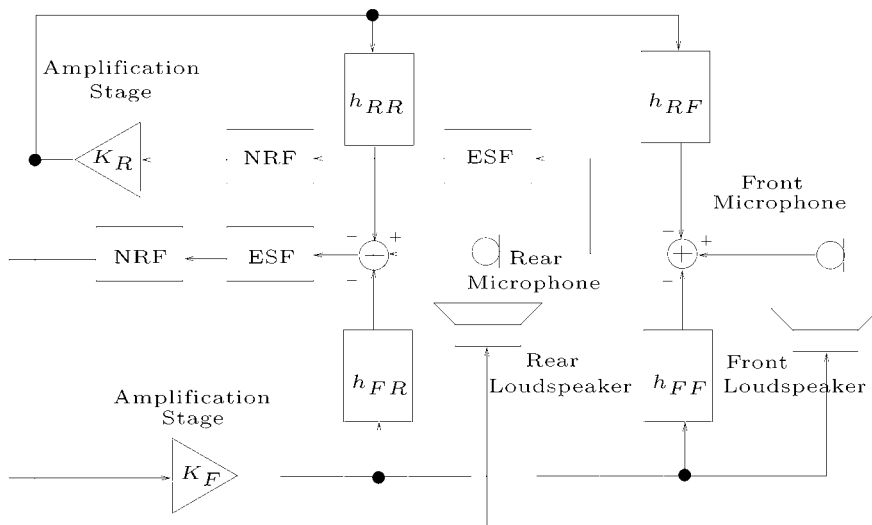


Figure 15-1. Schematic diagram of a two-channel speech reinforcement system for cars.

2. DESCRIPTION AND STABILITY STUDY OF THE TWO-CHANNEL SYSTEM

In order to make communications inside a car more comfortable, a two-channel speech reinforcement system is required. One channel must take the speech of the rear passengers to the front part of the car and the other one must take the speech of the front passengers to the rear seats. A block diagram of the two-channel system is presented in Fig. 15-1.

In a two-channel speech reinforcement system, for each channel, there must be two echo cancellers, an echo suppression filter, a Noise Reduction Filter (NRF) and an amplification stage.

The estimation of each Loudspeaker-Enclosure-Microphone (LEM) path performed by each adaptive filter is not enough to ensure the stability of the system. The inaccuracy of the estimation can make the system become unstable. The transfer function of the channel from microphone X (rear, R, or front, F) to loudspeaker Y (F or R) is

$$P_{XY}(e^{j\omega}) = \frac{K_Y W_Y(e^{j\omega}) [1 - K_X W_X(e^{j\omega}) \tilde{H}_{XY}(e^{j\omega})]}{D(e^{j\omega})} \quad (15.1)$$

where

$$\begin{aligned} D(e^{j\omega}) = & 1 - K_F W_F(e^{j\omega}) \tilde{H}_{FR}(e^{j\omega}) - K_R W_R(e^{j\omega}) \tilde{H}_{RF}(e^{j\omega}) \\ & - K_R K_F W_R(e^{j\omega}) W_F(e^{j\omega}) \tilde{H}_{FF}(e^{j\omega}) \tilde{H}_{RR}(e^{j\omega}) \\ & + K_R K_F W_R(e^{j\omega}) W_F(e^{j\omega}) \tilde{H}_{RF}(e^{j\omega}) \tilde{H}_{FR}(e^{j\omega}) \end{aligned} \quad (15.2)$$

and $\tilde{H}_{XY}(e^{j\omega})$ is the difference between the LEM path transfer function $H_{FR}(e^{j\omega})$ and its corresponding adaptive filter transfer function $\hat{H}_{XY}(e^{j\omega})$. $W_R(e^{j\omega})$ is the transfer function of the system composed of the ESF and the NRF for the front-rear channel and $W_F(e^{j\omega})$ for the rear-front channel. K_F and K_R are the gain factors for the rear-front channel and the front-rear respectively.

The optimal transfer function each channel is

$$P_{XY}(e^{j\omega}) = K_Y W_{Yn}(e^{j\omega}) \quad (15.3)$$

where $W_{Yn}(e^{j\omega})$ is the transfer function of the noise reduction filter of the corresponding channel.

Substituting (15.3) into (15.1), and considering (15.2), the optimal echo suppression filter follows

$$W_{Xe}(e^{j\omega}) = \frac{W_{Ye}(e^{j\omega})}{D_{Xe}(e^{j\omega})} \quad (15.4)$$

with

$$\begin{aligned} D_{Xe}(e^{j\omega}) = & 1 - K_Y W_{Yn}(e^{j\omega}) W_{Ye}(e^{j\omega}) \tilde{H}_{YX}(e^{j\omega}) \\ & + K_X W_{Xn}(e^{j\omega}) W_{Ye}(e^{j\omega}) \tilde{H}_{XY}(e^{j\omega}) \end{aligned} \quad (15.5)$$

The optimal expressions for both echo suppression filters are not independent and must be fulfilled simultaneously to ensure unconditional stability. This is only possible if

$$K_R \tilde{H}_{RF}(e^{j\omega}) = K_F \tilde{H}_{FR}(e^{j\omega}), \quad (15.6)$$

for each frequency, which implies that both filters must be equal to each other

$$W_{Re}(e^{j\omega}) = W_{Fe}(e^{j\omega}). \quad (15.7)$$

Condition (15.6), is not under the control of the designer, so it will not be always met.

3. ECHO SUPPRESSION FILTERS FOR THE CLOSED-LOOP SUBSYSTEMS AND THE OPEN-LOOP SUBSYSTEMS

One possible solution to increase the stability of the two-channel speech reinforcement system is to distinguish between open-loop subsystems and closed-loop subsystems applying specific treatment approaches to each one of them.

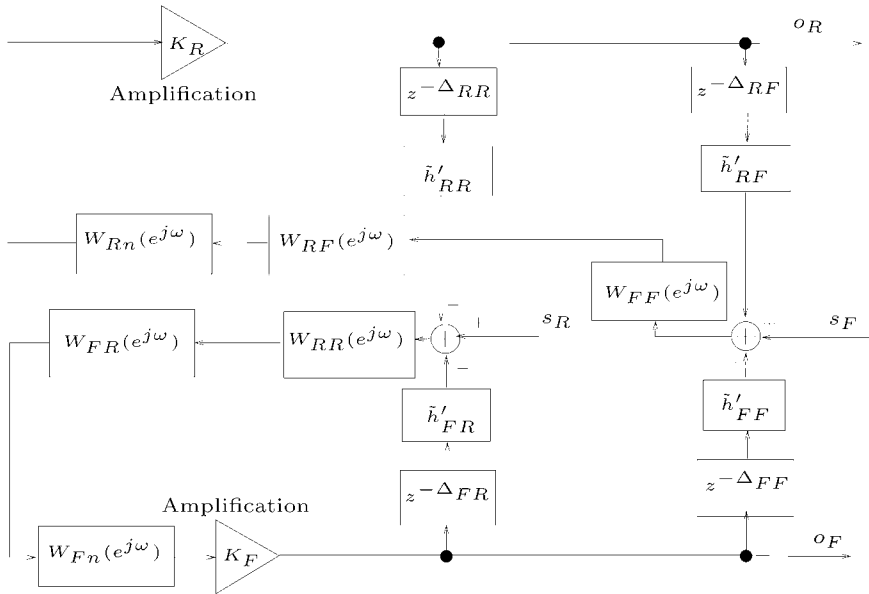


Figure 15-2. Two-channel speech reinforcement system with differentiated treatment techniques for closed-loop subsystems and for open-loop subsystems.

To cope with the residual echo remaining after the echo canceller for the open-loop subsystems, several approaches have been proposed in the literature [4],[5],[7]. The use of the filters $W_{FF}(e^{j\omega})$ and $W_{RR}(e^{j\omega})$, that follow a Wiener based approach, is proposed.

In order to increase the stability margin of the speech reinforcement system, we propose here the use of the echo suppression filters $W_{RF}(e^{j\omega})$ and $W_{FR}(e^{j\omega})$, specially designed for the closed-loop subsystems.

The proposed system is presented in Fig. 15-2. where s_R and s_F are the input signals for the rear-front channel and the front-rear channel respectively, o_R is the output signal of the front-rear channel and o_F is the output signal of the rear-front channel. Due to the propagation delay, the LEM path of each loudspeaker-microphone pair is modelled as a delay block of Δ_{XY} samples followed by a linear system with the same impulse response of the LEM path except for the first Δ_{XY} values. The first Δ_{XY} coefficients of its corresponding adaptive filter are also set to zero to compensate for the propagation delay.

According to Fig.15-2, the transfer functions of the system from microphone X to loudspeaker Y follow

$$P_{XY}(e^{j\omega}) = \frac{L_X(e^{j\omega})[1 - L_Y(e^{j\omega})W_{XXe}(e^{j\omega})\tilde{H}_{XY}(e^{j\omega})]}{D_2(e^{j\omega})} \quad (15.8)$$

with

$$L_X(e^{j\omega}) = K_X W_{XYe}(e^{j\omega}) W_{Xn}(e^{j\omega}) W_{YYe}(e^{j\omega}) \quad (15.9)$$

$$\begin{aligned} D_2(e^{j\omega}) = & 1 - L_F(e^{j\omega})\tilde{H}_{FR}(e^{j\omega}) - L_R(e^{j\omega})\tilde{H}_{RF}(e^{j\omega}) \\ & - L_F(e^{j\omega})L_R(e^{j\omega})\tilde{H}_{RR}(e^{j\omega})\tilde{H}_{FF}(e^{j\omega}) \\ & + L_F(e^{j\omega})L_R(e^{j\omega})\tilde{H}_{RF}(e^{j\omega})\tilde{H}_{FR}(e^{j\omega}) \end{aligned} \quad (15.10)$$

where

$$W_{XYe}(e^{j\omega}) = \frac{1}{1 + K_X W_{YYe}(e^{j\omega}) W_{Xn}(e^{j\omega}) \tilde{H}_{XY}(e^{j\omega})} \quad (15.11)$$

is the transfer function of the proposed ESF for the closed-loop subsystems.

Substituting (15.11) into (15.8), the system transfer functions satisfy

$$P_{XY}(e^{j\omega}) = \frac{K_X W_{Xn}(e^{j\omega}) W_{XXe}(e^{j\omega})}{D_3(e^{j\omega})} \quad (15.12)$$

where

$$\begin{aligned} D_3(e^{j\omega}) = & 1 - K_R W_{RRe}(e^{j\omega}) W_{Rn}(e^{j\omega}) \tilde{H}_{FF}(e^{j\omega}) \\ & \times K_F W_{FFe}(e^{j\omega}) W_{Fn}(e^{j\omega}) \tilde{H}_{RR}(e^{j\omega}) \end{aligned} \quad (15.13)$$

Thus, the stability of the reinforcement system, assuming that the echo suppression filters are working properly, depends only on the open-loop subsystems. That is, the stability depends on the misadjustment functions $\tilde{H}_{RR}(e^{j\omega})$ and $\tilde{H}_{FF}(e^{j\omega})$ that is intended to be minimized by the filters $W_{RRe}(e^{j\omega})$ and $W_{FFe}(e^{j\omega})$ respectively.

The echo suppression filters for the closed-loop subsystems, that increase the stability of the two-channel reinforcement system, depend on the misadjustment functions of the closed-loop subsystems that are a priori unknown. Assuming that the ESF for the open-loop subsystems are real valued functions, as well as the NRF for each channel, it can be shown that using the magnitude of the misadjustment function is the best option to increase the stability of the system [1].

The estimates of the magnitude of the misadjustment function for each closed-loop subsystem are obtained using estimates of the residual echo $r_{FF}(n)$ for the rear-front channel and estimates of the residual echo $r_{RR}(n)$ for the front-rear channel.

For the front-rear channel, the residual echo remaining after the closed-loop subsystem acoustic echo canceller, can be expressed as

$$r_{RF}(n) = o_R(n) * w_{FFe}(n) * \tilde{h}_{RF}(n) \quad (15.14)$$

where $w_{FFe}(n)$ is the impulse response of the ESF for the open-loop subsystem of the front-rear channel and $\tilde{h}_{RF}(n)$ is the inverse Fourier transform of the misadjustment function. Thus, the PSD of the residual echo can be expressed as

$$S_{r_{RF}}(e^{j\omega}) = S_{o_R}(e^{j\omega}) \cdot |W_{FFe}(e^{j\omega}) \tilde{H}_{RF}(e^{j\omega})|^2 \quad (15.15)$$

which depends on the PSD of the output signal that will be played back through the rear loudspeakers of the reinforcement system, $S_{o_R}(e^{j\omega})$, and on the squared magnitude of the misadjustment function, $|\tilde{H}_{RF}(e^{j\omega})|^2$, along with the squared magnitude of the ESF of the open-loop subsystem of the front-rear channel, $|W_{FFe}(e^{j\omega})|^2$.

According to (15.15), we can express the squared magnitude of the product of the open-loop subsystem ESF of the front-rear channel and the misadjustment function as

$$|W_{FFe}(e^{j\omega}) \tilde{H}_{RF}(e^{j\omega})|^2 = \frac{S_{r_{RF}}(e^{j\omega})}{S_{o_R}(e^{j\omega})} \quad (15.16)$$

The PSD of the rear output signal, according to Fig. 15-2, can be expressed as

$$S_{o_r}(e^{j\omega}) = S_{e_r}(e^{j\omega}) \cdot K_R^2 \cdot |W_{RFe}(e^{j\omega})W_{Rn}(e^{j\omega})|^2 \quad (15.17)$$

and thus, combining (15.16) and (15.17) and substituting into (15.11), we can obtain the expression for the closed-loop ESF of the front-rear channel that responds to

$$W_{RF}(e^{j\omega}) = 1 - \sqrt{\frac{S_{r_{RF}}(e^{j\omega})}{S_{e_r}(e^{j\omega})}} \quad (15.18)$$

which depends on the PSD of the residual echo remaining after the closed-loop subsystem of the front-rear channel, $S_{r_{RF}}(e^{j\omega})$, and on the PSD of the error signal of the front-rear channel, $S_{e_r}(e^{j\omega})$.

In the same way, we can obtain the expression for the ESF for the closed-loop subsystem of the rear-front channel that must follow

$$W_{FR}(e^{j\omega}) = 1 - \sqrt{\frac{S_{r_{FR}}(e^{j\omega})}{S_{e_f}(e^{j\omega})}} \quad (15.19)$$

4. PERFORMANCE MEASURES

In this section, a performance evaluation of the residual echo suppression filters for the closed-loop subsystems is presented. For the evaluation, we used four different impulse responses corresponding to four different real electro-acoustic paths measured in a medium-size car with 600 coefficients each, using a sampling rate of 8 kHz.

The misadjustment between the impulse response of the electro-acoustic path and the impulse response of the corresponding adaptive filter is controlled by adding a random noise to each one of the coefficients of the original impulse response. This estimation error can be measured by using the normalized l_2 norm of the weight misadjustment vector defined as

$$\|\mathcal{E}\|^2 = \frac{\sum_{k=0}^L |h'_k - \hat{h}'_k|^2}{\sum_{k=0}^L |h'_k|^2} \quad (15.20)$$

where h'_k is the k th coefficient of the impulse response of the real electro-acoustic path and \hat{h}'_k is the k th coefficient of its corresponding adaptive filter.

Several noise free speech recordings were used as passenger's speech adding real car noise, recorded while driving on a highway, as background noise resulting in a SNR around 20 dB. The length of each signal frame was 16 ms and to reduce the overall delay of the system, a time overlap of 75% was used.

In order to measure the benefit of using the ESF for the closed-loop subsystems, the isolation between channels is used. That is defined as the ratio between the power of the front-rear channel output and the power of the rear-front channel output when only the front passenger is talking.

$$I_{RF} = \frac{E \left[|o_R(n)|^2 \right]}{E \left[|o_F(n)|^2 \right]} \quad (15.21)$$

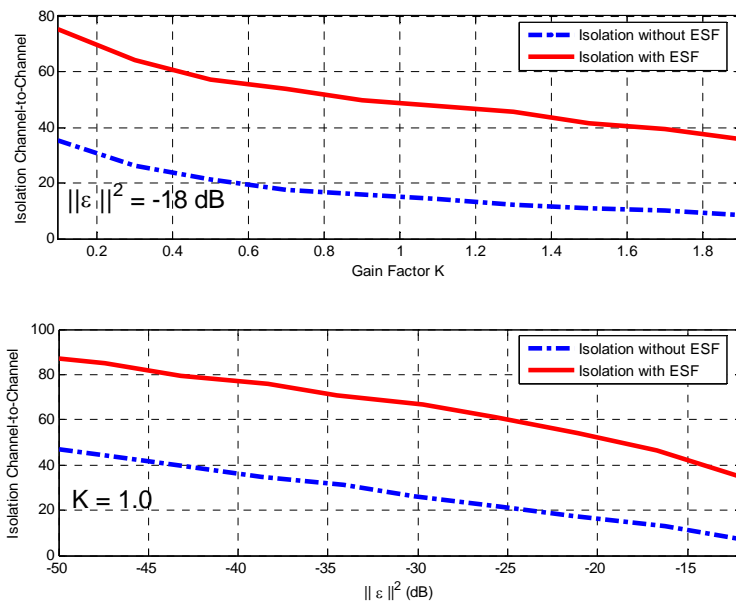


Figure 15-3. Isolation between channels with and without ESF in the closed-loop subsystems.

In the upper half of Fig. 15-3, the evolution of the isolation between channels with the gain factor K for $\|\varepsilon\|^2 = -18$ dB is presented. It can be seen that the increase is around 40 dB for almost every value of K . The evolution of the isolation between channels with the normalized l_2 norm of the weight misadjustment vector is plotted below for $K = 1.0$. The isolation increase ranges from 30 dB for high values of misadjustment (around -12 dB) to 40 dB for lower values of $\|\varepsilon\|^2$.

To show that there is no degradation in terms of system gain decrease or distortion increase, the evolution of the system gain with K for $\|\varepsilon\|^2 = -18$ dB, and the evolution of the system gain with $\|\varepsilon\|^2$ for $K = 1.0$ is presented in Fig. 15-4. In Fig. 15-5, the evolution with K for $\|\varepsilon\|^2 = -18$ dB and with $\|\varepsilon\|^2$ for $K = 1.0$ of the Itakura-Saito distance between the input signal and the corresponding output signal is presented.

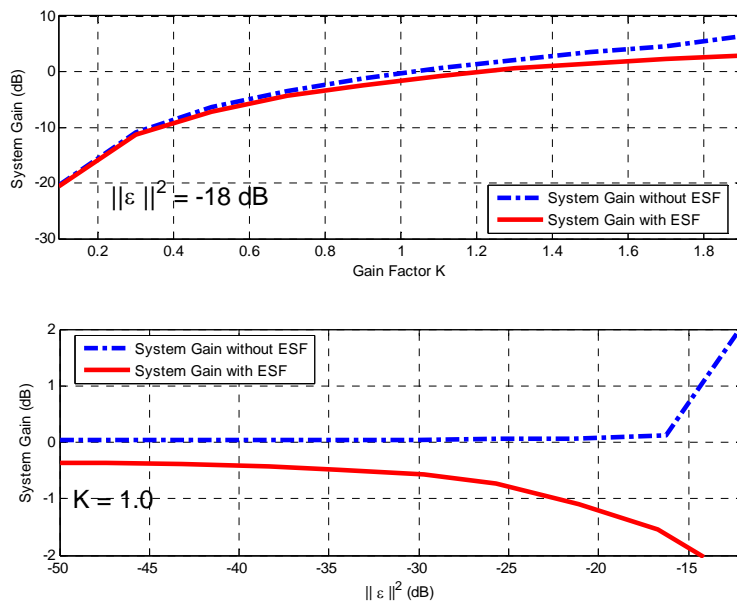


Figure 15-4. System gain with and without ESF in the closed-loop subsystems.

In the lower half of Fig. 15-4 it can be seen that the System Gain increases dramatically for values of $\|\varepsilon\|^2$ above -15 dB. The same effect can be observed in both parts of Fig. 15-5 regarding the distortion for high values of K or $\|\varepsilon\|^2$. This is due to the appearance of howling as the system

is very close to instability and strong tonal components are present in the output signal.

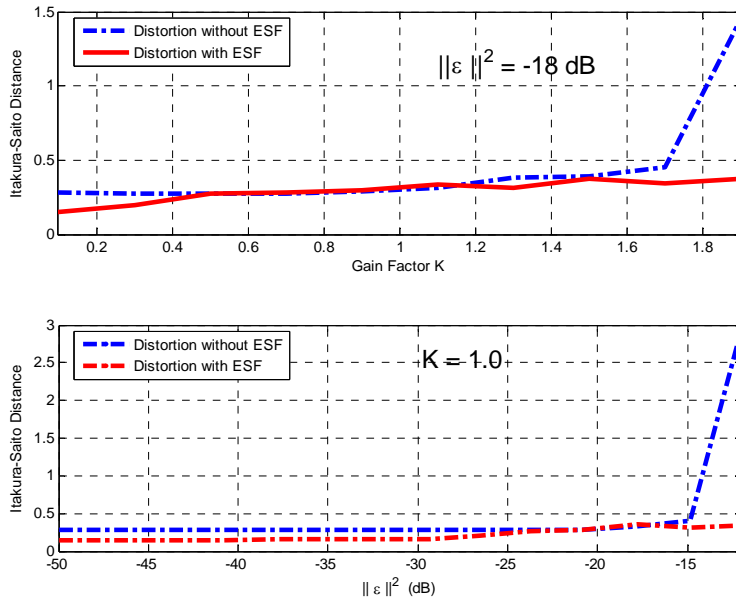


Figure 15-5. Itakura-Saito Distance between the input signal and the reinforced speech with and without ESF in the closed-loop subsystems.

5. CONCLUSIONS

A two-channel speech reinforcement system is required in order to make communications inside a car more comfortable. In a two-channel system, two subsystems can be distinguished for each channel, an open-loop and a closed-loop subsystem. The use of specific treatment for residual echo attenuation in the closed-loop subsystems has been presented, and the optimal expression for the transfer function of the Echo Suppression filter that ensures unconditional stability has been derived. Optimal echo suppression filters do not always exist and the existence of the optimal filters depends on the misadjustment function between the electro-acoustic path impulse response and the adaptive filter of the acoustic echo canceller which is not under the control of the designer. An alternative solution has been

proposed and evaluated. This solution is based on an estimation of the residual echo power spectral density. The performance evaluations show that there is an increase of around 40 dB in the isolation between channels when using the proposed Echo Suppression Filters, without decreasing the gain of the system or increasing the speech distortion.

6. REFERENCES

- [1] A. Ortega, E. Lleida, E. Masgrau, "Speech reinforcement system for car cabin communications" .IEEE Transactions on Speech and Audio Processing. vol. 13, no. 5. pp. 917-929. September 2005
- [2] F. Gallego, A. Ortega, E. Lleida, E. Masgrau "Method and system for suppressing echoes and noise in environments under variable acoustic and highly feedback conditions" patent WO 02/101728 A1.
- [3] C. Breining, P. Dreiseitel, E. Hänslér, 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 Mag., vol. 16, no. 4, pp. 42-69, July 1999.
- [4] S. Gustafsson, R. Martin and P. Vary, "Combined acoustic echo control and noise reduction for hands-free telephony", Signal Processing, no. 64, pp. 21-32, January 1998.
- [5] E. Hänslér and G. U. Schmidt, "Hands-free telephones - joint control of echo cancellation and postfiltering", Signal Processing, no. 80, pp. 2295-2305. January 2000.
- [6] A. Ortega, E. Lleida and E. Masgrau, "Residual echo power estimation for speech reinforcement systems in vehicles" in Proceedings of Eurospeech. September 2003.
- [7] G. Enzner, R. Martin and P. Vary, "Unbiased residual echo power estimation for hands-free telephony", in Proceedings of ICASSP, vol. 2, pp- 1893-1896. May 2002