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
{ortega,lleida,masgrau,lbuera,amiguel}@unizar.es

Abstract

A two-channel speech reinforcement system which has the goal of improving speech intelligibility inside cars is presented in this work. As microphones pick up not only the voice of the speaker but also the reinforced speech coming from the loudspeakers, feedback paths appear in a speech reinforcement system for vehicles. This 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 spectral estimation method for the power spectral density of the residual echo suppression filters is presented along with the derivation of the optimal echo suppression filter needed in the closed-loop subsystem. Results about the performance of the proposed system are also provided.

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.

According to Fig. 1, we can easily identify, on the one hand, two closed-loop subsystems, that are the electroacoustic path composed of the rear microphones, the reinforcement system, the front loudspeakers and the cabin, for the rear-front channel, and the path composed of the front microphones, the reinforcement system, the rear loudspeakers and the cabin for the front-rear channel. On the other hand, there are two open-loop

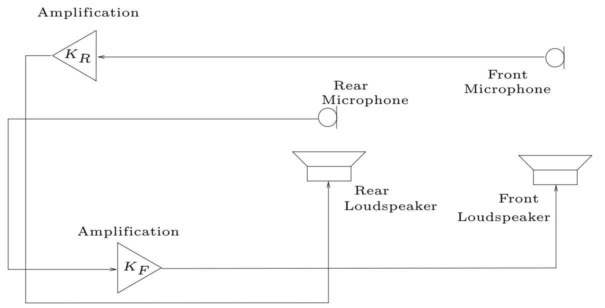

Figure 1: *Simplified Schematic diagram of a two-channel speech reinforcement system for cars.*

subsystems, the one composed of the reinforcement system, the front loudspeakers, the cabin and the front microphones, for the rear-rear channel and the one composed of the reinforcement system, the rear loudspeakers, the cabin and the rear microphones, for the front-front channel.

Acoustic 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 presented here. 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]. In this paper, the optimal ESF transfer function for the closed-loop subsystems in a two-channel speech reinforcement system is derived along with the residual echo Power Spectral Density (PSD) estimation method.

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.

This paper is organized as follows. A brief description and a stability study of the system will be presented in Section 2, along with the optimal expressions for the Echo Suppression Filters in the two-channel system. In Section 3, the proposed Echo Suppression Filters will be presented and in Section 4 the residual echo PSD estimation method for the closed-loop subsystems will be explained. In Section 5, performance measures and results will be shown and in Section 6, we present the conclussion along with a summary of the paper.

*This work was funded by the grant TIC2002-04103-C03-01 from the Spanish MEC

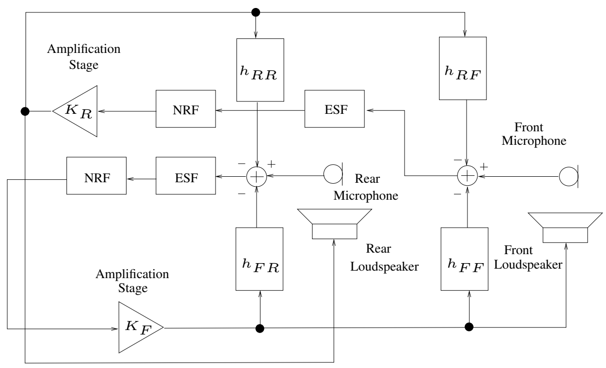

Figure 2: *Schematic diagram of a two-channel speech reinforcement system for cars.*

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. 2.

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 the echo cancellers adaptive filters 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 rear-front and front-rear channels in the two-channel system are

$$P_{RF}(e^{j\omega}) = \frac{K_F W_F(e^{j\omega})\left[1 - K_R W_R(e^{j\omega})\tilde{H}_{RF}(e^{j\omega})\right]}{D(e^{j\omega})} \tag{1}$$

$$P_{FR}(e^{j\omega}) = \frac{K_R W_R(e^{j\omega})\left[1 - K_F W_F(e^{j\omega})\tilde{H}_{FR}(e^{j\omega})\right]}{D(e^{j\omega})}, \tag{2}$$

where

$$\begin{aligned} D(e^{j\omega}) = 1 &- K_F W_F(e^{j\omega})\tilde{H}_{FR}(e^{j\omega}) \\ &- K_R K_F W_R(e^{j\omega}) W_F(e^{j\omega})\tilde{H}_{FF}(e^{j\omega})\tilde{H}_{RR}(e^{j\omega}) \\ &- K_R W_R(e^{j\omega})\tilde{H}_{RF}(e^{j\omega}) \\ &+ K_R K_F W_R(e^{j\omega}) W_F(e^{j\omega})\tilde{H}_{RF}(e^{j\omega})\tilde{H}_{FR}(e^{j\omega}) \end{aligned} \tag{3}$$

and $\tilde{H}_{XY}(e^{j\omega})$ is the difference between the LEM path transfer function $H_{XY}(e^{j\omega})$, from loudspeaker X to microphone Y, 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 two-channel speech reinforcement system is stable if and only if

$$\begin{aligned} |K_F W_F(e^{j\omega})\tilde{H}_{FR}(e^{j\omega}) &+ K_R W_R(e^{j\omega})\tilde{H}_{RF}(e^{j\omega}) \\ - K_R K_F W_R(e^{j\omega}) & W_F(e^{j\omega})\tilde{H}_{RF}(e^{j\omega})\tilde{H}_{FR}(e^{j\omega}) \\ + K_R K_F W_R(e^{j\omega}) & W_F(e^{j\omega})\tilde{H}_{FF}(e^{j\omega})\tilde{H}_{RR}(e^{j\omega})| < 1, \end{aligned} \tag{4}$$

at all frequencies with positive feedback, i.e.,

$$\begin{aligned} \angle\Big(& K_R K_F W_R(e^{j\omega}) W_F(e^{j\omega})\tilde{H}_{FF}(e^{j\omega})\tilde{H}_{RR}(e^{j\omega}) \\ & - K_R K_F W_R(e^{j\omega}) W_F(e^{j\omega})\tilde{H}_{RF}(e^{j\omega})\tilde{H}_{FR}(e^{j\omega}) \\ & + K_F W_F(e^{j\omega})\tilde{H}_{FR}(e^{j\omega}) \\ & + K_R W_R(e^{j\omega})\tilde{H}_{RF}(e^{j\omega})\Big) = m2\pi, \end{aligned} \tag{5}$$

with $m \in \mathbb{Z}$.

The optimal transfer functions for the two-channel speech reinforcement system are

$$P_{RF}(e^{j\omega}) = K_F W_{Fn}(e^{j\omega}) \tag{6}$$

$$P_{FR}(e^{j\omega}) = K_R W_{Rn}(e^{j\omega}), \tag{7}$$

where $W_{Fn}(e^{j\omega})$ and $W_{Rn}(e^{j\omega})$ are the transfer functions of the noise reduction filter of the rear-front channel and the front-rear channel respectively.

Substituting (6) and (7) into (1) and (2), and considering (3), the optimal echo suppression filters for each channel are

$$W_{Re}(e^{j\omega}) = \frac{W_{Fe}(e^{j\omega})}{D_{Re}(e^{j\omega})} \tag{8}$$

$$W_{Fe}(e^{j\omega}) = \frac{W_{Re}(e^{j\omega})}{D_{Fe}(e^{j\omega})}, \tag{9}$$

with

$$\begin{aligned} D_{Re}(e^{j\omega}) &= 1 - K_F W_{Fn}(e^{j\omega}) W_{Fe}(e^{j\omega})\tilde{H}_{FR}(e^{j\omega}) \\ &\quad + K_R W_{Fe}(e^{j\omega}) W_{Rn}(e^{j\omega})\tilde{H}_{RF}(e^{j\omega}) \end{aligned} \tag{10}$$

$$\begin{aligned} D_{Fe}(e^{j\omega}) &= 1 - K_R W_{Rn}(e^{j\omega}) W_{Re}(e^{j\omega})\tilde{H}_{RF}(e^{j\omega}) \\ &\quad + K_F W_{Re}(e^{j\omega}) W_{Fn}(e^{j\omega})\tilde{H}_{FR}(e^{j\omega}). \end{aligned} \tag{11}$$

Equations (8) and (9) are not independent and must be fulfilled simultaneously to ensure unconditional stability. This is only possible if

$$K_R \tilde{H}_{RF}(e^{j\omega}) = K_F \tilde{H}_{FR}(e^{j\omega}), \tag{12}$$

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

$$W_{Re}(e^{j\omega}) = W_{Fe}(e^{j\omega}). \tag{13}$$

Condition (12), 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.

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.

$$P_{FR}(e^{j\omega}) = \frac{K_R W_{RFe}(e^{j\omega}) W_{Rn}(e^{j\omega}) W_{FFe}(e^{j\omega}) \left[1 - K_F W_{Fn}(e^{j\omega}) W_{FFe}(e^{j\omega}) W_{FRe}(e^{j\omega}) W_{RRe}(e^{j\omega}) \tilde{H}_{FR}(e^{j\omega})\right]}{D_2(e^{j\omega})} \tag{14}$$

$$P_{RF}(e^{j\omega}) = \frac{K_F W_{FRe}(e^{j\omega}) W_{Fn}(e^{j\omega}) W_{RRe}(e^{j\omega}) \left[1 - K_R W_{Rn}(e^{j\omega}) W_{RRe}(e^{j\omega}) W_{RFe}(e^{j\omega}) W_{FFe}(e^{j\omega}) \tilde{H}_{RF}(e^{j\omega})\right]}{D_2(e^{j\omega})} \tag{15}$$

$$\begin{aligned} D_2(e^{j\omega}) \quad = \quad & 1 - K_F W_{FRe}(e^{j\omega}) W_{RRe}(e^{j\omega}) W_{Fn}(e^{j\omega}) \tilde{H}_{FR}(e^{j\omega}) \\ & - K_R K_F W_{RFe}(e^{j\omega}) W_{FRe}(e^{j\omega}) W_{Rn}(e^{j\omega}) W_{Fn}(e^{j\omega}) W_{RRe}(e^{j\omega}) W_{FFe}(e^{j\omega}) \tilde{H}_{RR}(e^{j\omega}) \tilde{H}_{FF}(e^{j\omega}) \\ & - K_R W_{RFe}(e^{j\omega}) W_{FFe}(e^{j\omega}) W_{Rn}(e^{j\omega}) \tilde{H}_{RF}(e^{j\omega}) \\ & + K_R K_F W_{RFe}(e^{j\omega}) W_{FRe}(e^{j\omega}) W_{RRe}(e^{j\omega}) W_{FFe}(e^{j\omega}) W_{Rn}(e^{j\omega}) W_{Fn}(e^{j\omega}) \tilde{H}_{RF}(e^{j\omega}) \tilde{H}_{FR}(e^{j\omega}) \end{aligned} \tag{16}$$

$$D_3(e^{j\omega}) = 1 - K_R K_F W_{FFe}(e^{j\omega}) W_{RRe}(e^{j\omega}) W_{Fn}(e^{j\omega}) W_{Rn}(e^{j\omega}) \tilde{H}_{RR}(e^{j\omega}) \tilde{H}_{FF}(e^{j\omega}) \tag{17}$$

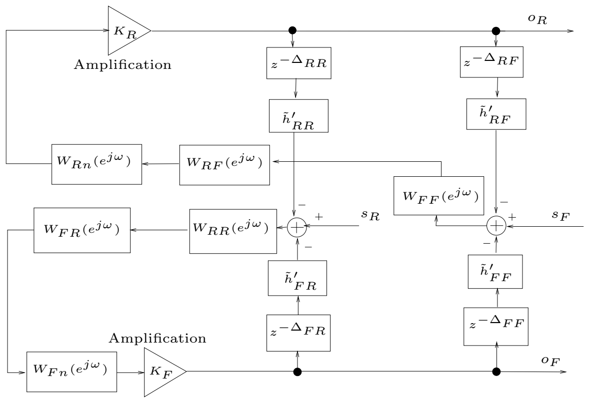

Figure 3: *Two-channel speech reinforcement system with diferentiated treatment techniques for closed-loop subsystems and for open-loop subsystems.*

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. 3 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 modeled as a delay block of Δ_{XY} samples (X refers to the loudspeakers, front or rear, and Y refers to the microphones, front or rear) 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. 3, the transfer funtions of the front-rear channel and the rear-front channel follow expressions (14) and (15), where

$$W_{RF}(e^{j\omega}) = \frac{1}{1 + K_R W_{FFe}(e^{j\omega}) W_{Rn}(e^{j\omega}) \tilde{H}_{RF}(e^{j\omega})} \tag{18}$$

$$W_{FR}(e^{j\omega}) = \frac{1}{1 + K_F W_{RRe}(e^{j\omega}) W_{Fn}(e^{j\omega}) \tilde{H}_{FR}(e^{j\omega})}. \tag{19}$$

are the transfer functions of the proposed ESF for the closed-loop subsystems.

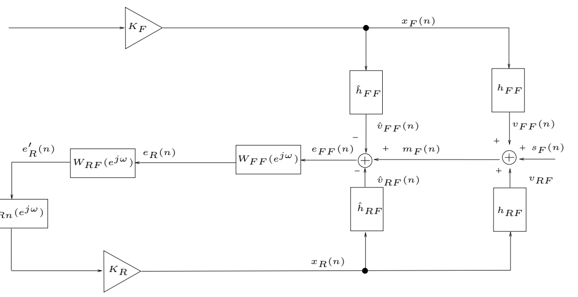

Figure 4: *Detailed diagram of the front-rear channel in a two-channel system with diferentiated treatment techniques for closed-loop subsystems and for open-loop subsystems.*

Substituting (18) and (19) into (14) and (15), the front-rear and rear-front transfer functions satisfy

$$P_{FR}(e^{j\omega}) = \frac{K_F W_{Fn}(e^{j\omega}) W_{FFe}(e^{j\omega})}{D_3(e^{j\omega})} \tag{20}$$

$$P_{RF}(e^{j\omega}) = \frac{K_R W_{Rn}(e^{j\omega}) W_{RRe}(e^{j\omega})}{D_3(e^{j\omega})}, \tag{21}$$

where $D_3(e^{j\omega})$ follows expression (17)

Thus, the stability of the reinforcement system, assuming that the ESF 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 misdjustment 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 [1], that using the magnitude of the misadjustment function is the best option to increase the stability of the system.

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, according to Fig. 4.

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

$$r_{RF}(n) = x_R(n) * w_{FFe}(n) * \tilde{h}_{RF}(n), \tag{22}$$

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

$$S_{r_{RF}}(e^{j\omega}) = S_{x_R}(e^{j\omega}) \cdot \left|W_{FFe}(e^{j\omega})\tilde{H}_{RF}(e^{j\omega})\right|^2, \quad (23)$$

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

According to (23), 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

$$\left|W_{FFe}(e^{j\omega})\tilde{H}_{RF}(e^{j\omega})\right|^2 = \frac{S_{r_{RF}}(e^{j\omega})}{S_{x_R}(e^{j\omega})}. \quad (24)$$

The PSD of the rear output signal, according to Fig. 4, can be expressed as

$$S_{x_R}(e^{j\omega}) = S_{e_R}(e^{j\omega}) \cdot K_R^2 \cdot \left|W_{RFe}(e^{j\omega})W_{Rn}(e^{j\omega})\right|^2, \quad (25)$$

and thus, combining (24) and (25) and substituting into (18), we can obtain the expression for the closed-loop ESF of the front-rear channel that responds to

$$W_{RF}(e^{j\omega}) = 1 - \sqrt{\frac{S_{r_{RF}}(e^{j\omega})}{S_{e_R}(e^{j\omega})}}, \quad (26)$$

wich 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

$$W_{FR}(e^{j\omega}) = 1 - \sqrt{\frac{S_{r_{FR}}(e^{j\omega})}{S_{e_F}(e^{j\omega})}}. \quad (27)$$

4. Residual Echo Power Spectral Density Estimation Method for the Closed-Loop Subsystems

The PSD of the error signals of each channel is directly accesible, so periodogram based methods can be used to obtain an estimate of $S_{e_R}(e^{j\omega})$ and $S_{e_F}(e^{j\omega})$. Nevertheless r_{FR} and r_{RF} are not directly accesible and alternative methods must be used. An estimate of the PSD of r_{FR} and r_{RF} can be obtained from the estimate of the PSD of $e_F(n)$ and $e_R(n)$ by using the iterative method described here.

Assuming stationarity of the speech signal on short periods of time (10-20 ms) we desire to obtain an estimate of the short-term PSD of the k-th segment of length L samples of residual echo, $S_{r_{FR}}(e^{j\omega})$ or $S_{r_{RF}}(e^{j\omega})$. For this, we use the optimal Wiener solution

$$H_{r_{FR}}(e^{j\omega};k) = \frac{S_{r_{FR}e_F}(e^{j\omega};k)}{S_{e_F}(e^{j\omega};k)} \quad (28)$$

$$H_{r_{RF}}(e^{j\omega};k) = \frac{S_{r_{RF}e_R}(e^{j\omega};k)}{S_{e_R}(e^{j\omega};k)}, \quad (29)$$

where $S_{r_{FR}e_R}(e^{j\omega};k)$ is the cross-power spectral density of residual echo and the error signal of the rear-front channel for the k-th segment and, $S_{r_{RF}e_F}(e^{j\omega};k)$ for the front-rear channel. We can express the short-time cross-power spectral densities as the Fourier transform of the short-time cross-correlation functions

$$R_{r_{FR}e_F}(m;kD) = E\left[r_{FR}(n;kD)e_F(n-m;kD)\right] \quad (30)$$

$$R_{r_{RF}e_R}(m;kD) = E\left[r_{RF}(n;kD)e_R(n-m;kD)\right], \quad (31)$$

where $E[\cdot]$ denotes expectation of the quantity within the brackets and $(n;kD)$ denotes n-th sample of the frame starting at sample kD.

The error signal of each channel, can be expressed as

$$e_R(n) = r_{RF}(n) + w_{FFe}(n) * [s_F(n) + r_{FF}(n) + b_F(n)] \quad (32)$$

$$e_F(n) = r_{FR}(n) + w_{RRe}(n) * [s_R(n) + r_{RR}(n) + b_R(n)] \quad (33)$$

where $r_{RF}(n)$ and $r_{FR}(n)$ are the residual echo associated with the electro-acoustic systems rear loudspeaker-enclosure-front microphone and front loudspeaker-enclosure-rear microphone respectively. $s_F(n)$ is the front speaker's voice signal and $s_R(n)$ is the rear speaker's one. $b_F(n)$ is the background noise signal picked up by the front microphones and $b_R(n)$, picked up by the rear ones. Thus, assuming that the background noise is statistically independent from the rest of the components of $e_R(n)$ or $e_F(n)$, equations (30) and (31) can be rewritten as

$$\begin{aligned} R_{r_{FR}e_F}(m;kD) = &\; E\left[r_{FR}(n;kD)r_{FR}(n-m;kD)\right] \\ &+E\left[r_{FR}(n;kD)s_R(n-m;kD)\right] * w_{RRe}(n) \\ &+E\left[r_{FR}(n;kD)r_{RR}(n-m;kD)\right] * w_{RRe}(n) \end{aligned} \quad (34)$$

$$\begin{aligned} R_{r_{RF}e_R}(m;kD) = &\; E\left[r_{RF}(n;kD)r_{RF}(n-m;kD)\right] \\ &+E\left[r_{RF}(n;kD)s_F(n-m;kD)\right] * w_{FFe}(n) \\ &+E\left[r_{RF}(n;kD)r_{FF}(n-m;kD)\right] * w_{FFe}(n). \end{aligned} \quad (35)$$

Modeling every LEM path impulse response as a delay block followed by a linear system according to Fig. 3

$$\begin{aligned} R_{r_{FR}e_F}(m;kD) = &\; E\left[r_{FR}(n;kD)r_{FR}(n-m;kD)\right] \\ &+K_F E\left[e_F(n-\Delta_{FR};kD)s_R(n-m;kD)\right] * w'_{RRe}(n) \\ &+K_F E\left[e_F(n-\Delta_{FR};kD)r_{RR}(n-m;kD)\right] * w'_{RRe}(n) \end{aligned} \quad (36)$$

$$\begin{aligned} R_{r_{RF}e_R}(m;kD) = &\; E\left[r_{RF}(n;kD)r_{RF}(n-m;kD)\right] \\ &+K_R E\left[e_R(n-\Delta_{RF};kD)s_F(n-m;kD)\right] * w'_{FFe}(n) \\ &+K_R E\left[e_R(n-\Delta_{RF};kD)r_{FF}(n-m;kD)\right] * w'_{FFe}(n), \end{aligned} \quad (37)$$

with

$$w'_{RRe}(n) = \tilde{h}'_{FR}(n) * w_{RRe}(n) * w_{FRe}(n) \quad (38)$$

$$w'_{FFe}(n) = \tilde{h}'_{RF}(n) * w_{FFe}(n) * w_{RFe}(n), \quad (39)$$

where

$$\tilde{h}'_{FR}(n) = \mathcal{F}^{-1}\left\{H'_{FR}(e^{j\omega}) - \hat{H}'_{FR}(e^{j\omega})\right\} \quad (40)$$

$$\tilde{h}'_{RF}(n) = \mathcal{F}^{-1}\left\{H'_{RF}(e^{j\omega}) - \hat{H}'_{RF}(e^{j\omega})\right\}. \quad (41)$$

Without loss of generality, let consider that the delay Δ_{RF} or Δ_{FR} are equal to p_{RF} and p_{FR} times D respectively, where D is the time shift between consecutive frames, and p_{RF} and p_{FR} are integers greater that one. Therefore, the k-th frame of residual echo $r_{RF}(n;k)$ and $r_{FR}(n;k)$ depends on previous frames of its respective error signal $e_R(n;(k-p_{RF}D))$ and $e_F(n;(k-p_{FR}D))$. Thus, equations (36) and (37) can be now expressed as

$$R_{r_{FR}e_F}(m;kD) = E\left[r_{FR}(n;kD)r_{FR}(n-m;kD)\right] + K_F E\left[e_F(n;(k-p_{FR})D)s_R(n-m;kD)\right] * w'_{RRe}(n) + K_F E\left[e_F(n;(k-p_{FR})D)r_{RR}(n-m;kD)\right] * w'_{RRe}(n) \tag{42}$$

$$R_{r_{RF}e_R}(m;kD) = E\left[r_{RF}(n;kD)r_{RF}(n-m;kD)\right] + K_R E\left[e_R(n;(k-p_{RF})D)s_F(n-m;kD)\right] * w'_{FFe}(n) + K_R E\left[e_R(n;(k-p_{RF})D)r_{FF}(n-m;kD)\right] * w'_{FFe}(n). \tag{43}$$

The second and the third term of (42) and (43) depend on the correlation between different frames, which is expected to be small since speech is not stationary. Therefore, these two terms can be considered negligible compared to the first one. The short-time cross-power spectral density of the residual echo and the error signal for each channel can be considered to be

$$S_{r_{RF}e_R}(e^{j\omega};k) = S_{r_{RF}}(e^{j\omega};k) \tag{44}$$

$$S_{r_{FR}e_F}(e^{j\omega};k) = S_{r_{FR}}(e^{j\omega};k). \tag{45}$$

Thus, an estimate of the Wiener filters of (29) and (28) can be obtained by using

$$\hat{H}_{r_{FR}}(e^{j\omega};k) = \frac{\hat{S}_{r_{FR}}(e^{j\omega};k)}{\hat{S}_{e_F}(e^{j\omega};k)} \tag{46}$$

$$\hat{H}_{r_{RF}}(e^{j\omega};k) = \frac{\hat{S}_{r_{RF}}(e^{j\omega};k)}{\hat{S}_{e_R}(e^{j\omega};k)}, \tag{47}$$

where $\hat{S}_{e_F}(e^{j\omega};k)$ and $\hat{S}_{e_R}(e^{j\omega};k)$ are estimates of the PSD of the front and rear error signal respectively and $\hat{S}_{r_{RF}}(e^{j\omega};k)$ and $\hat{S}_{r_{RF}}(e^{j\omega};k)$ are estimates of the PSD of the residual echo of the front-rear channel and the rear-front channel respectively, for the k-th frame.

An instantaneous estimate of the PSD of each channel residual echo for the next frame can be obtained from the PSD of its corresponding error signal by using the appropiate Wiener filter defined in (46) and (47)

$$\tilde{S}_{r_{FR}}(e^{j\omega};k+1) = \left[\lambda_e + (1-\lambda_e)\,\hat{H}_{r_{FR}}(e^{j\omega};k)\right]^2 \hat{S}_{e_R}(e^{j\omega};k) \tag{48}$$

$$\tilde{S}_{r_{RF}}(e^{j\omega};k+1) = \left[\lambda_e + (1-\lambda_e)\,\hat{H}_{r_{RF}}(e^{j\omega};k)\right]^2 \hat{S}_{e_F}(e^{j\omega};k), \tag{49}$$

where $0 \leq \lambda_e \leq 1$ is a bias term that avoids the clipping of any frequency to zero during the estimation process. Afterwards, we perform an exponential time averaging using a forgetting factor δ_e

$$\hat{S}_{r_{FR}}(e^{j\omega};k) = \delta_e \hat{S}_{r_{FR}}(e^{j\omega};k-1) + (1-\delta_e)\cdot\tilde{S}_{r_{FR}}(e^{j\omega};k) \tag{50}$$

$$\hat{S}_{r_{RF}}(e^{j\omega};k) = \delta_e \hat{S}_{r_{RF}}(e^{j\omega};k-1) + (1-\delta_e)\cdot\tilde{S}_{r_{RF}}(e^{j\omega};k). \tag{51}$$

Finally, according to (26) and (27) the ESF for each channel are computed using

$$W_{RF}(e^{j\omega};k) = 1 - \sqrt{\frac{\hat{S}_{r_{RF}}(e^{j\omega};k)}{\hat{S}_{e_R}(e^{j\omega};k)}} \tag{52}$$

$$W_{FR}(e^{j\omega};k) = 1 - \sqrt{\frac{\hat{S}_{r_{FR}}(e^{j\omega};k)}{\hat{S}_{e_F}(e^{j\omega};k)}}. \tag{53}$$

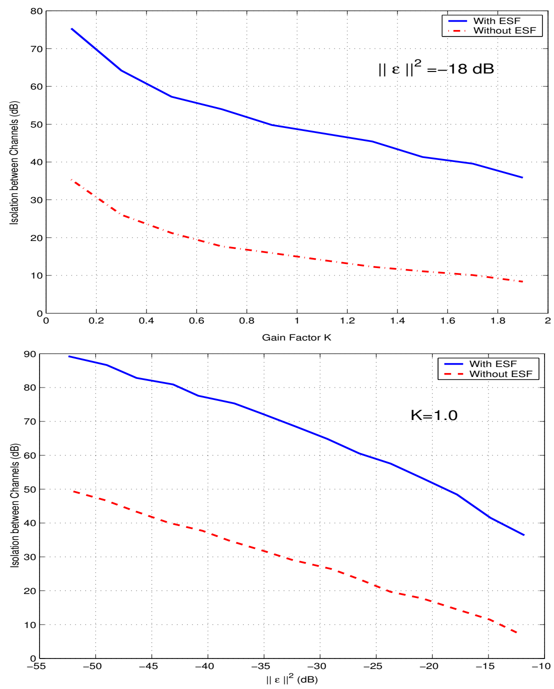

Figure 5: *Isolation between channels with and without ESF in the closed-loop subsystems.*

5. 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

$$\|\epsilon\|^2 = \frac{\sum_{k=0}^{L}\left|h'_k - \hat{h}'_k\right|^2}{\sum_{k=0}^{L}\left|h'_k\right|^2}, \tag{54}$$

where h'_k is the kth coefficient of the impulse response of the real electro-acoustico path and $\hat{h}'_k$ is the kth 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.

$$I_{RF} = \frac{E\left[|x_R(n)|^2\right]}{E\left[|x_F(n)|^2\right]}. \tag{55}$$

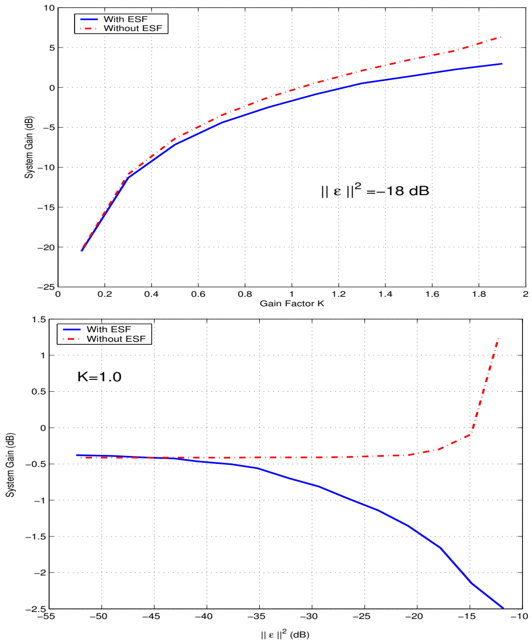

Figure 6: *System gain with and without ESF in the closed-loop subsystems.*

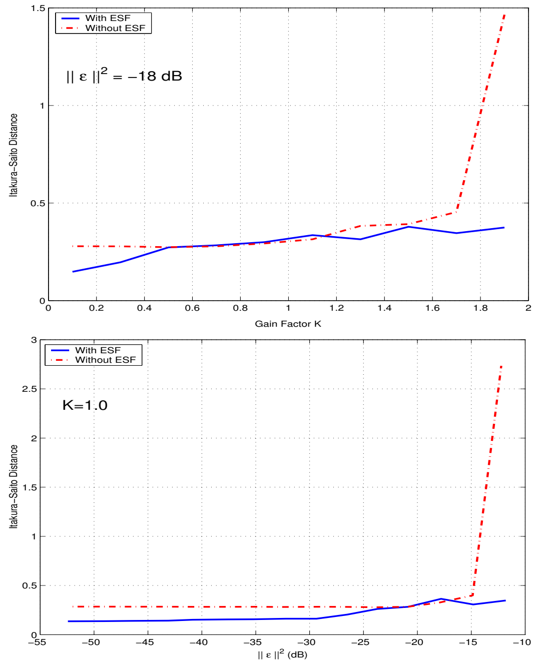

Figure 7: *Itakura-Saito Distance between the input signal and the reinforced speech with and without ESF in the closed-loop subsystems.*

In the upper half of Fig. 5, the evolution of the isolation between channels with the gain factor K for $||\epsilon||^2 = -18dB$ 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 $||\epsilon||^2$.

To shown that there is no degradation in terms of system gain decrease or distortion increase, the evolution of the system gain with K for $||\epsilon||^2 = -18dB$, and the evolution of the system gain with $||\epsilon||^2$ for $K = 1.0$ is presented in Fig. 6. In Fig. 7, the evolution with K for $||\epsilon||^2 = -18dB$ and with $|\epsilon||^2$ for $K = 1.0$ of the Itakura-Saito distance between the input signal and the corresponding output signal is presented.

In the lower half of Fig. 6 it can be seen that the System Gain increases dramatically for values of $||\epsilon||^2$ above -15 dB. The same effect can be observed in both parts of Fig. 7 regarding the distortion for high values of K or $||\epsilon||^2$. This is due to the appearence of howling as the system is very close to instability and strong tonal components are present in the output signal.

6. 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 in this paper, and the optimal expression for the transfer function of the Echo Suppression filter that ensures inconditional 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 is proposed and evaluated in this paper. This solution is based on an estimation of the residual echo power spectral density that is also presented here. 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 decresing the gain of the system or increasing the speech distortion.

7. References

[1] A. Ortega, E. Lleida, and E. Masgrau, "Speech reinforcement system for car cabin communications," 2005, accepted for publication in IEEE Transactions on Speech and Audio Processing.

[2] F. Gallego, E. Lleida, E. Masgrau, and A. Ortega, "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. 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 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. Hänsler 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, May 2002, pp. 1893–1896.