Acoustic Feedback and Echo Cancellation in Speech Communication Systems

Bruno Catarino Bispo

A thesis submitted for the degree of
Doctor of Philosophy

Department of Electrical and Computer Engineering
Faculdade de Engenharia da Universidade do Porto
Porto, Portugal

May, 2015

Abstract

During the last decades, signal processing techniques have been developed to attenuate the undesired effects caused by the acoustic coupling between loudspeaker and microphone in communication systems. In public address (PA) or sound reinforcement systems, the acoustic coupling causes the system to have a closed-loop transfer function that, depending on the amplification gain, may become unstable. Consequently, the maximum stable gain (MSG) of the system has an upper limit. In teleconference or hands-free communication systems, the acoustic coupling causes the speaker to receive back his/her voice signal after talking, which sounds like an echo and disturbs the communication.

The use of adaptive filters to identify the acoustic coupling path and estimate the resulting acoustic signal, which is subtracted from the microphone signal, is the state-of-art approach to remove the influence of the acoustic coupling in PA and teleconference systems. This approach is very attractive because, in theory, it would completely remove the effects caused by the acoustic coupling if the adaptive filter exactly matches the acoustic coupling path. And it has been applied to develop acoustic feedback cancellation (AFC) and acoustic echo cancellation (AEC) methods for PA and teleconference systems, respectively.

In a PA system, however, a bias is introduced in the adaptive filter coefficients if the traditional gradient-based or least-squares-based adaptive filtering algorithms are used. This issue occurs because the system input signal and the loudspeaker signal are highly correlated, mainly for colored signals as speech, and limits the performance of the AFC methods available in the literature. This work aims to primarily investigate the use of cepstral analysis to develop more effective AFC methods. It is proved that the cepstra of the microphone signal and the error signal may contain time domain information about the system, including its open-loop impulse response. Then, two new AFC methods are proposed: the AFC method based on the cepstrum of the microphone signal (AFC-CM) and the AFC method based on the cepstrum of the error signal (AFC-CE). The AFC-CM and AFC-CE methods estimate the feedback path impulse response from the cesptra of the microphone signal and error signal, respectively, to update the adaptive filter. Simulation results demonstrated that, for speech signals in a PA system with one microphone and one loudspeaker, the AFC-CM and AFC-CE methods can estimate the feedback path impulse

response with misalignment (MIS) of -9.8 and -25 dB, respectively, and increase the MSG of the PA system by 12 and 30 dB, respectively. And, for speech signals in a PA system with one microphone and four loudspeakers, the AFC-CM and AFC-CE methods can estimate the overall feedback path impulse response with MIS of -10.4 and -25 dB, respectively, and increase the MSG of the PA system by 11.3 and 30.6 dB, respectively.

The second theme of this work is related to AEC in teleconference systems. In the mono-channel case, the conventional AEC approach works quite well and any gradient-based or least-squares-based adaptive filtering algorithm can be used. In this work, the cepstral analysis, which is the basis of the proposed AFC methods, is applied in a different way to develop a new methodology for mono-channel AEC. This methodology estimates the cepstrum of the echo path through the cepstra of the microphone signal and the loudspeaker signal, and then computes an estimate of the echo path impulse response that is used to update the adaptive filter. Three new mono-channel AEC methods are proposed: the AEC method based on cepstral analysis with no lag (AEC-CA), the improved AEC-CA (AEC-CAI) and the AEC method based on cepstral analysis with lag (AEC-CAL). The AEC-CAI and AEC-CAL methods perform partially or completely the inverse of the overlap-and-add method using the adaptive filter as estimate of the echo path, respectively, in order to improve the computation of the frame of the microphone signal and thus the estimate of the echo path impulse response. The drawback of the AEC-CAL method is an estimation lag equal to the length of the echo path.

Simulation results demonstrated that the methods are sensitive to the ambient noise conditions and perform well in terms of MIS. However, they may perform worse than the traditional adaptive filtering algorithms in the first seconds of the Echo Return Loss Enhancement (ERLE) metric. In order to overcome this issue in the first seconds of ERLE, hybrid AEC methods that combine the AEC-CAI and AEC-CAL with two traditional adaptive filtering algorithms are also proposed. For speech signals and an echo-to-noise ratio (ENR) of 30 dB, the AEC-CAI and AEC-CAL methods can estimate the echo path impulse response with mean MIS of -18.7 and -18.6 dB, respectively, and attenuate the echo signal with mean ERLE of 32.4 and 36.1 dB, respectively. And the hybrid methods that use the AEC-CAI and AEC-CAL methods can estimate the echo path impulse response with mean MIS of -20 and -19.9 dB, respectively, and attenuate the echo signal with mean ERLE of 35.1 and 35.4 dB, respectively.

In stereophonic AEC (SAEC), a bias is introduced in the adaptive filter coefficients because of the high correlation between the loudspeaker signals if they are originated from the same sound source. Consequently, the adaptive filters converge to solutions that depend on impulse responses of the transmission room and the echo cancellation worsens if these impulse responses change. In order to overcome this problem, this work proposes two hybrid methods based on sub-band frequency shifting (FS) to decorrelate the loudspeaker signals before feeding them to the adaptive filters: Hybrid1 and Hybrid2. The Hybrid1 method applies a frequency shift of 5 Hz at the frequencies above 4 kHz and the traditional

half-wave rectifier (HWR) in the remaining frequencies. The Hybrid2 applies a frequency shift of 5 Hz at the frequencies above 4 kHz, a frequency shift of 1 Hz at the frequencies between 2 and 4 kHz and the HWR in the remaining frequencies. Simulation results demonstrated that the Hybrid1 and Hybrid2 methods cause the adaptive filters to estimate the impulse responses of the echo paths with MIS of -12.1 and -13 dB, respectively, thereby making the SAEC system less sensitive to variations in the transmission room. And the Hybrid1 and Hybrid2 methods produce stereo speech signals with a subjective sound quality of 85.4 and 87.2, respectively, in 100.

Resumo

Durante as últimas décadas, técnicas de processamento de sinal têm sido desenvolvidas para atenuar os indesejados efeitos causados pelo acoplamento acústico entre alto-falante e microfone em sistemas de comunicação. Em sistemas de comunicação ao público (PA) ou reforço sonoro, o acoplamento acústico faz o sistema ter uma função de transferência em malha fechada que, dependendo do ganho de amplificação, pode tornar-se instável. Consequentemente, o máximo ganho estável (MSG) do sistema tem um limite superior. Em sistemas de teleconferência ou comunicação com mãos livres, o acoplamento acústico faz o usuário receber de volta a sua própria voz logo após falar, a qual soa como um eco e perturba a comunicação.

O uso de filtros adaptativos para identificar o percurso de acoplamento acústico e estimar o resultante sinal acústico, o qual é subtraído do sinal do microfone, é a abordagem estado-da-arte para remover a influência do acoplamento acústico nos sistemas PA e de teleconferência. Essa abordagem é muito atrativa porque, na teoria, removeria completamente os efeitos causados pelo acoplamento acústico se o filtro adaptativo corresponder exatamente ao percurso de acoplamento acústico. E tem sido utilizada para desenvolver métodos de cancelamento de realimentaçãa acústica (AFC) e de cancelamento de eco acústico (AEC) para sistemas PA e de teleconferência, respectivamente.

Em um sistema PA, entretanto, um viés é introduzido nos coeficientes do filtro adaptativo se os tradicionais algoritmos de filtragem adaptativa baseados no gradiente descendente ou mínimos quadrados forem utilizados. Isso ocorre porque o sinal de entrada do sistema e o sinal do alto-falante são altamente correlacionais, principalmente para sinais coloridos como voz, e limita o desempenho dos métodos AFC disponíveis na literatura. Esse trabalho objetiva principalmente investigar o uso da análise cepstral para desenvolver métodos AFC mais eficazes. Prova-se que os cepstros do sinal do microfone e do sinal de erro podem conter informação no domínio do tempo sobre o sistema, incluindo a sua resposta ao impulso em malha aberta. Em seguida, dois novos métodos AFC são propostos: o método AFC baseado no cepstro do sinal do microfone (AFC-CM) e o método AFC baseado no cepstro do sinal de erro (AFC-CE). Os métodos AFC-CM e AFC-CE estimam a resposta ao impulso do percurso de realimentação a partir dos

cepstros do sinal do microfone e do sinal de erro, respectivamente, para atualizar o filtro adaptativo. Resultados de simulações demonstraram que, para sinais de voz em sistemas PA com um microfone e um alto-falante, os métodos AFC-CM e AFC-CE podem estimar a resposta ao impulso do percurso de realimentação com desalinhamento (MIS) de -9.8 e -25 dB, respetivamente, e aumentar o MSG do sistema PA em 12 e 30 dB, respetivamente. E, para sinais de voz em sistemas PA com um microfone e quatro alto-falantes, os métodos AFC-CM e AFC-CE podem estimar a resposta ao impulso do percurso geral de realimentação com MIS de -10.4 and -25 dB, respetivamente, e aumentar o MSG do sistema PA em 11.3 e 30.6 dB, respectivamente.

O segundo tema desse trabalho está relacionado com AEC em sistemas de teleconferência. No caso mono-canal, a abordagem AEC convencional funciona muito bem e qualquer algoritmo de filtragem adaptativa baseado no gradiente descendente ou mínimos quadrados pode ser utilizado. Nesse trabalho, a análise cepstral, que é a base dos métodos AFC propostos, é aplicado de uma maneira diferente para desenvolver uma nova metodologia para AEC mono-canal. Essa metodologia estima o cepstro do percurso de eco através dos cepstros do sinal do microfone e do sinal do alto-falante, e em seguida calcula uma estimativa da resposta ao impulso do percurso de eco que é utilizada para atualizar o filtro adaptativo. Três novos métodos AEC mono-canal são propostos: o método AEC baseado em análise cesptral sem atraso (AEC-CA), o AEC-CA melhorado (AEC-CAI) e o método AEC baseado em análise cesptral com atraso (AEC-CAL). Os métodos AEC-CAI e AEC-CAL realizam de maneira parcial e completa o inverso do método de sobreposição-e-soma, respectivamente, para melhorar o cálculo da janela do sinal do microfone e assim a estimativa da resposta ao impulso do percurso de eco. A desvantagem do método AEC-CAL é um atraso de estimação igual ao comprimento do percurso de eco.

Resultados de simulações demonstraram que os métodos são sensíveis às condições de ruído ambiente e têm um bom desempenho em termos de MIS. No entanto, eles podem apresentar um desempenho pior que os tradicionais algoritmos de filtragem adaptativa nos primeiros segundos do métrica *Echo Return Loss Enhancement* (ERLE). Com o intuito de superar esse problema nos primeiros segundos do ERLE, métodos AEC híbridos que combinam os AEC-CAI e AEC-CAL com dois tradicionais algoritmos de filtragem adaptativa são propostos. Para sinais de voz e uma razão eco-ruído de 30 dB, os métodos AEC-CAI e AEC-CAL podem estimar a resposta ao impulso do percurso de eco com MIS médio de -18.7 e -18.6 dB, respectivamente, e atenuar o sinal de eco com ERLE médio de 32.4 e 36.1 dB, respectivamente. E os métodos híbridos que utilizam AEC-CAI e AEC-CAL podem estimar a resposta ao impulso do percurso de eco com MIS médio de -20 e -19.9 dB, respectivamente, e atenuar o sinal de eco com ERLE médio de 35.1 e 35.4 dB, respectivamente.

Em AEC estéreo (SAEC), um viés é introduzido nos coeficientes dos filtros adaptativos por causa da alta correlação entre os sinais dos alto-falantes se eles foram gerados da mesma fonte sonora. Consequentemente, os filtros adaptativos convergem para soluções

que dependem de respostas ao impulso na sala de transmissão e o cancelamento de eco piora se essas respostas ao impulso mudam. Com o intuito de superar esse problema, esse trabalho propõe dois métodos híbridos baseados em deslocamento frequencial em sub-bandas para descorrelacionar os sinais dos alto-falantes antes de usá-los nos filtros adaptativos: Híbrido1 e Híbrido2. O método Híbrido1 aplica um descolamento de 5 Hz nas frequências maiores que 4 kHz e o tradicional retificador de meia-onda (HWR) nas restantes frequências. O método Híbrido2 aplica um descolamento de 5 Hz nas frequências maiores que 4 kHz, um descolamento de 1 Hz nas frequências entre 2 e 4 kHz e o tradicional retificador de meia-onda (HWR) nas restantes frequências. Resultados de simulações demonstraram que os métodos Híbrido1 e Híbrido2 fazem os filtros adaptativos estimarem as respostas ao impulso dos percursos de eco com MIS de -12.1 e -13 dB, respectivamente, tornando assim o sistema SAEC menos sensível às variações na sala de transmissão. E os métodos Híbrido1 e Híbrido2 produzem sinais de voz estéreos com qualidade subjetiva de 85.4 e 87.2, respectivamente, em 100.

Acknowledgements

I would like to thank my advisor, Professor Diamantino Rui da Silva Freitas, for providing me with the guidance and support throughout the PhD journey.

I am grateful to Professors Rui Manuel Esteves Araújo and Aníbal João de Sousa Ferreira for their motivation and helpful advices. I also thank my PhD colleagues Pedro Miguel de Luís Rodrigues and João Neves Moutinho for having contributed to my personal and professional time at Porto and for their valuable discussion. I would like to express my gratitude to Ricardo Jorge Pinto de Castro, a friend who accompanied me during the PhD studies.

I would like to acknowledge the financial support provided by the FCT (Fundação para a Ciência e a Tecnologia) through the scholarship SFHR/BD/49038/2008, which was fundamental to carrying out this research.

Finally, I would like to thank my family, in particular my parents Nelson and Maria Alice, for all their love, encouragement and invaluable support.

Contents

List of Abbreviations

ΔMSG Increase in Maximum Stable Gain

A/D Analog-to-Digital

AEC Acoustic Echo Cancellation

AEC-CA Acoustic Echo Cancellation based on Cepstral Analysis with no Lag

AEC-CAI Acoustic Echo Cancellation based on Cepstral Analysis with no Lag - Improved

AEC-CAL Acoustic Echo Cancellation based on Cepstral Analysis with Lag

AEQ Automatic Equalization

AES Acoustic Echo Suppression

AFC Acoustic Feedback Cancellation

AFC-CE Acoustic Feedback Cancellation Method based on the Cepstrum of the Error Signal

AFC-CM Acoustic Feedback Cancellation Method based on the Cepstrum of the Microphone Signal

AGC Automatic Gain Control

AIF Adaptive Inverse Filtering

CQS Continuous Quality Scale

D/A Digital to Analog

DCR Degradation Category Rating

DM Delay Modulation

DTD Double-talk Detector

ENR Echo-to-Noise Ratio

ERLE Echo Return Loss Enhancement

FEP Feedback Existence Probability

FFT Fast Fourier Transform

FIR Finite Impulse Response

FS Frequency Shifting

HA Hearing Aid

HWR Half-Wave Rectifier

IFFT Inverse Fast Fourier Transform

IIR Infinite Impulse Response

IMSD Interframe Magnitude Slope Deviation

IPMP Interframe Peak Magnitude Persistence

LPTV Linear Periodically Time-Varying

LSB Lower SideBand

LTV Linear Time-Varying

MIS Misalignment

MOS Mean Opinion Score

MSBG Maximum Stable Broadband Gain

MSE Mean Square Error

MSG Maximum Stable Gain

NGC Nyquist's Gain Condition

NHS Notch-Filter-based Howling Suppression

PA Public Address

PAPR Peak-to-Average Power Ratio

PEM-AFC Prediction Error Method based Adaptive Feedback Canceller

PEM-AFROW Prediction Error Method based on Adaptive Filtering with Row Operations

PHPR Peak-to-Harmonic Power Ratio

PM Phase Modulation

PNPR Peak-to-Neighboring Power Ratio

PTPR Peak-to-Threshold Power Ratio

SAEC Stereophonic Acoustic Echo Cancellation

SNR Signal-to-Noise Ratio

SSB Single Sideband

USB Upper Sideband

VAD Voice Activity Detector

W-PESQ Wideband Perceptual Evaluation of Speech Quality

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Research Motivation

Communication is a necessity of human beings and speech is their most fundamental communication tool, carrying not only a linguistic information but also an emotional expression [1]. With current technologies, speech communication systems have been established in order to fulfill this need and make life easier. Invariably, the communication systems use microphones and loudspeakers to pick up and play back the speech signals, respectively.

Figure 1.1 illustrates a communication system with N_l loudspeakers and N_m microphones operating in the same acoustic environment. The acoustic coupling between a loudspeaker and a microphone cause the signal of the kth loudspeaker, which is hereafter called loudspeaker signal $x_k(n)$, to be picked up by the lth microphone after going through several paths, which constitute the corresponding acoustic coupling path, and thus return into the communication system.

The acoustic coupling path includes the direct path, if it exists, and a large number of paths given by reflections. These paths cause a delay and an attenuation in the signal. As the attenuation typically increases with path length, only a finite number of paths need to be considered. For simplicity, the feedback path also includes the characteristics of the D/A converter, loudspeaker, microphone and A/D converter. Although some non-linearities may occur, for example because of loudspeaker saturation, it is almost always considered that these devices have unit responses and the feedback path is linear. Therefore, the acoustic coupling path between the kth loudspeaker and the lth microphone is usually defined as a finite impulse response (FIR) filter $F_{kl}(q, n)$.

Let the system input signal $u_l(n)$ be the source signal $v_l(n)$ added to the ambient noise signal $r_l(n)$, i.e., $u_l(n) = v_l(n) + r_l(n)$, and, for simplicity, also include the characteristics of the microphone and A/D converter. The resulting microphone signal $y_l(n)$ is defined

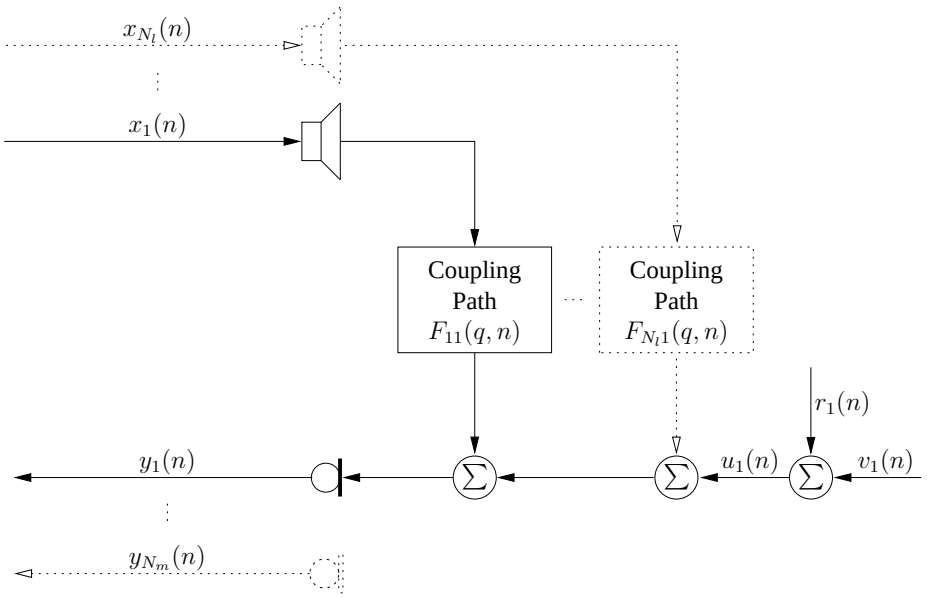


Figure 1.1: Acoustic couplings between loudspeakers and microphones.

as

$$y_l(n) = u_l(n) + \sum_{k=1}^{N_l} \mathbf{f}_{kl}(n) * x_k(n), \quad \text{for } l = 1, \ldots, N_m. \tag{1.1}$$

The microphone signal $y_l(n)$ is the system input signal $u_l(n)$ added to a sum of N_l undesired signals originating from the acoustic couplings, which are hereafter called coupling signals. The sum of the N_l coupling signals is hereafter called overall coupling signal. The existence of the acoustic coupling is inevitable and can generate some annoying effects which can disturb the communication or even make it impossible [2, 3, 4, 5, 6].

1.1.1 Public Address or Sound Reinforcement Systems

In a public address (PA) system, a speaker employs microphones and loudspeakers along with an amplification system to apply a gain on his/her voice signal aiming to be heard by a large audience in the same acoustic environment. Considering only one microphone, the microphone captures the desired system input signal, the microphone signal is amplified and then sent to the loudspeakers [2]. Because of the acoustic couplings, the loudspeaker signals are unavoidably fed back into the microphone, thereby leading to the so-called problem of acoustic feedback [2]. In this case, the acoustic couplings, acoustic coupling paths, coupling signals and overall coupling signal are called acoustic feedbacks, acoustic feedback paths, feedback signals and overall feedback signal, respectively. Therefore, a closed signal loop is created which causes the system input signal to circulate in the PA system and be played back several times by the loudspeakers. As the time delay caused by the amplification system is generally small, the overall feedback signal generally cannot be audibly distinguished from the system input signal and just sounds like reverberation.

The acoustic feedback limits the performance of a PA system in two ways. First and most important, depending on the amplification gain, the closed-loop transfer function of the PA system may become unstable resulting in a howling artifact, a phenomenon known as Larsen effect [3, 4]. This howling will be very annoying for all the audience and the amplification gain generally has to be reduced. As a consequence, the maximum stable gain (MSG) of the PA system has an upper limit [3, 4]. Second, even if the MSG is not exceeded, the sound quality is affected by excessive reverberation or ringing.

In order to overcome the Larsen effect, several methods have been developed over the last 50 years [2]. Among them, two approaches have been widely used: frequency shifting (FS) and notch-filter-based howling suppression (NHS). The former shifts the entire spectrum of the microphone signal by a few Hz so that its spectral peaks fall into spectral valleys of the feedback path after few loops [7, 8, 9, 10, 11, 12, 13, 14]. The latter detects the frequency components that may generate instability and then decreases the amplification gain applied to them by means of notch filters [2, 15, 16].

The FS and NHS methods smooth the gain of the open-loop transfer function of the PA system [2, 13, 14]. The amount of achievable smoothness depends on the magnitude difference between the peaks and valleys of the open-loop frequency response. When the amplification system is a broadband gain, the waveform of the open-loop frequency response will depend only on the feedback path frequency response. The Schroeder's statistics analysis of a feedback path frequency response states that, if the open-loop gain could be perfectly smoothed, a maximum increase in the MSG of about 10 dB may be achieved [10]. Some references reported increases in the MSG up to 14 dB [2, 13, 14].

However, the FS and NHS methods change not only the overall feedback signal but also the system input signal, which implies a fidelity loss of the PA system, and do not remove the reverberation caused by the acoustic feedback. Moreover, the FS methods may insert audible degradations depending on the amount of frequency shift employed [2, 10, 13, 14]. The NHS is a pre-active approach that first needs the occurrence of the Larsen effect to hereupon detect the frequency component responsible for the howling, compute the notch filter and remove the frequency component from the system. During the inherent processing time, the audience is exposed to the howling [3]. In fact, both methods assume the existence of the Larsen effect and only concern to control it.

Nowadays, the results obtained by the FS and NHS methods are becoming less acceptable and they are being replaced by the acoustic feedback cancellation (AFC) approach [3]. The AFC approach uses adaptive filters to identify the feedback paths and estimate the feedback signals, which are subtracted from the microphone signal [2, 3]. Ideally, if the adaptive filters exactly match the feedback paths, the overall feedback signal is completely removed from the microphone signal and thus the PA system has no longer a closed-loop transfer function. As a consequence, the MSG can be infinite. In practice, the AFC methods stand out for producing the best results with regard to MSG and sound quality [2].

Nevertheless, owing to the amplification system, the system input and loudspeaker signals are highly correlated, mainly when the source signal is colored as speech. Since the system input signal acts as interference to the adaptive filter, a bias is introduced in the adaptive filter coefficients if the traditional gradient-based or least-squares-based adaptive filtering algorithms are used [2, 17, 18, 19]. Consequently, the adaptive filter only partially cancels the feedback signal and applies distortion to the system input signal. Mostly, the solutions available in the literature to overcome the bias problem try to reduce the correlation between the system input and loudspeaker signals but still using the traditional adaptive filtering algorithms [2]. However, the additional processing to accomplish this decorrelation must not perceptually affect the quality of the signals [2]. Therefore, the challenge is to develop AFC methods that achieve unbiased estimates of the feedback paths without affecting the quality of the signals. And as AFC is a recent approach, there may be room for improvement.

1.1.2 Teleconference or Hands-Free Communication Systems

In a teleconference system, individuals or groups employ microphones and loudspeakers along with a VoIP system to communicate remotely. Each individual or group is located at one acoustic environment with one or more microphones to pick up its own voice signal and one or more loudspeakers to play back the voice signals of the others. For a specific individual or group, its acoustic environment is called transmission room while the acoustic environments of the others are called reception rooms. The acoustic couplings in the reception rooms may cause that, after talking, a speaker receives back his/her own voice signal in the transmission room. Owing to the delay of hundreds of milliseconds caused by the communication channel, the overall coupling signal is audibly distinguished from the speaker's signal and thus is called as echo. The occurrence of this acoustic echo is annoying and should be eliminated or, at least, attenuated [5, 6].

Although a closed signal loop may exist because of the couplings paths in both transmission and reception rooms, it is considered that the coupling paths in the transmission rooms do not occur or are eliminated. This is the difference between the acoustic echo and feedback problems: in acoustic echo, there is no closed signal loop and thereby the system may not become unstable. Therefore, the acoustic coupling limits the performance of a teleconference system only with regard to sound quality, which is affected by echoes. Moreover, in the acoustic echo problem, the loudspeaker signals, source signals, ambient noise signals, coupling paths and coupling signals are commonly called far-end speaker signals, near-end speaker signals, near-end ambient noise signals, echo paths and echo signals, respectively.

In order to attenuate the acoustic echo, two approaches have been developed over the last 20 years: acoustic echo suppression (AES) and acoustic echo cancellation (AEC). The former, also denominated loss control, attenuates the loudspeaker and/or microphone signals depending on the comparison between their energies with pre-defined thresholds

and between themselves [12, 20]. Similarly to AFC, the latter estimates the echo signal, usually by means of an adaptive filter, and subtracts it from the microphone signal [12, 21].

The operation of AES is straightforward. If only the loudspeaker signals are active, it attenuates the microphone signals in order to avoid the transmission of acoustic echoes. If only the source signal is active, it attenuates the loudspeaker signals in order to avoid the reception of noise. The problem occurs when both loudspeaker and source signals are simultaneously active, which is defined as a double-talk situation [12]. In this case, the method decides which signal, of the loudspeaker or microphone, is attenuated or not. Therefore, AES methods preclude full-duplex communication [12]. In fact, the AES approach assumes the existence of the acoustic echo and only concerns to control it.

Nowadays, the AES approach is practically in disuse and the teleconference and hands-free communication systems widely use the AEC approach. Although the standard [21] does not specify a technique to estimate the echo signals, adaptive filters are commonly used to identify the echo paths and estimate the echo signals, which are subtracted from the microphone signals. Ideally, if the adaptive filters exactly match the echo paths, the overall echo signal is completely removed from the microphone signal. The drawback compared to the AES approach is a higher computational complexity.

In the mono-channel case, the AEC methods work quite well and the only concern is not updating the adaptive filters in the absence of echo signals and in the presence of double-talk. For the first case, voice activity detectors (VAD) are used. For the second, double-talk detectors (DTD). However, in the multi-channel case, a bias is introduced in the adaptive filter coefficients because of the strong correlation between the loudspeaker signals if they are originated from the same sound source [5, 6, 22]. As undesirable consequences, the adaptive filters converge to solutions that depend on conditions of the transmission room and the cancellation worsens if these conditions change [5, 6, 22]. The solutions available in the literature to overcome the bias problem try to decorrelate the loudspeaker signals. However, the additional processing to accomplish this decorrelation must not perceptually affect the quality of the multi-channel signals, including modifications in the spatial image of the sound source, which is particularly difficult to achieve. Therefore, the challenge is to develop AEC methods that achieve unbiased estimates of the echo paths without affecting the perceptual quality of the signals.

1.2 Research Goals

In the light of the above discussion, it is clear that the use of adaptive filters to cancel the effects of the acoustic feedback/echo is a trend. The theoretical and practical advantages in performance have justified the continuous development of methods based on adaptive filtering and their applications in real-world products. The drawback is a high computational complexity because the adaptive filters generally require a few thousand coefficients in order to model the acoustic feedback/echo paths with sufficient accuracy.

However, the use of only adaptive filters generally is not sufficient to produce satisfactory results. In mono-channel AEC, control mechanisms are necessary to avoid disturbances in the adaptive filter update. In multi-channel AEC, besides the control mechanisms, it is also necessary additional processing to decrease the cross-correlation between the loudspeaker signals in order to improve the performance of adaptive filtering. In AFC, even in mono-channel case, additional processing is also required to decrease the cross-correlation between the loudspeaker and system input signals.

The present work is primarily concerned with AFC in PA systems for speech signals as source signals. A new approach will be proposed to update the adaptive filter in order to avoid the bias problem in the adaptive filter coefficients and thus increase the MSG of a PA system. Following this approach, two new AFC methods will be developed. Unlike the traditional AFC methods, it will be not necessary to apply any processing to the signals that travel in the system other than the adaptive filter. Therefore, for an AFC method, the fidelity of the PA system and the quality of the system signals will be as high as possible because they will only depend on the accuracy of the adaptive filter. The performance of the proposed methods will be evaluated considering single and multiple feedback paths.

Secondly, this work will address AEC in teleconference systems for speech signals as source signals. In the mono-channel case, it is possible to find in the literature the application of time-domain, time-domain block, fullband frequency-domain and subband frequency-domain adaptive filtering algorithms. The basis of the proposed AFC methods will be used to develop a new approach for mono-channel AEC. Based on this approach, three new AEC methods will be developed. In the stereo case, two new pre-processors will be proposed to reduce the bias problem in the adaptive filter coefficients and then improve the performance of the stereophonic acoustic echo cancellation (SAEC) .

1.3 Outline and Contributions

The focus of the present work is concentrated in the development of signal processing techniques to improve the performance of AFC and AEC systems for speech signals. The organization and contribution of this work are as follows:

Chapter 2, Acoustic Feedback Control, introduces the problem of the acoustic feedback in PA systems and presents the basic principles behind several approaches to control the Larsen effect. Due to historical reasons, the FS and NHS approaches are discussed in detail. The added value of this chapter consists of a survey of the results available in the literature for these approaches.

Chapter 3, Acoustic Feedback Cancellation, addresses the acoustic feedback control based on adaptive filtering. The specific bias problem of AFC in PA systems caused by the strong correlation between the system input signal and loudspeaker signal is discussed.

The solutions available in the literature to overcome the bias problem and thus improve the performance of AFC systems are described. The state-of-art method is discussed and evaluated. The added value of this chapter consists of a survey of the methods available in the literature to overcome the specific problem of AFC in PA systems and a complete evaluation of the state-of-art method.

Chapter 4, Acoustic Feedback Cancellation Based on Cepstral Analysis, presents a complete cepstral analysis of PA and AFC systems. The contributions of this chapter is twofold: first, it is proved that the cepstra of the system signals contain time-domain information about the systems if some gain conditions are fulfilled; and second, two AFC methods based on the cepstral analysis of the system signals are proposed. The findings of this chapter were disseminated in the following publications:

[I] B. C. Bispo and D. R. S. Freitas, "On the use of cepstral analysis in acoustic feedback cancellation," *Digital Signal Processing*, 2015, http://dx.doi.org/10.1016/j.dsp.2015.03.003.

[II] D. R. S. Freitas and B. C. Bispo, "Acoustic feedback cancellation based on cepstral analysis," *Patent Application WO 2015/044915, PCT/IB2014/06883*, April 2015.

[III] B. C. Bispo, P. M. L. Rodrigues and D. R. S. Freitas, "Acoustic feedback cancellation based on cepstral analysis," in *Proceedings of 17th IEEE Conference on Signal Processing Algorithms, Architectures, Arrangements and Applications*, Poznan, Poland, September 2013, pp. 205–209.

Chapter 5, Acoustic Feedback Cancellation with Multiple Feedback Paths, is concerned with the evaluation of the proposed AFC methods in a PA system with multiple feedback paths. This is a practical situation that occurs when, for example, a PA system with one microphone, responsible for picking up the speaker signal, and several loudspeakers placed in different positions, responsible for playback and distributing the voice signal in the acoustic environment so that everyone in the audience can hear it, is used. This chapter formed the basis for the following publications:

[IV] B. C. Bispo and D. R. S. Freitas, "Performance evaluation of acoustic feedback cancellation methods in single-microphone and multiple-loudspeakers public address systems," in *Lecture Notes - Communications in Computer and Information Science*. Springer, to be published in 2015.

[V] B. C. Bispo and D. R. S. Freitas, "Evaluation of acoustic feedback cancellation with multiple feedback paths," in *Proceedings of 11th International Conference on Signal Processing and Multimedia Applications*, Vienna, Austria, August 2014, pp. 127–133.

Chapter 6, Acoustic Echo Cancellation, introduces the problem of the acoustic echo in teleconference systems. The cepstral analysis, which is the basis for the AFC methods proposed in the Chapter 4, is applied in a different way to develop a new approach to update the adaptive filters in mono-channel AEC. Then, three new mono-channel AEC methods are proposed. This study was published in:

[VI] B. C. Bispo and D. R. S. Freitas, "Acoustic echo cancellation based on cepstral analysis," in *Proceedings of 17th IEEE Conference on Signal Processing Algorithms, Architectures, Arrangements and Applications*, Poznan, Poland, September 2013, pp. 210–214.

Chapter 7, Multi-channel Acoustic Echo Cancellation, deals with AEC in multi-channel teleconference systems. The specific bias problem of multi-channel AEC caused by the strongly correlation between the loudspeaker signals is discussed. The solutions available in the literature to overcome the bias problem and then improve the performance of multi-channel AEC systems are described. Two new sub-band decorrelation methods are proposed. This research was explored in the following publication:

[VII] B. C. Bispo and D. R. S. Freitas, "Hybrid pre-processor based on frequency shifting for stereophonic acoustic echo cancellation," in *Proceedings of 20th European Signal Processing Conference*, Bucharest, Romania, August 2012, pp. 2447–2451.

Chapter 8 reports the final remarks and establishes plans for future work.

During the present research, the following additional articles were also published:

[VIII] P. M. L. Rodrigues, B. C. Bispo, D. R. S. Freitas, J. P. Teixeira and A. Carrere, "Evaluation of EEG spectral features in alzheimer disease discrimination," in *Proceedings of 21th European Signal Processing Conference*, Marrakech, Marocco, September 2013, pp. 1–5.

[IX] B. C. Bispo, P. A. A. Esquef, L. W. P. Biscainho, A. A. de Lima, F. P. Freeland, R. A. de Jesus, A. Said, B. Lee, R. Schafer, A. Kalker, "EW-PESQ: A quality assessment method for speech signals sampled at 48 kHz," *Journal of the Audio Engineering Society*, vol. 58, no. 4, pp. 251–268, April 2010.

[X] A. A. de Lima, S. L. Netto, L. W. P. Biscainho, F. P. Freeland, B. C. Bispo, R. A. de Jesus, R. Schafer, A. Said, B. Lee, A. Kalker, "Quality evaluation of reverberation in audioband speech signals," in *e-Business and Telecommunications - Communications in Computer and Information Science*, J. Filipe and M. S. Obaidat, Eds. Springer, 2009, vol. 48, pp. 384–396.

1.4 Notation

The discrete-time index is denoted by n. The superscript T denotes vector/matrix transpose. The symbol f_s denotes the sampling frequency while $T_s = \frac{1}{f_s}$ corresponds to the sampling period. The delay operator is denoted by q^{-1} such that $q^{-1}x(n) = x(n-1)$. A time-varying discrete-time filter with length L_F is represented by the polynomial [2, 23]

$$\begin{aligned} F(q,n) &= f_0(n) + f_1(n)q^{-1} + \ldots + f_{L_F-1}(n)q^{-(L_F-1)} \\ &= [f_0(n) \;\; f_1(n) \;\; \ldots \;\; f_{L_F-1}(n)] \begin{bmatrix} 1 \\ q^{-1} \\ \vdots \\ q^{-(L_F-1)} \end{bmatrix} \\ &= \mathbf{f}^T(n)\mathbf{q} \end{aligned} \tag{1.2}$$

or, alternatively, by its impulse response $\mathbf{f}(n)$. The vector $\mathbf{f}(n)$ has a constant length L_F but all of its values may vary over time n. The filter $F(q)$ refers to a time-invariant discrete-time filter with length L_F and impulse response $\mathbf{f}$. The filtering operation of a signal $x(n)$ with $F(q,n)$ is denoted as

$$F(q,n)x(n) = \mathbf{f}(n) * x(n) = \sum_{m=0}^{L_F-1} f_m(n)x(n-m). \tag{1.3}$$

Although the term transfer function should be reserved for the z-transform of $\mathbf{f}(n)$, $F(q,n)$ shall be called the transfer function of the linear system in (1.3) as in [2, 23].

The discrete-time Fourier Transform of $F(q,n)$, or $\mathbf{f}(n)$, and $x(n)$ are denoted by $F(e^{j\omega},n)$ and $X(e^{j\omega},n)$, respectively, where $\omega \in [0,\pi]$ is the normalized angular frequency, e is the Euler's number and j is the imaginary number.

Part I

Acoustic Feedback Cancellation

Chapter 2

Acoustic Feedback Control

2.1 Introduction

This chapter introduces the problem of acoustic feedback in PA systems. The acoustic feedback causes the PA system to have a closed-loop transfer function that, depending on the amplification gain, may become unstable resulting in a howling artifact, a phenomenon known as Larsen effect. This howling will be very annoying for all the audience and the amplification gain generally has to be reduced. As a consequence, the MSG of the PA system has a upper limit. Moreover, even if the MSG is not exceeded, the acoustic feedback causes the sound quality to be affected by excessive reverberation or ringing.

During the past years, several methods have been developed to control the Larsen effect and an overview of them is presented in this chapter. The FS and NHS methods are addressed in detail because they are the most widely used methods not only in the literature but also as in commercial products and for historic reasons. The FS method was proposed in the early 60's and consists in shifting, at each loop, the spectrum of the microphone signal by a few Hz. The NHS method consists in detecting the candidate frequencies to generate instability and then apply notch filters in order to remove these frequencies from the microphone signal. Both methods smooth the gain of the open-loop transfer function of the PA system and, in theory, can increase the MSG around 10 dB.

A survey of the results available in the literature for these approaches is presented and increases in the MSG up to 14 dB are reported. However, the FS and NHS methods change not only the feedback signal but also the system input signal, which implies a fidelity loss of the PA system, and do not remove the excessive reverberation caused by the acoustic feedback. Moreover, the FS methods may insert audible degradations depending on the amount of frequency shift. And the NHS methods first need the occurrence of the Larsen effect before removing the frequency component responsible for the howling. During the inherent processing time, the audience may be exposed to the howling. In fact, both methodologies assume the existence of the Larsen effect and only concern to control it.

2.2 The Acoustic Feedback Problem

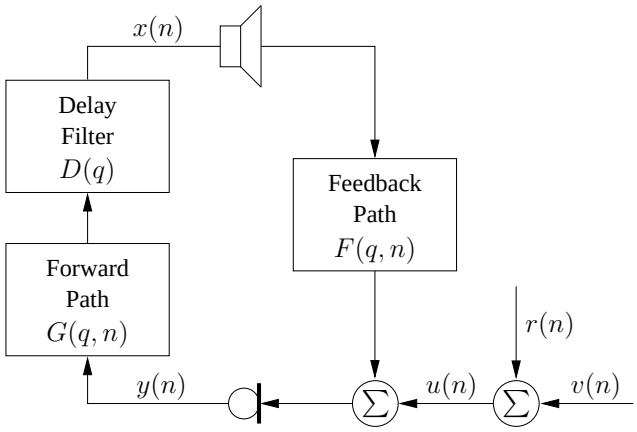


Figure 2.1: Acoustic feedback in a public address system.

A typical PA system with one microphone and one loudspeaker is illustrated in Figure 2.1. The loudspeaker signal $x(n)$ is fed back into the microphone through the feedback path $F(q,n)$. The feedback signal $\mathbf{f}(n) * x(n)$ is added to the source signal $v(n)$ and the ambient noise $r(n)$, generating the microphone signal

$$y(n) = \mathbf{f}(n) * x(n) + v(n) + r(n). \tag{2.1}$$

The forward path includes the characteristics of the amplifier and any other signal processing device inserted in that part of the signal loop, such as an equalizer. Although some non-linearities may exist, for example because of compression, the forward path is usually assumed to be linear and defined as a FIR filter

$$\begin{aligned} G(q,n) &= g_0(n) + g_1(n)q^{-1} + \ldots + g_{L_G-1}(n)q^{-(L_G-1)} \\ &= \mathbf{g}^T(n)\mathbf{q} \end{aligned} \tag{2.2}$$

with length L_G.

As it is sometimes found in the literature, a forward delay is represented separately by the delay filter

$$\begin{aligned} D(q) &= d_{L_D-1}q^{-(L_D-1)} \\ &= \mathbf{d}^T(n)\mathbf{q} \end{aligned} \tag{2.3}$$

with length L_D, which will be exploited further. For closed-loop analysis, $L_D > 1$.

Let the system input signal $u(n)$ be the source signal $v(n)$ added to the ambient noise signal $r(n)$, i.e., $u(n) = v(n) + r(n)$, and, for simplicity, also include the characteristics of the microphone and A/D converter. The system input signal $u(n)$ and the loudspeaker

signal $x(n)$ are related by the closed-loop transfer function of the PA system as

$$x(n) = \frac{G(q,n)D(q)}{1 - G(q,n)D(q)F(q,n)}u(n). \tag{2.4}$$

It is worth mentioning that, differently from the acoustic echo problem, the system input signal $u(n)$ and the loudspeaker signal $x(n)$ are directly related.

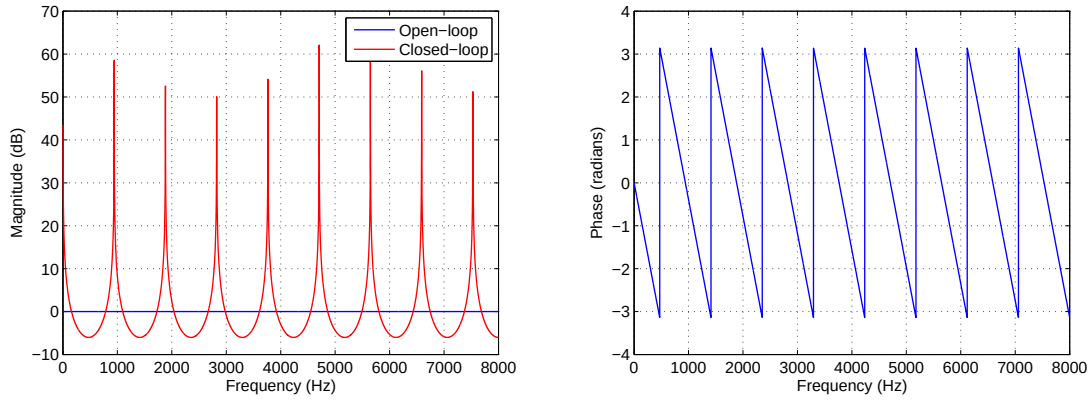
According to the Nyquist's stability criterion, the closed-loop system is unstable if there is at least one frequency ω for which [2, 24]

$$\begin{cases} \left|G(e^{j\omega},n)D(e^{j\omega})F(e^{j\omega},n)\right| \geq 1 \\ \angle G(e^{j\omega},n)D(e^{j\omega})F(e^{j\omega},n) = 2k\pi,\ k \in \mathbb{Z}. \end{cases} \tag{2.5}$$

Considering $f_s = 16$ kHz, Figure 2.2 shows the open-loop and closed-loop frequency responses for a PA system with $F(q) = q^{-1}$, $D(q) = q^{-16}$ and $G(q) = 1$. The closed-loop frequency response has peaks and valleys in locations that correspond to phase shifts equal to 0 and 180 degrees, respectively. The peaks are in theory infinite values and represent the instability of the PA system. This example shows that, as stated by the Nyquist's stability criterion, even though all the frequencies fulfill the gain condition of (2.5), only the frequencies that fulfill the phase condition of (2.5) generate instability. The conditions in (2.5) are essential because any acoustic feedback control method attempts to prevent either one or both of these conditions from being met [2].

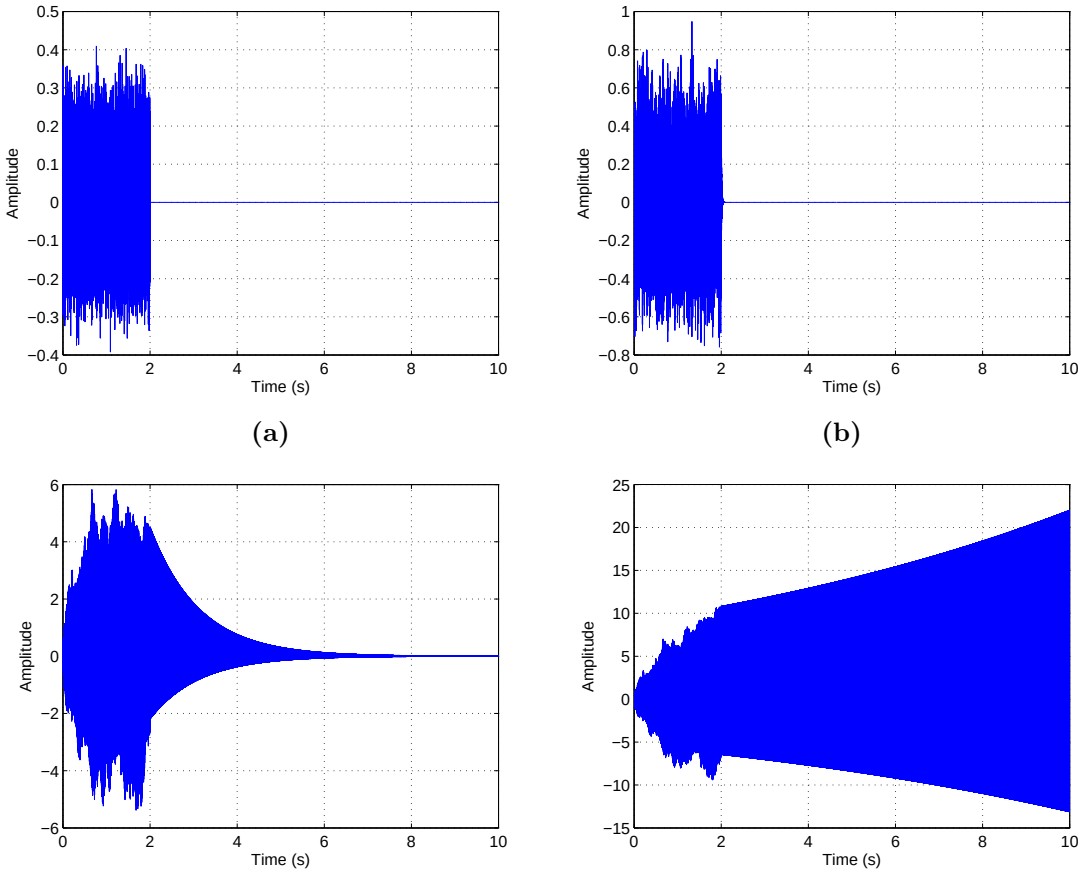
Figure 2.3 exemplifies the stability of the PA system as a function of the system gain through the waveform of the loudspeaker signal $x(n)$ over time. The system input signal $u(n)$ was a white noise with duration of 2 s followed by 8 s of silence and is showed in Figure 2.3a. The choice of the white noise was to excite the PA system at all frequencies and equally. And the use of the silence interval was to observe the behavior of the loudspeaker signal $x(n)$ after the end of the system input signal $u(n)$. Considering again $F(q) = q^{-1}$ and $D(q) = q^{-16}$, Figure 2.3b shows the loudspeaker signal $x(n)$ when $G(q) = 0.9$. Since $\left|G(e^{j\omega},n)D(e^{j\omega})F(e^{j\omega},n)\right| = 0.9$, the system is relatively far from the instability causing the loudspeaker signal $x(n)$ to end immediately after the system input signal $u(n)$.

When $G(q) = 0.999$, $\left|G(e^{j\omega},n)D(e^{j\omega})F(e^{j\omega},n)\right| = 0.999 \approx 1$ and the system is very close to instability, which causes the loudspeaker signal $x(n)$ to take some time to disappear after the end of the system input signal $u(n)$, as can be observed in Figure 2.3c. It is noteworthy that, after the end of $u(n)$, $x(n)$ is basically formed by audible howling but the system is stable because it naturally disappears. Finally, Figure 2.3d shows the loudspeaker signal $x(n)$ when $G(q) = 1.0001$. Since $\left|G(e^{j\omega},n)D(e^{j\omega})F(e^{j\omega},n)\right| = 1.0001 > 1$, the system is unstable which causes the loudspeaker signal $x(n)$ to never disappear from the system and its magnitude to increase every iteration such that $|x(n)| \to \infty$.



(a) **(b)**

Figure 2.2: Open-loop and closed-loop frequency responses for $F(q) = q^{-1}$, $G(q) = 1$ and $D(q) = q^{-16}$: (a) magnitude; (b) phase.



(a) **(b)**

(c) **(d)**

Figure 2.3: Illustration of the stability of a PA system when $F(q) = q^{-1}$ and $D(q) = q^{-16}$: (a) $u(n)$; (b),(c),(d) $x(n)$; (b) $G(q) = 0.9$; (c) $G(q) = 0.999$; (d) $G(q) = 1.0001$.

Indeed, the Nyquist's stability criterion states that if a frequency component is amplified with a phase shift equal to an integer multiple of 2π after going through the system open-loop transfer function, $G(q,n)D(q)F(q,n)$, this frequency component will never disappear from the system. After each loop through the system, its amplitude will increase resulting in a howling at that frequency, a phenomenon known as Larsen effect [2, 3]. This howling will be very annoying for the audience and the amplification gain at that frequency generally has to be reduced. As a consequence, the stable gain of the PA system at that frequency has an upper limit due to the acoustic feedback [2, 3, 4].

In general, the stable gain of the PA system is strictly limited as follows

$$\left|G(e^{j\omega},n)\right| < \frac{1}{|D(e^{j\omega})F(e^{j\omega},n)|},\ \omega \in P(n), \tag{2.6}$$

where $P(n)$ denotes the set of frequencies that fulfill the phase condition in (2.5), also called critical frequencies of the PA system, that is

$$P(n) = \left\{\omega | \angle G(e^{j\omega},n)D(e^{j\omega})F(e^{j\omega},n) = 2k\pi,\ k \in \mathbb{Z}\right\}. \tag{2.7}$$

It is worth emphasizing that the stable gain of the PA system has an upper limit at the frequencies $\omega \in P(n)$. For $\omega \notin P(n)$, the gain may be, in theory, infinite.

With the aim of quantifying the achievable amplification in a PA system, it is customary to define a broadband gain $K(n)$ of the forward path as the average magnitude of the forward path frequence response [2], i.e.,

$$K(n) = \frac{1}{2\pi}\int_0^{2\pi} |G(e^{j\omega},n)|\ d\omega \tag{2.8}$$

and extract it from the forward path $G(q,n)$ as follows

$$G(q,n) = K(n)J(q,n). \tag{2.9}$$

Assuming that $J(q,n)$ is known and $K(n)$ can be varied, the maximum stable gain (MSG) of the PA system is defined as [2]

$$\begin{aligned} \mathrm{MSG}(n)(\mathrm{dB}) &= 20\log_{10} K(n) \\ \text{such that } \max_{\omega\in P(n)} &\left|G(e^{j\omega},n)D(e^{j\omega})F(e^{j\omega},n)\right| = 1, \end{aligned} \tag{2.10}$$

resulting in

$$\mathrm{MSG}(n)(\mathrm{dB}) = -20\log_{10}\left[\max_{\omega\in P(n)} \left|J(e^{j\omega},n)D(e^{j\omega})F(e^{j\omega},n)\right|\right]. \tag{2.11}$$

In order to eliminate or, at least, to control the Larsen effect and thus to increase the

MSG of the PA system, several methods have been developed over the past 50 years and they can be divided in four main groups [2]. These groups, their main members and a brief description of each method are resumed below:

1. Phase-Modulation Methods: methods that insert in the system open-loop a processing device to change, at each loop, the phase of the system open-loop frequency response in order to prevent any frequency component from fulfilling the phase condition of the Nyquist's stability criterion during several loops.

    - Frequency Shifting (FS) [2, 7, 8, 9, 10, 11, 12, 13, 14, 25, 26, 27, 28]: the spectrum of the microphone signal is shifted so that its spectral peaks fall into spectral valleys of the feedback path.
    - Phase Modulation (PM) [2, 13, 14]: phase modulation is applied to the microphone signal with the aim of bypassing the phase condition of the Nyquist's stability criterion.
    - Delay Modulation (DM) [2, 13, 14]: the time delay of the microphone signal is varied around a time delay offset in order to bypass the phase condition of the Nyquist's stability criterion.

2. Gain Reduction Methods: methods that attempt to automatically act as a human operator controlling a system conducive to the Larsen effect. These actions are usually restricted to reduce the gain of the system open-loop so that the gain condition of the Nyquist's stability criterion is no longer fulfilled.

    - Automatic Gain Control (AGC) [2, 12, 29]: the gain is reduced equally in the entire frequency range by decreasing the broadband gain $K(n)$ defined in (2.8).
    - Automatic Equalization (AEQ) [2, 12]: the gain reduction is applied in subbands of the entire frequency range, namely in those subbands in which the gain is close to unity.
    - Notch Howling Suppression (NHS) [2, 12, 15, 16, 30, 31]: the gain is reduced in narrow bands of the entire frequency range around frequencies at which the gain is close to unity.

3. Spatial Filtering Methods [2, 32, 33, 34, 35, 36]: methods that use a microphone array that has maximum spatial response in the direction of the source signal and minimum spatial response in the direction of the loudspeaker, and/or a loudspeaker array that has maximum spatial response in the direction of the audience and minimum spatial response in the direction of the microphone, in order to enhance the source signal in the microphone while attenuating the feedback signal.

4. Room Modeling Methods: methods that attempt to identify the acoustic feedback path and then remove its influence from the PA system.

- Adaptive Inverse Filtering (or Adaptive Equalizer) (AIF) [2, 37, 38]: the inverse of the acoustic feedback path is identified and inserted in the system open-loop in order to equalize the microphone signal.
- Acoustic (or Adaptive) Feedback Cancellation (AFC) [2, 3, 4, 39, 40, 41, 42]: the acoustic feedback path is identified and used to estimate the feedback signal, which is subtracted from the microphone signal.

All the methods are well described in the literature, except the gain reduction methods which are mainly formed by patents, and reference [2] provides a thorough discussion about most of them as well as simulation results of several methods.

The phase modulation, spatial filtering and room modeling methods are proactive that attempt to prevent the Larsen effect before it occurs. On the other hand, the gain reduction methods are mostly reactive in the sense that the Larsen effect must first occur to hereupon be detected and eliminated. This is a disadvantage because, during the time between occurrence, detection and elimination of the Larsen effect, the audience is exposed to the howling [3].

Except for the spatial filtering and AFC methods, all the methods modify not only the feedback signal $\mathbf{f}(n) * x(n)$ but also the system input signal $u(n)$, which implies a fidelity loss of the PA system. However, this fidelity loss may be neglected if the methods do not perceptually affect the quality of the system signals, what is particularly difficult to achieve. The spatial filtering methods do not apply any processing to the system signals but constrain the placement of the microphone and/or loudspeaker.

The AFC methods, in theory, may modify only the feedback signal, thereby ensuring the fidelity of the PA system. In advantage over the spatial filtering methods, the AFC methods do not constrain the placement of the microphone and/or loudspeaker. Moreover, the AFC methods stand out for producing the best results and for being a recent technique [2, 3, 4], which may allow a large room for improvement.

For these reasons, the present work will focus on AFC methods. However, the FS and NHS methods will also be addressed because they are widely used not only in literature but also in commercial products and for historic reasons.

2.3 Frequency Shifting

One of the first approaches proposed to control the acoustic feedback in PA systems consists in frequency shifting (FS), at each loop, the microphone signal $y(n)$ by a few Hz, as illustrated in Figure 2.3. It was introduced by Schroeder in the early 60's and exploits the fact that the average spacing between large peaks and adjacent valleys in the frequency response $F(e^{j\omega})$ of large rooms is about 5 Hz [10]. Nevertheless, in general, this average spacing is related to the reverberation time of the room [8].

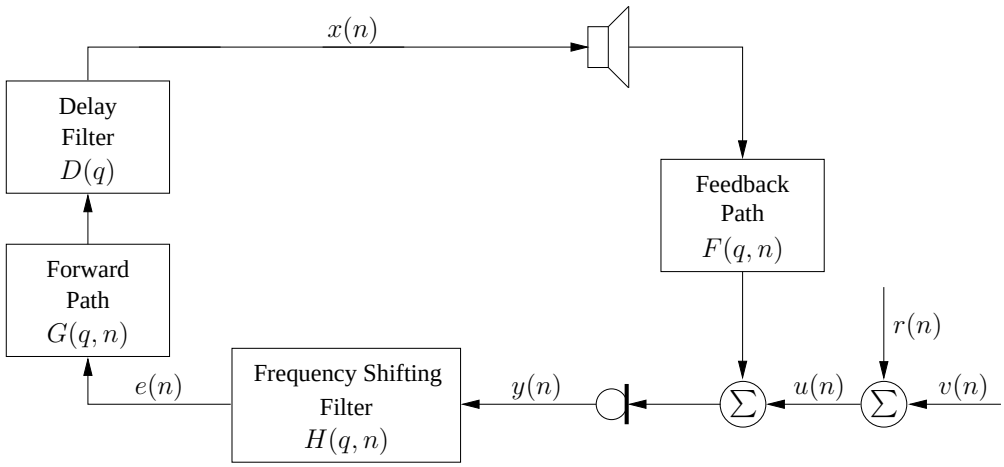


Figure 2.4: Acoustic feedback control using frequency shifting.

Considering that the PA system is close to instability and the forward path $G(q,n)$ is a gain, the howling will appear first at the critical frequency of the PA system where $|F(e^{j\omega},n)|$ is maximum. However, some loops through the system are necessary to make the howling audible. Then, in each loop, the spectrum $Y(e^{j\omega},n)$ of the microphone signal is shifted by a few cycles so that the frequency component responsible for the howling falls into a valley of $F(e^{j\omega},n)$ after a few loops and, thus, is attenuated before the howling becomes audible. As a consequence, the MSG of the PA system is expected to increase.

In fact, the FS smoothes the open-loop gain $\left|G(e^{j\omega},n)D(e^{j\omega})F(e^{j\omega},n)\right|$ of the PA system [2, 13, 14] such that, ideally, the MSG of PA system is determined by its average magnitude rather than peaks magnitude [2, 10]. A statistical analysis of frequency responses of large rooms was carried out in [10] and show that the highest peak exceeds the average level by about 10 dB. Therefore, if the open-loop gain could be perfectly smoothed, a maximum increase in the MSG of about 10 dB may be achieved [10]. Posteriorly, a similar analysis was done in [26] confirming these results.

The statistical analysis in [10] also states that the optimum frequency shift is equal to the average spacing between large peaks and adjacent valleys of the room frequency response, which is typically 5 Hz, or about $4/T_{60}$ Hz, where T_{60} is the reverberation time of the room. Practical experiments in [10, 13] confirmed the theory by showing that frequency shifts higher than the optimum value did not give any significant improvement and, in some cases, are even less effective. However, in practice, the optimum value of the frequency shift can be slightly different from the theory [13]. Moreover, there is no significant consistent difference between positive and negative shifts [10, 11, 13]. And, although the FS approach has the drawback of not preserving the harmonic relations between tonal components in voiced speech and music signals [2], a frequency shift of 5 Hz is inaudible both for speech and music signals [10].

As observed in [2, 13, 14], the behavior of a FS filter can be analyzed using the theory of linear time-varying (LTV) systems explored in [43]. From this analysis, the FS filter

can be interpreted as a linear periodically time-varying (LPTV) filter [2, 13, 14] and has, for a frequency shift of $f_0 = \omega_0(f_s/2\pi)$ Hz, the following frequency response [2]

$$H(e^{j\omega}, n) = e^{j\omega_0 n}. \tag{2.12}$$

The closed-loop transfer function of the system depicted in Figure 2.4 is defined as

$$\frac{x(n)}{u(n)} = \frac{G(q,n)D(q)H(q,n)}{1 - G(q,n)D(q)H(q,n)F(q,n)} \tag{2.13}$$

and, according to the Nyquist's stability criterion, is unstable if there is at least one frequency ω for which

$$\begin{cases} \left|G(e^{j\omega}, n)D(e^{j\omega})H(e^{j\omega}, n)F(e^{j\omega}, n)\right| \geq 1 \\ \angle G(e^{j\omega}, n)D(e^{j\omega})H(e^{j\omega}, n)F(e^{j\omega}, n) = 2k\pi,\ k \in \mathbb{Z}. \end{cases} \tag{2.14}$$

Then, considering the broadband gain $K(n)$ of the forward path defined in (2.8), the MSG of the PA system with an FS method is defined as

$$\begin{gathered} \mathrm{MSG}(n)(\mathrm{dB}) = 20\log_{10} K(n) \\ \text{such that } \max_{\omega \in P_H(n)} \left|G(e^{j\omega}, n)D(e^{j\omega})H(e^{j\omega}, n)F(e^{j\omega}, n)\right| = 1, \end{gathered} \tag{2.15}$$

resulting in

$$\mathrm{MSG}(n)(\mathrm{dB}) = -20\log_{10}\left[\max_{\omega \in P_H(n)} \left|J(e^{j\omega}, n)D(e^{j\omega})H(e^{j\omega}, n)F(e^{j\omega}, n)\right|\right]. \tag{2.16}$$

where $P_H(n)$ is the set of frequencies that fulfill the phase condition in (2.14), that is

$$P_H(n) = \left\{\omega | \angle G(e^{j\omega}, n)D(e^{j\omega})H(e^{j\omega}, n)F(e^{j\omega}, n) = 2k\pi,\ k \in \mathbb{Z}\right\}. \tag{2.17}$$

The increase in the MSG provided by the FS method is defined as

$$\Delta\mathrm{MSG}(n)(\mathrm{dB}) = -20\log_{10}\left[\frac{\max_{\omega \in P_H(n)} \left|J(e^{j\omega}, n)D(e^{j\omega})H(e^{j\omega}, n)F(e^{j\omega}, n)\right|}{\max_{\omega \in P(n)} \left|J(e^{j\omega}, n)D(e^{j\omega})F(e^{j\omega}, n)\right|}\right]. \tag{2.18}$$

2.3.1 Frequency Shifter

A digital frequency shifter can be implemented by means of a single sideband (SSB) modulator which uses cosine and sine as modulation functions along with a Hilbert filter [12, 26]. Consider a discrete-time signal $x(n)$ with a band-limited spectrum $X(e^{j\omega})$ that can be decomposed into negative and positive frequencies as follows

$$X(e^{j\omega}) = X_-(e^{j\omega}) + X_+(e^{j\omega}), \tag{2.19}$$

where $X_-(e^{j\omega})$ is the signal spectrum in the negative frequencies, lower sideband (LSB), and $X_+(e^{j\omega})$ is the spectrum in the positive frequencies, upper sideband (USB).

The frequency shift will be denoted by ω_0. If $\omega_0 > 0$, $X_-(e^{j\omega})$ will be shifted towards the normalized frequency π and $X_+(e^{j\omega})$ towards $-\pi$, yielding an LSB modulator. If $\omega_0 < 0$, the spectra will be shifted in opposite directions resulting in a USB modulator.

Aiming to generate the desired spectrum, the algorithm creates a first carrier signal by modulating the input signal $x(n)$ with a cosine function according to

$$x_{cos}(n) = x(n)\cos(n\omega_0). \tag{2.20}$$

In the frequency domain, the modulation results in two shifted versions of the input spectrum as follows

$$X_{cos}(e^{j\omega}) = \frac{1}{2}X\left(e^{j(\omega+\omega_0)}\right) + \frac{1}{2}X\left(e^{j(\omega-\omega_0)}\right), \tag{2.21}$$

which by replacing (2.19) in (2.21) becomes

$$\begin{aligned} X_{cos}(e^{j\omega}) =& \frac{1}{2}\left[X_-\left(e^{j(\omega-\omega_0)}\right) + X_+\left(e^{j(\omega+\omega_0)}\right)\right. \\ &\left. + X_-\left(e^{j(\omega+\omega_0)}\right) + X_+\left(e^{j(\omega-\omega_0)}\right)\right]. \end{aligned} \tag{2.22}$$

For an LSB modulator, the first and second terms on the right-hand side of (2.22) are the desired movements of the positive and negative frequencies of the input spectrum. However, the third and fourth terms on the right-hand side of (2.22) are undesired components that were shifted into the opposite directions. In order to eliminate them, the algorithm creates a second carrier signal by applying an Hilbert filter with impulse response $\mathbf{h}_{hil}$ to the input signal $x(n)$ according to

$$x_{hil}(n) = x(n) * \mathbf{h}_{hil}. \tag{2.23}$$

The frequency response of the Hilbert filter is defined as

$$H_{hil}(e^{j\omega}) = -j\,\mathrm{sgn}(\omega), \tag{2.24}$$

which means that the Hilbert filter shifts the phase of $X_-(e^{j\omega})$ by $\pi/2$ and the phase of $X_+(e^{j\omega})$ by $-\pi/2$. Then, in the frequency domain, (2.23) implies

$$X_{hil}(e^{j\omega}) = -j\,\mathrm{sgn}(\omega)X(e^{j\omega}). \tag{2.25}$$

The Hilbert filtered signal $x_{hil}(n)$ is modulated with a sine function leading to

$$x_{sin}(n) = x_{hil}(n)\sin(n\omega_0). \tag{2.26}$$

In the frequency domain, the modulation results in two shifted and multiplied versions of $X_{hil}(e^{j\omega})$ as follows

$$X_{sin}(e^{j\omega}) = j\frac{1}{2}X_{hil}\left(e^{j(\omega+\omega_0)}\right) - j\frac{1}{2}X_{hil}\left(e^{j(\omega-\omega_0)}\right), \tag{2.27}$$

which by replacing (2.19) and (2.25) in (2.27) becomes

$$\begin{aligned} X_{sin}(e^{j\omega}) =& \frac{1}{2}\left[X_-\left(e^{j(\omega-\omega_0)}\right) + X_+\left(e^{j(\omega+\omega_0)}\right)\right. \\ & \left. -X_-\left(e^{j(\omega+\omega_0)}\right) - X_+\left(e^{j(\omega-\omega_0)}\right)\right]. \end{aligned} \tag{2.28}$$

As in (2.22), the resulting spectrum in (2.28) is formed by two desired movements of the positive and negative frequencies of the input spectrum and two undesired components that were shifted into the opposite directions. But now, the undesired components have opposite signs compared to those from (2.22).

Therefore, the frequency shifted signal $x'(n)$ is obtained by adding the two modulated signals according to

$$\begin{aligned} X'(e^{j\omega}) &= X_{cos}(e^{jw}) + X_{sin}(e^{jw}) \\ &= X_-\left(e^{j(\omega-\omega_0)}\right) + X_+\left(e^{j(\omega+\omega_0)}\right). \end{aligned} \tag{2.29}$$

The block diagram of the digital frequency shifter is depicted in Figure 2.5, where the definition of $\mathbf{h}_{hil}$ and the need for the delay $q^{-N_{hil}}$ are explained in the following section.

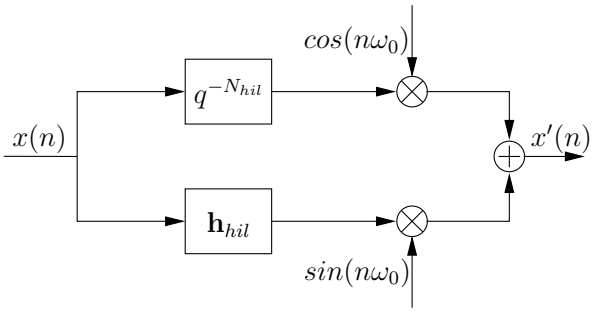


Figure 2.5: Block diagram of the frequency shifter.

2.3.2 Hilbert Filter

The impulse response of the Hilbert filter can be calculated by applying the inverse Fourier transform on (2.25), resulting in

$$h_{hil_m} = \begin{cases} 0, & \text{if } m \text{ is even,} \\ \dfrac{2}{m\pi}, & \text{else,} \end{cases} \tag{2.30}$$

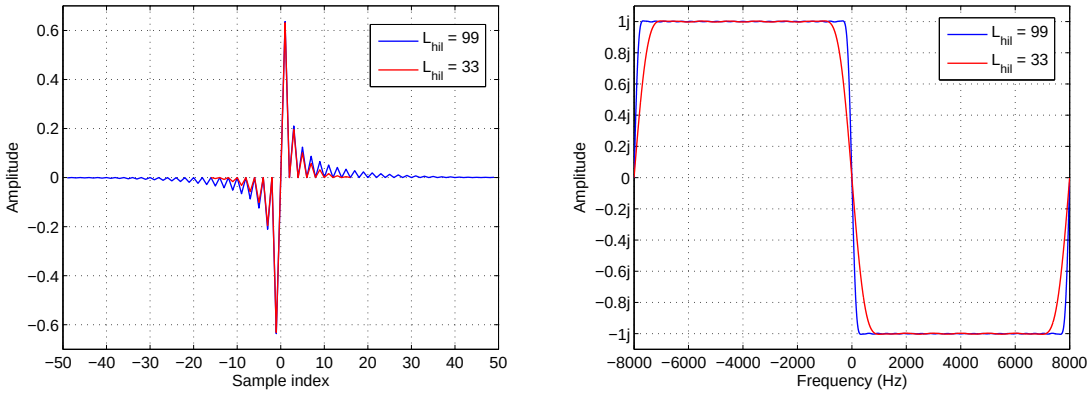
where m is the sample index.

The problem of (2.30) is twofold: $\mathbf{h}_{hil}$ is infinitely long and non-causal. Therefore, it must first be truncated to a range $m = -N_{hil}, \ldots, N_{hil}$ by means of a window function. And, second, it is necessary to shift the truncated solution by N_{hil} coefficients and, consequently, to delay the cosine modulated signal in (2.20) by N_{hil} samples. The resulting Hilbert filter is denoted by $\hat{\mathbf{h}}_{hil}$ and has length $L_{hil} = 2N_{hil} + 1$.

It is evident that the efficiency of this implementation of the frequency shifter depends on the length of the Hilbert filter: higher values of N_{hil} provide more accurate solutions but, at the same time, insert longer delays in the output signal. Fortunately, since the filter coefficients tend to zero as $|m|$ increases, the values of N_{hil} do not need to be very high in order for the filter $\hat{\mathbf{h}}_{hil}$ to have an accurate solution.

This trade-off between efficiency and filter length is illustrated in Figure 2.6 for $L_{hil} =$ 33 and 99 samples when $f_s = 16$ kHz and a Hamming window is used as windowing function. The frequency response $H_{hil}(e^{j\omega})$ of the Hilbert filter when $L_{hil} = 33$ presents transition bands with a considerable bandwidth, which causes the frequency components in these bands not to be properly shifted. This consequence can be softened by using higher order filters as $L_{hil} = 99$, resulting in shorter transition bands. However, the drawback is the higher intrinsic delay N_{hil}. Moreover, because of the Gibbs phenomenon [44], filters with sharper transition bands generate oscillations in the spectrum of its output signal around their cutoff frequencies which, if in the human audible range, may be perceptible.

One important property of the Hilbert transform is the orthogonality between its input and output signals [45, 46]. A discrete-time signal $x(n)$ with duration N and its



(a) **(b)**

Figure 2.6: Hilbert filter for different L_{hil} values and using a Hamming window: a) impulse response; b) frequency response.

corresponding Hilbert transformed signal $x_H(n)$ are orthogonal if and only if [45, 46]

$$\frac{1}{f_s}\sum_{n=0}^{N-1} x(n)x_H(n) = 0. \tag{2.31}$$

In order to verify this principle, an experiment was made using 100 speech signals with duration of 4 s, $f_s = 16$ kHz and a Hilbert filter $\hat{\mathbf{h}}_{hil}$ with $L_{hil} = 641$ (corresponding to a delay of 20 ms). The values on the left-hand side of (2.31) were calculated for each signal and are shown in Figure 2.7. Although non-zero, the resulting very low values confirm that the orthogonality is preserved.

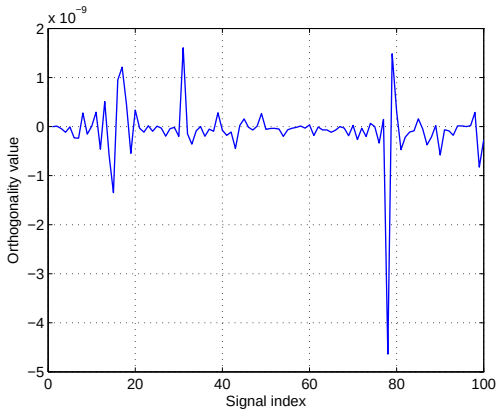


Figure 2.7: Orthogonality of the Hilbert transform.

2.3.3 Results of FS Systems in the Literature

In this section, results available in the literature about the use of FS to control the acoustic feedback in PA systems will be presented. Results from practical experiments where the increase in the MSG of the PA system, ΔMSG, was obtained by increasing the gain of the forward path $G(q, n)$ until instability occurred are presented in [7, 8, 10, 11, 12, 25, 26]. Following the same approach, results from simulated experiments are reported in [13, 14, 27]. Considering also simulated experiments, results where ΔMSG was mathematically calculated are presented in [2].

The evaluations carried out by Schroeder in [7, 8, 10] do not explain the nature of the source signal $v(n)$ used. Absolute values of frequency shifts up to 20 Hz were considered and the results confirmed the theoretical analysis about the optimum shift frequency present in [10] and previously discussed in this section. Values of ΔMSG up to 12 dB were achieved in a large auditorium and soundproof booth while ΔMSG values up to 11 dB were achieved in medium-size room. However, the subjectively acceptable value of ΔMSG was limited to 6 dB because of audible beating effects [7, 8, 10]. In [25], an analog frequency shifter is described in detail and the same subjectively acceptable ΔMSG of

6 dB is reported. Another analog implementation of the frequency shifter is described in [11], where an usable value of ΔMSG equal to 8 dB is presented.

Average values of ΔMSG obtained using speech signals at different power levels as the source signals $v(n)$ and three different rooms are presented in [12]. The frequency shifts were 6, 9 and 12 Hz, the frequency shifter was the one described in Section 2.3.1 and the forward path $G(q, n)$ was a gain. The average values of ΔMSG are in the range 1-2 dB in a lecture room, 3-4 dB in an entrance hall and 5-6 dB in an echoic chamber, which is a room of an acoustical research department that has a reverberation time more than one second. The maximum value of ΔMSG was obtained with a frequency shift of 9 Hz in the lecture room and with 12 Hz in the other rooms. Artifacts were audible for frequency shifts larger than 12 Hz.

ΔMSG values are reported in [26] considering two different rooms and several microphone configurations. The frequency shifter was the one described in Section 2.3.1, the frequency shift was 6 Hz, and the forward path $G(q, n)$ was a gain. Although this paper emphasizes the efficiency of the FS when the source signal $v(n)$ was speech and attenuation in the very low frequencies when $v(n)$ was audio due to the highpass nature of the Hilbert filter, the nature of the source signal used in the measurements is not clarified. The ΔMSG values are in the range 0.4-7 dB and no artifacts are noticeable.

Using 18 different microphone positions, average values of ΔMSG are presented in [27]. The frequency shifts were 2, 4, 6 and 8 Hz. In a simulated environment, the feedback path $F(q, n)$ was measured for each position of the microphone, the forward path $G(q, n)$ was a gain and the source signal $v(n)$ was white noise. The average values of ΔMSG are in the range 1.6-3.6 dB and the performance always improved as the frequency shift increased.

In [13, 14], ΔMSG values obtained with frequency shifts of $\pm\{0.5, 1, 2, 3, 4, 5\}$ Hz are reported. The source signal $v(n)$ was noise and the feedback path $F(q, n)$ was an electronic reverberation unity. In a first configuration, the forward path $G(q, n)$ was a gain followed by a electronic equalizer. In a second, the previous $G(q, n)$ was also followed by an electronic reverberation unity. The gain of $G(q, n)$ was increased while keeping the PA system stable and the loudspeaker signal $x(n)$ was monitored. In the first configuration, the ΔMSG values are in the range 5-9 dB and the maximum value was obtained with frequency shifts of ±2 Hz. In the second configuration, the ΔMSG values are on the range 8-15 dB and the maximum value was obtained with frequency shifts of ±4 Hz.

In a simulated environment, results obtained with frequency shifts of 5 Hz are presented in [2]. The source signals $v(n)$ were one speech signal with duration of $T = 30$ s and $f_s = 16$ kHz and one audio signal with duration of $T = 60$ s and $f_s = 44.1$ kHz. The feedback path $F(q, n)$ was a measured room impulse response until $t = {}^{3T}/_{4}$ s and then it was changed for other measured impulse response of the same room. The broadband gain of the forward path $G(q, n)$ was initialized to a value such that the PA system had an initial gain margin of 3 dB and remained at this value until $t = {}^{T}/_{4}$ s. During the next $t = {}^{T}/_{4}$ s, it was increased linearly (in dB scale) by 3 dB and remained at this value until the end of the

simulation. As said previously, in [2], the ΔMSG was mathematically calculated at each iteration which enabled display the values of ΔMSG over time, ΔMSG(n). Considering only the last $T/2$ s of simulation, the FS achieved an average ΔMSG of 1.1 dB and a maximum ΔMSG of 4.1 dB.

2.4 Notch Howling Suppression

Other widely used approach to control the acoustic feedback in PA systems is the notch-filter-based howling suppression (NHS). The NHS approach, depicted in Figure 2.8, consists of two stages: howling detection and notch filter design. The howling detection stage is responsible for detecting the frequencies that generate howling and providing a set of design parameters D_H. The notch filter design stage uses the parameter set D_H to design a bank of adjustable notch filters $H(q, n)$ that is inserted in the open-loop transfer function in order to remove, or attenuate, these frequency components from the microphone signal $y(n)$. As a consequence, the MSG of the PA system is expected to increase.

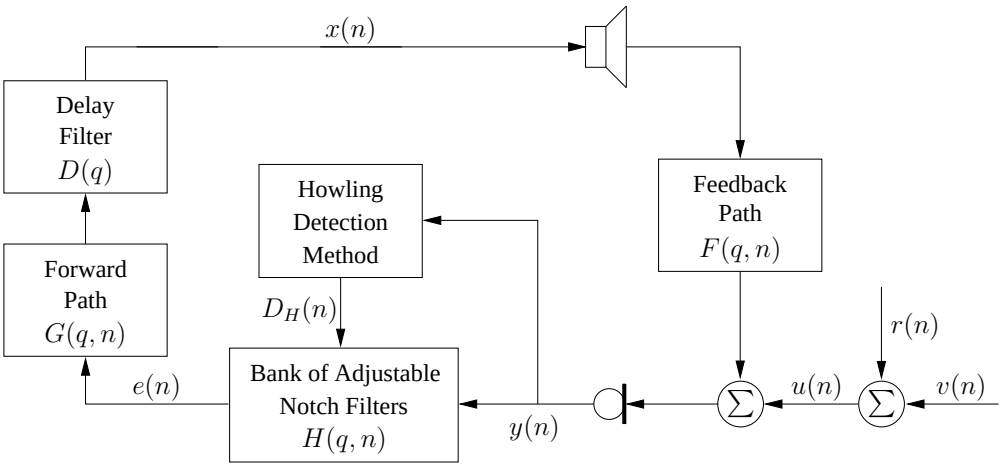


Figure 2.8: Acoustic feedback control using notch filters.

As previously mentioned, even when the PA system is close to instability, some loops through the system are necessary to make the howling audible. In the meantime, the NHS method should correctly detect the frequencies that generate howling, design and apply the notch filters. Otherwise, the audience will be exposed to the howling even if only for a short time.

In fact, as the FS approach discussed in Section 2.3, the NHS approach also smoothes the open-loop gain $\left|G(e^{j\omega}, n)D(e^{j\omega})F(e^{j\omega}, n)\right|$ of the PA system such that, ideally, the MSG of the PA system is determined by its average magnitude rather than peak magnitude [2]. If the open-loop gain could be perfectly smoothed, a maximum increase in the MSG of about 10 dB may be achieved as before [2, 10]. Apart from the $H(q, n)$ contents, the

system depicted in Figure 2.8 is equivalent to the one in Figure 2.4. Hence, its closed-loop transfer function, stability criterion, MSG(n), critical frequencies and ΔMSG(n) are defined, respectively, according to (2.13), (2.14), (2.15), (2.17) and (2.18).

The NHS literature mainly consists of patents and few experimental results have been reported [15]. Nevertheless, references [2, 15, 16] unified the framework for howling detection and provided a comparative evaluation of several howling detection criteria.

2.4.1 Howling Detection

The first stage of the NHS methods detects the frequencies $\tilde{\omega}_c$ that are candidates to generate howling and provides a set of design parameters D_H. It is assumed that the howling detection is performed on frames of the microphone signal $y(n)$ that, at discrete-time n, is defined as [2, 15, 16]

$$\mathbf{y}(n) = [y(n+P-M) \;\; y(n+P-M-1) \;\; \ldots \;\; y(n+P-1)]\,, \tag{2.32}$$

where M is the length and P is the hop size of the frame. The short-term spectrum $Y(e^{j\omega}, n)$ of the microphone signal is calculated using the Fast Fourier Transform (FFT) and usually includes a windowing function to reduce the spectral leakage.

The choice of the framing parameters M and P has a great influence on the performance of the howling detection methods. Small values of the frame length M provide a very fast howling detection such that the howling may be detected before it is really perceived. On the other hand, large values allow a better frequency resolution in the microphone signal spectrum which is very useful when working with narrowband notch filters. Values corresponding to 4.2, 85.3 and 92.9 ms have already been used in the literature [15, 16].

With respect to the frame hop size P, small values increase the computational complexity since the howling detection methods are applied more often. On the other hand, large values may result in a lag time between the howling detection and the application of the notch filters, unless the cascade $D(q)G(q,n)$ generates a delay of at least P samples [15, 16]. Generally, a good compromise is obtained with $25-50\%$ frame overlap [16].

A pre-defined number N_p of peaks are selected from the spectrum magnitude $|Y(e^{j\omega}, n)|$ of the microphone signal, where usually $1 \leq N_p \leq 10$ [15, 16]. These N_p frequency components are called candidate howling components and their angular frequency values form the set

$$D_{\tilde{\omega}_c}(n) = \{\tilde{\omega}_k\}_{k=1}^{N_p}. \tag{2.33}$$

A spectral peaking algorithm is usually applied to find the candidate howling frequencies but more advanced techniques, as detecting the frequency components that present increasing magnitude in successive frames, are also used. Thereafter, spectral and/or temporal features are calculated and combined in a howling detection criterion to determine

whether a candidate howling component really corresponds to a howling component or only to a tonal component of the source signal $v(n)$ [15, 16].

2.4.1.1 Signal Features

After detecting the candidate howling components and forming the set $D_{\tilde{\omega}_c}(n)$, some features of the microphone signal are calculated and used to classify them as real howling components or not. To this purpose, six spectral and time features have already been proposed to be used individually or together in order to establish howling detection criteria. Their definitions and brief explanations about them are listed below:

1. Peak-to-Threshold Power Ratio (PTPR) [2, 15, 16]: a spectral feature that determines the ratio between the power $\left|Y(e^{j\tilde{\omega}_k}, n)\right|^2$ of the candidate howling component and a fixed power threshold P_0, i.e.,

$$\text{PTPR}(\tilde{\omega}_k, n)\ [\text{dB}] = 10\log_{10}\frac{\left|Y(e^{j\tilde{\omega}_k}, n)\right|^2}{P_0}. \tag{2.34}$$

   The use of the PTPR feature in howling detection is explained by the fact that a howling should be suppressed only when it occurs with a minimum loudness. Thus, relatively large values for the PTPR feature are expected in howling components. The value of the power threshold P_0 is usually dependent on the sound reinforcement scenario.

2. Peak-to-Average Power Ratio (PAPR) [2, 15, 16, 31]: a spectral feature that determines the ratio between the power $\left|Y(e^{j\tilde{\omega}_k}, n)\right|^2$ of the candidate howling component and the average power $\hat{P}_y(n)$ of the microphone signal, i.e.,

$$\text{PAPR}(\tilde{\omega}_k, n)\ [\text{dB}] = 10\log_{10}\frac{\left|Y(e^{j\tilde{\omega}_k}, n)\right|^2}{\hat{P}_y(n)}, \tag{2.35}$$

   where

$$\hat{P}_y(n) = \frac{1}{M}\sum_{i=0}^{M-1}\left|Y(e^{j\omega_i}, n)\right|^2. \tag{2.36}$$

   The reason for the PAPR feature is that the power of howling components may be large when compared to the power of speech and audio components present in the microphone signal. Then, relatively large values for the PAPR feature are expected in howling components.

3. Peak-to-Harmonic Power Ratio (PHPR) [2, 15, 16]: a spectral feature that determines the ratio between the power $\left|Y(e^{j\tilde{\omega}_k}, n)\right|^2$ of the candidate howling component

and the power $\left|Y(e^{j\tilde{\omega}_k m}, n)\right|^2$ of its mth harmonic component, i.e,

$$\mathrm{PHPR}(\tilde{\omega}_k, n, m)\ [\mathrm{dB}] = 10\log_{10}\frac{\left|Y(e^{j\tilde{\omega}_k}, n)\right|^2}{\left|Y(e^{j\tilde{\omega}_k m}, n)\right|^2}. \tag{2.37}$$

The PHPR feature exploits the fact that, unlike voiced speech and tonal audio components, the howling does not have a harmonic structure unless saturation occurs on microphone or loudspeaker. Hence, relatively large values for the PHPR feature are expected in howling components.

4. Peak-to-Neighboring Power Ratio (PNPR) [2, 15, 16]: a spectral feature that determines the ratio between the power $\left|Y(e^{j\tilde{\omega}_k}, n)\right|^2$ of the candidate howling component and the power $\left|Y(e^{j(\tilde{\omega}_k+2\pi m/M)}, n)\right|^2$ of its mth neighbors frequency components, i.e,

$$\mathrm{PNPR}(\tilde{\omega}_k, n, m)\ [\mathrm{dB}] = 10\log_{10}\frac{\left|Y(e^{j\tilde{\omega}_k}, n)\right|^2}{\left|Y(e^{j(\tilde{\omega}_k+2\pi m/M)}, n)\right|^2}. \tag{2.38}$$

Voiced speech and tonal audio can be represented, in the time-domain, as damped sinusoids. In the frequency domain, they have non-zero bandwidth and their power is spread over several DFT bins around a spectral peak. On the other hand, a howling is, in the time domain, a pure sinusoid and its spectrum is supposed to be concentrated in a single DFT bin. Therefore, relatively large values for the PNPR feature are expected in howling components.

- Peakness [2, 15, 16]: the peakness feature reflects the time-averaged probability over 8 signal frames that the PNPR, averaged over 6 neighboring frequency bins on both sides of $\tilde{\omega}_k$ (excluding the closest neighbor on both sides), exceeds a 15 dB threshold, and is defined as

$$\begin{aligned}\mathrm{peakness}\,(\tilde{\omega}_k, n) = \sum_{j=0}^{7}\frac{1}{16}\Bigg\{&\left[\frac{1}{6}\sum_{m=2}^{7}\mathrm{PNPR}(\tilde{\omega}_k, n-jP, m) \geq 15\ \mathrm{dB}\right]\\ &+\left[\frac{1}{6}\sum_{m=-7}^{-2}\mathrm{PNPR}(\tilde{\omega}_k, n-jP, m) \geq 15\ \mathrm{dB}\right]\Bigg\}.\end{aligned} \tag{2.39}$$

5. Interframe Peak Magnitude Persistence (IPMP) [2, 15, 16, 31]: a temporal feature that, considering Q_M past frames, counts in how many frames the frequency $\tilde{\omega}_k$ is in the set of candidate howling components, and is defined as

$$\mathrm{IPMP}(\tilde{\omega}_k, n) = \frac{\sum_{j=0}^{Q_M-1}\left[\tilde{\omega}_k \in Y(e^{j\tilde{\omega}_k}, n-jP)\right]}{Q_M}. \tag{2.40}$$

The IPMP feature is based on that a howling component usually persists during a longer time than voiced speech and tonal audio components. Then, relatively large values for the IPMP feature are expected in howling components.

6. Interframe Magnitude Slope Deviation (IMSD) [2, 15, 16]: a temporal feature that determines the deviation over Q_M successive signal frames of a specific slope. First, an average difference (in dB scale) of the candidate howling component power in the $Q_M - 1$ most recent frames and the Q_M-th previous frame is performed. Second, an average difference (in dB scale) of the candidate howling component power in recent frames is carried out. Then, the slope is defined by the average difference of these two values as follows

$$\mathrm{IMSD}(\tilde{\omega}_k, n) = \frac{1}{Q_M - 1} \sum_{m=1}^{Q_M - 1} \left\{ \frac{1}{Q_M} \sum_{j=0}^{Q_M - 1} \frac{1}{Q_M - j} \left[20 \log_{10} \frac{\left| Y(e^{j\tilde{\omega}_k}, n - jP) \right|}{\left| Y(e^{j\tilde{\omega}_k}, n - Q_M P) \right|} \right] - \frac{1}{m} \sum_{j=0}^{m-1} \frac{1}{m - j} \left[20 \log_{10} \frac{\left| Y(e^{j\tilde{\omega}_k}, n - jP) \right|}{\left| Y(e^{j\tilde{\omega}_k}, n - mP) \right|} \right] \right\}. \quad (2.41)$$

Howling components have a nearly linear (in dB scale) increase in magnitude over time and thus their slope tends to be nearly constant. Thus, relatively small values for the IMSD feature are expected in howling components.

- Slopeness [15, 16]: the slopeness feature is a non-linear mapping of the IMSD feature which is not explicitly in the original proposal but in [15, 16] was defined as

$$\text{slopeness}\,(\tilde{\omega}_k, n) = e^{-|\mathrm{IMSD}(\tilde{\omega}_k, n)|}. \quad (2.42)$$

2.4.1.2 Detection Criteria

After their calculation, the values of the signal features are analyzed based on some criteria to classify each candidate howling component as a real howling component or not. Generally, the rule of the detection criteria is to compare the values of one or more signal features with pre-defined thresholds. Depending on the comparison results, a situation of howling is declared or not. The angular frequencies of the candidate howling components that are classified as real howling components form the set $D_{\tilde{\omega}_r} \subset D_{\tilde{\omega}_c}$.

The single-feature howling detection criteria found in literature are listed below:

1. PTPR criterion [2, 15, 16]:

$$\mathrm{PTPR}(\tilde{\omega}_k, n) \geq T_{\mathrm{PTPR}}\ [\mathrm{dB}] \Rightarrow \text{Howling detected}, \tilde{\omega}_k \in D_{\tilde{\omega}_r}(n). \quad (2.43)$$

2. PAPR criterion [2, 15, 16]:

$$\mathrm{PAPR}(\tilde{\omega}_k, n) \geq T_{\mathrm{PAPR}}\ [\mathrm{dB}] \Rightarrow \text{Howling detected}, \tilde{\omega}_k \in D_{\tilde{\omega}_r}(n). \tag{2.44}$$

3. PHPR criterion [2, 15, 16]:

$$\bigwedge_{m \in \mathcal{M}_{\mathrm{PHPR}}} \left[\mathrm{PHPR}(\tilde{\omega}_k, n, m) \geq T_{\mathrm{PHPR}}\ [\mathrm{dB}]\right] \Rightarrow \text{Howling detected}, \tilde{\omega}_k \in D_{\tilde{\omega}_r}(n), \tag{2.45}$$

where the symbol $\wedge$ denotes the logical conjunction operator.

4. PNPR criterion [2, 15, 16]:

$$\bigwedge_{m \in \mathcal{M}_{\mathrm{PNPR}}} \left[\mathrm{PNPR}(\tilde{\omega}_k, n, m) \geq T_{\mathrm{PNPR}}\ [\mathrm{dB}]\right] \Rightarrow \text{Howling detected}, \tilde{\omega}_k \in D_{\tilde{\omega}_r}(n) \tag{2.46}$$

5. IPMP criterion [2, 15, 16]:

$$\mathrm{IPMP}(\tilde{\omega}_k, n) \geq T_{\mathrm{IPMP}} \Rightarrow \text{Howling detected}, \tilde{\omega}_k \in D_{\tilde{\omega}_r}(n). \tag{2.47}$$

6. IMSD criterion [2, 15, 16]:

$$|\mathrm{IMSD}(\tilde{\omega}_k, n)| \leq T_{\mathrm{IMSD}}\ [\mathrm{dB}] \Rightarrow \text{Howling detected}, \tilde{\omega}_k \in D_{\tilde{\omega}_r}(n). \tag{2.48}$$

The signal features can be combined to achieve howling detection criteria that perform better than the single-feature ones. A very simple approach is to use logical conjunctions of single-feature howling detection criteria. The multiple-feature howling detection criteria found in literature are listed below:

1. Feedback existence probability (FEP) criterion [2, 15, 16, 30]:

$$\mathrm{FEP}\left(\tilde{\omega}_k, n\right) \geq T_{\mathrm{FEP}} \Rightarrow \text{Howling detected}, \tilde{\omega}_k \in D_{\tilde{\omega}_r}(n), \tag{2.49}$$

where the FEP feature is defined as

$$\mathrm{FEP}\left(\tilde{\omega}_k, n\right) = 0.7 \cdot \mathrm{slopeness}\left(\tilde{\omega}_k, n\right) + 0.3 \cdot \mathrm{peakness}\left(\tilde{\omega}_k, n\right). \tag{2.50}$$

2. PHPR & IPMP criterion [2, 15, 16]:

$$\left\{ \bigwedge_{m \in \mathcal{M}_{\text{PHPR}}} \left[\text{PHPR}(\tilde{\omega}_k, n, m) \geq T_{\text{PHPR}} \text{ [dB]} \right] \right\} \wedge \left\{ \text{IPMP}(\tilde{\omega}_k, n) \geq T_{\text{IPMP}} \text{ [dB]} \right\} \Rightarrow \text{Howling detected}, \tilde{\omega}_k \in D_{\tilde{\omega}_r}(n). \tag{2.51}$$

3. PHPR & PNPR criterion [15, 16]:

$$\left\{ \bigwedge_{m \in \mathcal{M}_{\text{PHPR}}} \left[\text{PHPR}(\tilde{\omega}_k, n, m) \geq T_{\text{PHPR}} \text{ [dB]} \right] \right\} \wedge \left\{ \bigwedge_{m \in \mathcal{M}_{\text{PNPR}}} \left[\text{PNPR}(\tilde{\omega}_k, n, m) \geq T_{\text{PNPR}} \text{ [dB]} \right] \right\} \Rightarrow \text{Howling detected}, \tilde{\omega}_k \in D_{\tilde{\omega}_r}(n). \tag{2.52}$$

4. PHPR & IMSD criterion [15, 16]:

$$\left\{ \bigwedge_{m \in \mathcal{M}_{\text{PHPR}}} \left[\text{PHPR}(\tilde{\omega}_k, n, m) \geq T_{\text{PHPR}} \text{ [dB]} \right] \right\} \wedge \left\{ |\text{IMSD}(\tilde{\omega}_k, n)| \leq T_{\text{IMSD}} \text{ [dB]} \right\} \Rightarrow \text{Howling detected}, \tilde{\omega}_k \in D_{\tilde{\omega}_r}(n). \tag{2.53}$$

5. PNPR & IMSD criterion [15, 16]:

$$\left\{ \bigwedge_{m \in \mathcal{M}_{\text{PNPR}}} \left[\text{PNPR}(\tilde{\omega}_k, n, m) \geq T_{\text{PNPR}} \text{ [dB]} \right] \right\} \wedge \left\{ |\text{IMSD}(\tilde{\omega}_k, n)| \leq T_{\text{IMSD}} \text{ [dB]} \right\} \Rightarrow \text{Howling detected}, \tilde{\omega}_k \in D_{\tilde{\omega}_r}(n). \tag{2.54}$$

6. PHPR & PNPR & IMSD criterion [15, 16]:

$$\left\{ \bigwedge_{m \in \mathcal{M}_{\text{PHPR}}} \left[\text{PHPR}(\tilde{\omega}_k, n, m) \geq T_{\text{PHPR}} \text{ [dB]} \right] \right\} \wedge \left\{ \bigwedge_{m \in \mathcal{M}_{\text{PNPR}}} \left[\text{PNPR}(\tilde{\omega}_k, n, m) \geq T_{\text{PNPR}} \text{ [dB]} \right] \right\} \wedge \left\{ |\text{IMSD}(\tilde{\omega}_k, n)| \leq T_{\text{IMSD}} \text{ [dB]} \right\} \Rightarrow \text{Howling detected}, \tilde{\omega}_k \in D_{\tilde{\omega}_r}(n). \tag{2.55}$$

After detecting the real howling components, the howling detection method should provide the set of design parameters $D_H(n)$ to the notch filter design stage. The set $D_H(n)$ should contain $D_{\tilde{\omega}_r}(n)$, the set of the angular frequencies of the real howling components, and $\left|Y(e^{j\omega}, n)\right|_{\omega \in D_{\tilde{\omega}_r}(n)}$, the magnitude values of the microphone signal spectrum at these frequency components.

2.4.2 Notch Filter Design

The second stage of the NHS methods designs a bank of notch filters in order to suppress the howling components and thus to maintain the closed-loop system stable. In NHS, the most used structure of digital notch filters is the second-order infinite impulse response (IIR) filter defined, for the kth howling component, as [2, 15, 16]

$$H_k(q, n) = \frac{b_{k,0}(n) + b_{k,1}(n)q^{-1} + b_{k,2}(n)q^{-2}}{1 + a_{k,1}(n)q^{-1} + a_{k,2}(n)q^{-2}}. \tag{2.56}$$

Thus, the bank of adjustable notch filters, which is inserted in the open-loop system as shown in Figure 2.8, is defined as a cascade of $N_H \leq N_p$ notch filters according to [2, 15, 16]

$$H(q, n) = \prod_{k=1}^{N_H} H_k(q, n). \tag{2.57}$$

The notch filter design receives, from the howling detection method, the set of design parameters $D_H(n)$ and converts it into a set of six filter specifications: the center frequency ω_c, the bandwidth B, the notch gain G_c, the gain at the band edges G_B, the gain at DC level G_0, and the gain at Nyquist frequency G_π. The latter two specifications can be fixed according to $G_0 = G_\pi = 0$ dB. Moreover, the gain at band edges may be defined as $G_B = G_c + 3$ dB in case of $G_c \leq -6$ dB, or as $G_B = G_c/2$ dB in case of $G_c \geq -6$ dB [2, 15, 16].

For the kth howling component, a notch filter with center frequency $\omega_{c,k}$ corresponding to the howling frequency should be designed and applied. Its notch gain $G_{c,k}$ can be calculated based on $\left|Y(e^{j\omega_{c,k}}, n)\right|$, the magnitude value of the microphone signal spectrum at the howling frequency. However, a common and simple approach is to work with fixed notch gain values that are independent of $\left|Y(e^{j\omega_{c,k}}, n)\right|$ [2, 15, 16]. When a new howling component is detected, a new notch filter is designed with an initial notch gain $G^0_{c,k}$, for example, $G^0_{c,k} = -3$ dB or $G^0_{c,k} = -6$ dB [2, 15, 16]. If the howling persists or occurs at a frequency close to a previously identified howling frequency, then the notch gain is decreased with $\Delta G_{c,k}$, for example, $\Delta G_{c,k} = -3$ dB or $\Delta G_{c,k} = -6$ dB [2, 15, 16]. The notch filter bandwidth B_k is usually chosen proportional to the center frequency in order to obtain a constant quality factor [2, 15, 16].

Aiming to complete the notch filter design, the set of filter specifications $\{\omega_{c,k}, B_k, G_{c,k}\}$ have to be translated to a set of filter coefficients $\{b_{k,0}(n), b_{k,1}(n), b_{k,2}(n), a_{k,1}(n), a_{k,2}(n)\}$.

To this end, some method should be applied as, for example, the bilinear transform of the notch filter transfer function or pole-zero placement techniques [2, 15, 16, 47].

2.4.3 Results of NHS Systems in the Literature

In this section, results available in the literature about the use of NHS in acoustic feedback control of PA systems will be presented. A study was carried out in [15, 16] about the efficiency of several howling detection criteria as a function of the values of their signal feature parameters and their decision thresholds. The performance of some NHS methods in terms of the increase in MSG and sound quality was analyzed in [2, 15, 16].

In [15, 16], an evaluation of the howling detection criteria described in Section 2.4.1.2 was performed by measuring their probabilities of detection and false alarm. As usual, for each frame of the microphone signal, N candidate howling components were selected from the spectrum magnitude $|Y(e^{j\omega}, n)|$ of the microphone signal by a peak algorithm. At the end of the signal, the total of N_T candidate howling components were obtained. In this procedure, it was assumed that the N_P frequencies components that really correspond to a howling (positive realizations) are known as well as the N_N frequency components that do not (negative realizations), where $N_T = N_P + N_N$.

Then, the probability of detection was defined as [15, 16]

$$P_D = \frac{N_{TP}}{N_P}, \tag{2.58}$$

where N_{TP} is the number of howling components that each method correctly detected (true positives). Similarly, the probability of false alarm was defined as [15, 16]

$$P_{FA} = \frac{N_{FP}}{N_N}, \tag{2.59}$$

where N_{FP} is the number of howling components that each method incorrectly detected (false positives).

In a PA system, high values of P_D are required in order to correctly remove the howling components and increase the MSG by activating appropriate notch filters. On the other hand, low values of P_{FA} are desired in order to not degrade the sound quality of the system signals by removing tonal components and prevent unnecessary activations of notch filters. The last observation is specifically important because the deactivation of notch filters is still an open problem in the NHS literature [15, 16]. Then once activated, a notch filter $H_k(q, n)$ remains activated until the end of the simulation affecting the sound quality and reducing the number of available notch filters that can be applied when a howling occurs. The trade-off between P_D and P_{FA} is controlled by the value of the detection threshold.

A classical approach to evaluate the performance of binary classifiers as a function of their discriminant threshold is to draw the receiver operating characteristic (ROC) curve [48]. The ROC corresponds to a P_D vs. P_{FA} curve where each point is obtained using

a different value of the discriminant threshold. For multiple-feature detection criteria, different ROCs should be drawn for each discriminant threshold.

ROC curves for the howling detection criteria described in Section 2.4.1.2 are shown in [15, 16]. For the same values of signal feature parameters and decision thresholds, the use of logical conjunctions of single-feature detection criteria results in a multiple-feature detection criterion that will not have P_D and P_{FA} higher than the corresponding single-feature detection criteria. Since a high P_D value is considered more important than a low P_{FA} value in terms of the overall performance of acoustic feedback control, multiple-feature detection criteria should combine single-feature detection criteria that present high P_D values regardless of their P_{FA} values [15, 16]. Some of the multiple-feature howling detection criteria described in Section 2.4.1.2 were proposed based on this idea.

Values of parameters and thresholds of several howling detection criteria that result in a minimum P_{FA} for $P_D = 95\%$ are provided in [15, 16]. In these experiments, the feedback path $F(q, n)$ was a measured room impulse response with duration of 100 ms and the forward path $G(q, n)$ was a broadband gain followed by a saturation function. The broadband gain was chosen slightly above the MSG of the PA system. The source signal $v(n)$ was an audio signal with duration of 10 s, $N = 3$, $N_P = 166$ and $N_N = 482$. A summary of the results is shown in Table 2.1.

It can be noticed that, except for the PHPR & IPMP criterion, the multiple-feature howling detection criteria achieved lower P_{FA} values than the single-feature ones. The PHPR & PNPR & IMSD and PNPR & IMSD criteria stood out by, for a $P_D = 95\%$, achieving P_{FA} equal to 3 and 5%, respectively. On the other hand, the PTPR and PAPR criteria obtained the worst results with $P_{FA} > 60\%$, which probably explains why they were not used on multiple-feature detection criteria.

With regard to the performance of NHS methods, average values of the achievable increase in MSG, ΔMSG, are presented in [2, 15, 16]. All these results were obtained in a simulated environment using the same configuration of the PA system but different howling detection criteria. The ΔMSG was mathematically calculated at each iteration which enabled display the values of the ΔMSG over time, $\Delta\text{MSG}(n)$. Considering a simulation runtime of T s, the feedback path $F(q, n)$ was a measured room impulse response until $t = {}^{3T}/_{4}$ s and then it was changed for other measured impulse response of the same room. The broadband gain of the forward path $G(q, n)$ was initialized to a value such that the PA system had an initial gain margin of 3 dB and remained at this value until $t = {}^{T}/_{4}$ s. During the next ${}^{T}/_{4}$ s, it was increased linearly (in dB scale) by 5 dB and remained at this value until the end of the simulation. It is noteworthy that the increase in the broadband gain achieved by the NHS methods (5 dB) was higher than that by the FS method (3 dB), described in Section 2.3.3, which indicates a superior performance of the NHS methods.

In [15], the NHS methods used only howling detection criteria capable of achieving a probability of detection $P_D > 65\%$ at a probability of false alarm as low as $P_{FA} = 1\%$. They were the FEP, PHPR & PNPR, PHPR & IMSD and PHPR & PNPR & IMSD.

Table 2.1: Comparison of P_{FA} values of several howling detection criteria for $P_D = 95\%$.

Detection criterion	Parameter and threshold values	P_{FA}
PTPR	$P_0 = 0$ dB, $T_{\text{PTPR}} = 34$ dB	70%
PAPR	$T_{\text{PAPR}} = 35$ dB	63%
PHPR	$\mathcal{M}_{\text{PHPR}} = \{2, 3\}$, $T_{\text{PHPR}} = 27$ dB	37%
PNPR	$\mathcal{M}_{\text{PNPR}} = \{\pm2, \pm3, \pm4\}$, $T_{\text{PNPR}} = 14$ dB	31%
IPMP	$Q_M = 20$, $T_{\text{IPMP}} = 0.3$	53%
IMSD	$Q_M = 32$, $T_{\text{IMSD}} = 0.25$ dB	40%
PHPR & IPMP	$\mathcal{M}_{\text{PHPR}} = \{0.5, 1.5, 2, 3, 4\}$, $T_{\text{PHPR}} = 10$ dB $Q_M = 5$, $T_{\text{IPMP}} = 0.4$	65%
FEP	$Q_M = 16$, $T_{\text{FEP}} = 0.7$	24%
PHPR & PNPR	$\mathcal{M}_{\text{PHPR}} = \{2, 3\}$, $T_{\text{PHPR}} = 30$ dB $\mathcal{M}_{\text{PNPR}} = \{\pm1, \pm2, \pm3, \pm4\}$, $T_{\text{PNPR}} = 6$ dB	14%
PHPR & IMSD	$\mathcal{M}_{\text{PHPR}} = \{2, 3\}$, $T_{\text{PHPR}} = 27$ dB $Q_M = 16$, $T_{\text{IMSD}} = 1$ dB	25%
PNPR & IMSD	$\mathcal{M}_{\text{PNPR}} = \{\pm2, \pm3, \pm4\}$, $T_{\text{PNPR}} = 12$ dB $Q_M = 16$, $T_{\text{IMSD}} = 0.5$ dB	5%
PHPR & PNPR & IMSD	$\mathcal{M}_{\text{PHPR}} = \{2, 3\}$, $T_{\text{PHPR}} = 23$ dB $\mathcal{M}_{\text{PNPR}} = \{\pm2, \pm3, \pm4\}$, $T_{\text{PNPR}} = 8$ dB $Q_M = 16$, $T_{\text{IMSD}} = 0.5$ dB	3%

The number of available notch filters was $N_H = 12$ and the source signal $v(n)$ was an audio signal with duration of $T = 60$ s. Table 2.2 shows the specific parameter and threshold values and, considering only the last $^T/_2$ s of simulation, the mean and maximum values of $\Delta\text{MSG}(n)$ obtained in each case. The results show an average ΔMSG(n) around 5 dB for all detection methods with a slight advantage to the PHPR & PNPR & IMSD. However, this method presented the worst sound quality because of its higher number of false alarms [15]. The NHS system based on the FEP criterion was the only one that presented some howling [15].

In [16], aiming to improve the performance of the NHS methods, the howling detection criteria were further restricted to those capable of achieving a probability of detection $P_D > 85\%$ (instead of 65% as in [15]) at a probability of false alarm as low as $P_{FA} = 1\%$ (same as in [15]). They were the FEP, PHPR & IMSD and PHPR & PNPR & IMSD. The PHPR & PNPR criterion, used in [15], was excluded. No information about the number

Table 2.2: Performance comparison of NHS systems with $P_D > 65\%$ and $P_{FA} = 1\%$.

Detection criterion	Parameter and threshold values	ΔMSG (dB)
FEP	$Q_M = 8$, $T_{\text{FEP}} = 0.9$	mean: 4.7 max: 6.2
PHPR & PNPR	$\mathcal{M}_{\text{PHPR}} = \{2, 3\}$, $T_{\text{PHPR}} = 42$ dB $\mathcal{M}_{\text{PNPR}} = \{\pm1, \pm2, \pm3, \pm4\}$, $T_{\text{PNPR}} = 6$ dB	mean: 4.8 max: 6.5
PHPR & IMSD	$\mathcal{M}_{\text{PHPR}} = \{2, 3\}$, $T_{\text{PHPR}} = 36$ dB $Q_M = 16$, $T_{\text{IMSD}} = 0.5$ dB	mean: 5 max: 5.8
PHPR & PNPR & IMSD	$\mathcal{M}_{\text{PHPR}} = \{2, 3\}$, $T_{\text{PHPR}} = 30$ dB $\mathcal{M}_{\text{PNPR}} = \{\pm1, \pm2, \pm3, \pm4\}$, $T_{\text{PNPR}} = 6$ dB $Q_M = 16$, $T_{\text{IMSD}} = 0.5$ dB	mean: 5.6 max: 6

Table 2.3: Performance comparison of NHS systems with $P_D > 85\%$ and $P_{FA} = 1\%$.

Detection criterion	Parameter and threshold values	ΔMSG (dB)
FEP	$Q_M = 16$, $T_{\text{FEP}} = 0.95$	mean: 6
PHPR & IMSD	$\mathcal{M}_{\text{PHPR}} = \{2, 3\}$, $T_{\text{PHPR}} = 42$ dB $Q_M = 16$, $T_{\text{IMSD}} = 1$ dB	mean: 5.8
PHPR & PNPR & IMSD	$\mathcal{M}_{\text{PHPR}} = \{2, 3\}$, $T_{\text{PHPR}} = 36$ dB $\mathcal{M}_{\text{PNPR}} = \{\pm2, \pm3, \pm4\}$, $T_{\text{PNPR}} = 12$ dB $Q_M = 16$, $T_{\text{IMSD}} = 0.1$ dB	mean: 6

N_H of available notch filters was provided and the source signal $v(n)$ was the same audio signal with duration of $T = 60$ s used in [15]. Probably, $N_H = 12$ as in [15] and as in the most cases in [2].

For each method, results were obtained with 3 different threshold values. Table 2.3 shows the mean values of $\Delta\text{MSG}(n)$, considering only the last $T/2$ s of simulation, achieved by the overall best performance of each method and the values of the corresponding parameters and thresholds. The results demonstrate an average ΔMSG around 6 dB for all detection methods with a slight disadvantage to the PHPR & IMSD, which also presented worse detection lag (resulting in longer instability intervals) and higher number of false alarms. Moreover, the results show that loose threshold values reduce the detection lag but increase the number of false alarms. On the other hand, strict threshold values decrease the number of false alarms but increase the detection lag. In both cases the resulting effect on the sound quality may be detrimental [16].

In [2], an evaluation of NHS methods using the howling detection criteria PHPR & IMSD, PAPR and FEP was carried out. The source signal $v(n)$ was an audio signal with $T = 60$ s and $f_s = 44.1$ kHz, the same used in [15, 16], and a speech signal with $T = 30$ s and $f_s = 16$ kHz. Table 2.4 summarizes the specific parameter and threshold values and, considering only the last $T/2$ s of simulation, the mean and maximum values of $\Delta\text{MSG}(n)$ obtained in each case. It is worth mentioning that a four times greater number of notch filters ($N_H = 48$) were made available for the PAPR method because of its higher probability of false alarm [2], which was already observed in [15, 16] and in Table 2.1.

When $v(n)$ was an audio signal, the PAPR achieved the best performance in terms of ΔMSG(n), obtaining a mean ΔMSG(n) of 7.1 dB, but by far the worst performance in terms of sound quality. Both results are explained by the higher number of notch filters available for the PAPR method. The PHPR & IMSD method achieved an average ΔMSG(n) of 5.7 dB but it also achieved a poor sound quality due to the high number of activated notch filters [2]. However, it is noteworthy that both PAPR and PHPR & IMSD methods obtained a high probability of false alarm, P_{FA}, in the evaluation of howling detection criteria performed in [15, 16] and whose results are shown in Table 2.1. The FEP method achieved a mean ΔMSG(n) of 4.8 dB, which is very close to that obtained in [15] and also close to that obtained [16] as can be observed in Tables 2.2 and 2.3. It also obtained the best results in terms of sound quality.

When $v(n)$ was a speech signal, all howling detection methods presented similar performances with respect to ΔMSG with a slight advantage to the FEP method, which achieved a mean $\Delta\text{MSG}(n)$ of 5 dB. It should be observed that, as regards $\Delta\text{MSG}(n)$, all detection methods performed worse when the source signal $v(n)$ was speech than when it was audio. In terms of sound quality, the FEP and PHPR & IMSD methods had similar performances and were slightly superior to the PAPR method. All detection methods performed, as regards sound quality, much better when the source signal $v(n)$ was speech than when it was audio and no comment was made about any specific problem in sound quality. The results reported in [2] are very interesting because they are the only published evaluation of NHS methods for speech signals as the source signal $v(n)$.

NHS methods based on the FEP howling detection criterion are among the most efficient methods, if not the best, considering all results presented in the NHS literature. Moreover, it achieves similar performances when the source signal $v(n)$ was audio or speech, although only 3 NHS methods had been evaluated for both natures of source signal. However, it is important to keep in mind that all presented results were obtained using only one signal (audio or speech), which is statistically insufficient to accurately infer the efficiency of any method.

Table 2.4: Performance comparison of NHS systems.

Signal	Detection criterion	Parameter and threshold values	ΔMSG (dB)
Speech	PHPR & IPMP	$\mathcal{M}_{\text{PHPR}} = \{0.5, 1.5, 2, 3, 4\}$, $T_{\text{PHPR}} = 30$ dB $Q_M = 5$, $T_{\text{IPMP}} = 0.6$ $N_H = 12$, $N_p = 3$, $M = 2048$, $P = 1024$	mean: 4.5 max: 5.2
	PAPR	$T_{\text{PAPR}} = 33$ dB $N_H = 48$, $N_p = 3$, $M = 2048$, $P = 1024$	mean: 4.5 max: 5.2
	FEP	$Q_M = 7$, $T_{\text{FEP}} = 0.7$ $N_H = 12$, $N_p = 3$, $M = 2048$, $P = 1024$	mean: 5 max: 5.6
Audio	PHPR & IPMP	$\mathcal{M}_{\text{PHPR}} = \{0.5, 1.5, 2, 3, 4\}$, $T_{\text{PHPR}} = 30$ dB $Q_M = 5$, $T_{\text{IPMP}} = 0.6$ $N_H = 12$, $N_p = 3$, $M = 4096$, $P = 2048$	mean: 5.7 max: 6.1
	PAPR	$T_{\text{PAPR}} = 55$ dB $N_H = 48$, $N_p = 3$, $M = 4096$, $P = 2048$	mean: 7.1 max: 8.6
	FEP	$Q_M = 7$, $T_{\text{FEP}} = 0.95$ $N_H = 12$, $N_p = 3$, $M = 4096$, $P = 2048$	mean: 4.8 max: 6

2.5 Conclusion

This chapter presented the problem of the acoustic feedback in a PA system. The acoustic feedback causes the PA system to have a closed-loop that, depending on the amplification gain, may become unstable resulting in a howling artifact, a phenomenon known as Larsen effect. This howling will be very annoying for all the audience and the amplification gain generally has to be reduced. As a consequence, the MSG of the PA system has a upper limit. Moreover, even if the MSG is not exceeded, the acoustic feedback causes the sound quality to be affected by excessive reverberation or ringing.

During the past years, several methods have been developed to eliminate or, at least, to control the Larsen effect. These methods can be divided in four main groups: phase-modulation, gain reduction, spatial filtering and room modeling methods. This chapter briefly described their main members and, then, addressed in detail the FS and NHS methods because they are the most widely used methods not only in literature but also as in commercial products and for historic reasons.

The FS method consists in shifting, at each loop, the spectrum of the microphone signal by a few Hz. It exploits the fact the average spacing between large peaks and adjacent valleys in the frequency response of large rooms is about 5 Hz. Then, in each loop, the spectrum of the microphone signal is shifted by a few cycles so that the frequency component responsible for the howling falls into a spectral valley of the feedback path after a few loops and, thus, is attenuated before the howling becomes audible. Increases up to 15 dB in the MSG due to the use of FS methods are reported in the literature but, in general, the subjectively acceptable increase is lower because of audible distortions.

The NHS method consists in detecting the candidate frequencies to generate instability and then apply notch filters in order to remove, or attenuate, these frequencies from the microphone signal. The major challenge of this method is to correctly accomplish these tasks before the howling becomes audible. The howling detection methods available in the literature were presented and briefly discussed. Increases up tp 8.6 dB in the MSG due to the use of NHS methods are reported.

However, in general, acoustic feedback control methods assume the existence of the acoustic feedback and only concern to control it. Moreover, they inevitably change not only the feedback signal but also the system input signal, which implies a fidelity loss of the PA system. This fidelity loss is undesired but may be neglected if the methods do not perceptually affect the sound quality, which is particularly difficult to achieve. Finally, they are not able to remove the excessive reverberation caused by the acoustic feedback.

Chapter 3

Acoustic Feedback Cancellation

3.1 Introduction

This chapter addresses the topic of acoustic feedback cancellation in PA systems. The AFC approach uses an adaptive filter to identify the acoustic feedback path and estimate the feedback signal, which is subtracted from the microphone signal. If the adaptive filter exactly matches the feedback path, the feedback signal would be completely removed from the microphone signal and thus the PA system would no longer have a closed-loop transfer function. As a consequence, the MSG would be infinite. In theory, the AFC approach offers a clear advantage over acoustic feedback control methods.

However, due to the amplification system, the system input and loudspeaker signals will be highly correlated, mainly when the source signal is colored as speech signal. Then, if the traditional gradient-based or least-squares-based adaptive filtering algorithms are used, a bias will be introduced adaptive filter coefficients. Hence, the adaptive filter will only partially cancel the feedback signal and also apply distortion to the system input signal. Therefore, in practice, the performance of the AFC approach is limited.

During the past years, several AFC methods have been developed to overcome the bias problem in AFC and an overview of them is presented in this chapter. The PEM-AFROW, which is the state-of-art method, is described in detail. It considers that the system input signal, which acts as noise to the estimation of the feedback path, is modeled by a filter whose input is white noise. Thus, the PEM-AFROW method prefilters the loudspeaker and microphone signals with the inverse source model, in order to create their whitened versions, before feeding them to a traditional adaptive filtering algorithm.

An evaluation of the PEM-AFROW method is carried out in a simulated environment using a measured room impulse response as the feedback path impulse response, a time-varying forward path broadband gain and two different ambient noise conditions. Its ability to estimate the feedback path and increase the MSG of a PA system is measured as well as the spectral degradations inserted in the microphone signal.

3.2 Acoustic Feedback Cancellation

As discussed in Chapter 2, among all methods developed to control the Larsen effect, the acoustic feedback cancellation (AFC) methods stand out for achieving the best overall performances. The AFC methods identify and track the acoustic feedback path $F(q,n)$ using an adaptive filter that is generally defined as an FIR filter

$$\begin{aligned} H(q,n) &= h_0(n) + h_1(n)q^{-1} + \ldots + h_{L_H-1}(n)q^{-(L_H-1)} \\ &= \mathbf{h}^T(n)\mathbf{q} \end{aligned} \tag{3.1}$$

with length L_H.

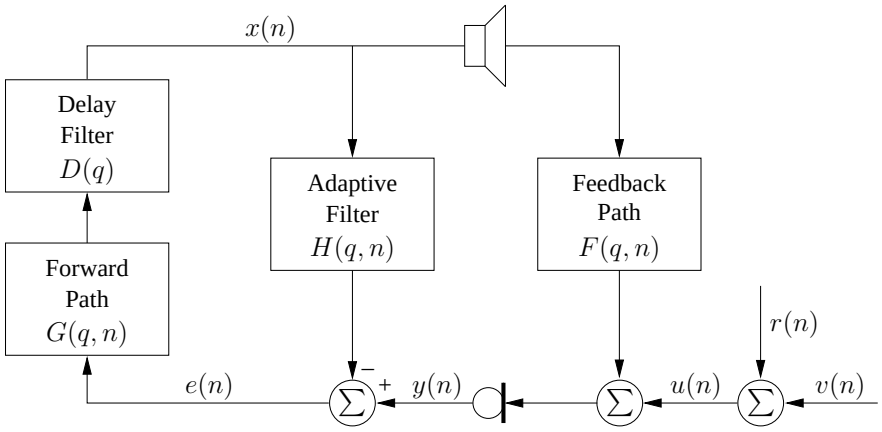


Figure 3.1: Acoustic feedback cancellation.

Then, an estimate of the feedback signal $\mathbf{f}(n) * x(n)$ is calculated as $\mathbf{h}(n) * x(n)$ and subtracted from the microphone signal $y(n)$, generating the error signal

$$\begin{aligned} e(n) &= y(n) - \mathbf{h}(n) * x(n) \\ &= u(n) + \mathbf{f}(n) * x(n) - \mathbf{h}(n) * x(n) \\ &= u(n) + [\mathbf{f}(n) - \mathbf{h}(n)] * x(n) \end{aligned} \tag{3.2}$$

which is effectively the signal fed to the forward path $G(q,n)$. Such a scheme is shown in Figure 3.1 [2, 3].

The closed-loop transfer function of a PA system with a AFC method, hereafter called AFC system, is defined as

$$\frac{x(n)}{u(n)} = \frac{G(q,n)D(q)}{1 - G(q,n)D(q)\left[F(q,n) - H(q,n)\right]} \tag{3.3}$$

and, according to the Nyquist's stability criterion, it is unstable if there is at least one frequency ω for which

$$\begin{cases} \left|G(e^{j\omega},n)D(e^{j\omega})\left[F(e^{j\omega},n)-H(e^{j\omega},n)\right]\right| \geq 1 \\ \angle G(e^{j\omega},n)D(e^{j\omega})\left[F(e^{j\omega},n)-H(e^{j\omega},n)\right] = 2k\pi,\ k\in\mathbb{Z}. \end{cases} \tag{3.4}$$

Then, considering the broadband gain $K(n)$ of the forward path defined in (2.8), the MSG of the AFC system is defined as [2]

$$\begin{gathered} \mathrm{MSG}(n)(\mathrm{dB}) = 20\log_{10}K(n) \\ \text{such that } \max_{\omega\in P_H(n)}\left|G(e^{j\omega},n)D(e^{j\omega})\left[F(e^{j\omega},n)-H(e^{j\omega},n)\right]\right| = 1, \end{gathered} \tag{3.5}$$

resulting in

$$\mathrm{MSG}(n)(\mathrm{dB}) = -20\log_{10}\left[\max_{\omega\in P_H(n)}\left|J(e^{j\omega},n)D(e^{j\omega})\left[F(e^{j\omega},n)-H(e^{j\omega},n)\right]\right|\right], \tag{3.6}$$

where $P_H(n)$ denotes the set of frequencies that fulfill the phase condition in (3.4), also called critical frequencies of the AFC system, that is

$$P_H(n) = \left\{\omega|\angle G(e^{j\omega},n)D(e^{j\omega})\left[F(e^{j\omega},n)-H(e^{j\omega},n)\right] = 2k\pi,\ k\in\mathbb{Z}\right\}. \tag{3.7}$$

The increase in the MSG achieved by the AFC method is defined as

$$\Delta\mathrm{MSG}(n)(\mathrm{dB}) = -20\log_{10}\left[\frac{\max_{\omega\in P_H(n)}\left|J(e^{j\omega},n)D(e^{j\omega})\left[F(e^{j\omega},n)-H(e^{j\omega},n)\right]\right|}{\max_{\omega\in P(n)}\left|J(e^{j\omega},n)D(e^{j\omega})F(e^{j\omega},n)\right|}\right]. \tag{3.8}$$

It should be noticed that the adaptive filter $H(q,n)$ must be initialized so that (3.4) is not fulfilled and thus the system closed-loop transfer function, defined in (3.3), is stable. Commonly, $H(q,n)$ is initialized with zeros, i.e. $H(q,0)=0$, in order to represent the lack of knowledge about the system to be identified, $F(q,n)$.

From (3.8), it can be concluded that the achievable ΔMSG increases as the match between the frequency responses of the adaptive filter and feedback path at the critical frequencies of the AFC system gets better. If $H(e^{j\omega},n) = F(e^{j\omega},n)$, $\forall\omega\in P_H(n)$, the MSG of the AFC system is infinite. However, in this case, some reverberation may still exist in the error signal $e(n)$ due to the frequency components that were not perfectly matched.

But if the adaptive filter exactly matches the feedback path, i.e. $H(q,n)=F(q,n)$, in addition to achieving an infinite MSG, it follows from (3.2) that the acoustic feedback will be totally cancelled because $e(n)=u(n)$. Hence, the system will no longer have a closed signal loop because (3.3) will become $x(n)=G(q,n)D(q)u(n)$, which means that only the system input signal $u(n)$ will be fed to the forward path $G(q,n)$, as desired.

The concept of AFC is very similar to acoustic echo cancellation (AEC) commonly used in teleconference systems [2, 3]. But in AFC, owing to the cascade $G(q,n)D(q)$, the system input signal $u(n)$ and the loudspeaker signal $x(n)$ are highly correlated, mainly when the source signal $v(n)$ is colored as speech. Since the system input signal $u(n)$ acts as interference to the adaptive filter $H(q,n)$, if the traditional gradient-based or least-squares-based adaptive filtering algorithms are used, a bias is introduced in $H(q,n)$ [2, 17, 18, 39, 40]. Consequently, the adaptive filter $H(q,n)$ only partially cancels the feedback signal $\mathbf{f}(n) * x(n)$, thereby achieving a limited increase in the MSG of the PA system, and also degrades the system input signal $u(n)$ [2, 4].

Mostly, the solutions available in the literature to overcome the bias in the adaptive filter $H(q,n)$ attempt to reduce the correlation between the loudspeaker signal $x(n)$ and system input signal $u(n)$ but still using the traditional adaptive filtering algorithms to update $H(q,n)$ [2, 49]. They can be divided in two main groups. The first group contains the methods that insert a processing device in the system open-loop in order to change the waveform of the loudspeaker signal $x(n)$. Even if the feedback signal is totally cancelled, this implies a fidelity loss of the PA system that, however, may be neglected if the added processing device does not perceptually affect the sound quality of the system, which is particularly difficult to achieve. The second group is formed by the methods that do not apply any processing to the signals that travel in the system other than the adaptive filter $H(q,n)$, and thereby keep the fidelity of the PA system as high as possible.

The AFC methods belonging to the first group can be divided in:

1. Noise injection [2, 49, 50, 51]: the AFC methods based on noise injection add a white signal $w(n)$ to the loudspeaker signal $x(n)$ such that

$$x(n) = G(q,n)D(q)e(n) + w(n). \tag{3.9}$$

   Then, the adaptive filter $H(q,n)$ can be updated in two ways. First, the loudspeaker signal $x(n)$ (including the added white noise $w(n)$) is used as the input signal to $H(q,n)$. In this case, the purpose of white noise signal $w(n)$ is to reduce the cross-correlation between the loudspeaker signal $x(n)$ and source signal $v(n)$ and, consequently, decrease the bias in $H(q,n)$. Second, only the white noise signal $w(n)$ is used as the input signal to $H(q,n)$ which leads to an unbiased estimate of the feedback path. But, in this case, the convergence of the adaptive filter will be rather slow because not only the system input signal $u(n)$ but also its component in feedback signal $\mathbf{f}(n) * x(n)$ will act as estimation noise to the adaptive filter $H(q,n)$. The drawback of noise injection is the degradation in the sound quality, which can be reduced by shaping the noise spectrum so that its effect is less perceptible to the human hearing. Unfortunately, the decorrelation effect caused by such shaped noises decreases making the noise injection less effective in reducing the bias in $H(q,n)$.

2. Time-varying processing [2, 49, 50, 52, 53]: the AFC methods based on time-varying processing insert an LPTV filter $L(q,n)$ in the system open-loop, as in Section 2.3, such that

$$x(n) = G(q,n)D(q)L(q,n)e(n). \tag{3.10}$$

Sinusoidal FM and PM, and FS filters have already been used to decorrelate the loudspeaker signal $x(n)$ and the source signal $v(n)$. The audible degradation on sound quality appear to be acceptable for speech signals but become more severe for audio signals. It is noteworthy that a beneficial effect of using LPTV filters as deccorelation filters is that they also contribute to stabilize the closed-loop system, as discussed in Section 2.3 for FS, by smoothing the open-loop gain.

3. Non-linear processing [2, 50]: in stereophonic AEC, the correlation between the loudspeaker signals leads to a bias in the estimate of the acoustic echo path, which can be reduced by adding to the loudspeaker signals nonlinearly processed versions of themselves [6, 22]. The same approach can be used to reduce the correlation between loudspeaker signal $x(n)$ and the system input signal $u(n)$ in an AFC system. In particular, the half-wave rectifier function has already been applied as follows

$$x(n) = G(q,n)D(q)\left[e(n) + \alpha\left(\frac{e(n)+|e(n)|}{2}\right)\right], \tag{3.11}$$

where α is the parameter that controls the amount of added nonlinearity and, consequently, the trade-off between decorrelation and audible signal distortions.

The AFC methods belonging to the second group can be divided in:

1. Forward path delay [2, 17, 18, 49]: the correlation between the loudspeaker signal $x(n)$ and the system input signal $u(n)$ can be reduced by the time delay caused by the cascade $G(q,n)D(q)$. Then, a very simple idea exploits the delay filter $D(q)$ in order to insert a delay of $L_D - 1$ samples in the cascade $G(q,n)D(q)$. This approach is particularly useful for source signals $v(n)$ that have an autocorrelation function that decays quickly, e.g., unvoiced segments of speech signals. If $u(n)$ is white noise, the cross-correlation vanishes with a unity delay. Moreover, the use of $D(q)$ as a decorrelation filter can be easily combined with any decorrelation approach. Obviously, the time delay must not impair the dynamics of real-time applications.

2. Cancellation path delay [2, 17, 18, 49, 54]: the idea of the previous item can be similarly implemented by inserting a delay filter $D_2(q)$, with length L_{D_2}, in the cancellation path so that the input signal to the adaptive filter $H(q,n)$ is the loudspeaker signal $x(n)$ delayed by $L_{D_2} - 1$ samples. In fact, it is the same to use an adaptive filter with length $L_H + L_{D_2} - 1$ samples where its first $L_{D_2} - 1$ samples are 0. Thus, if the cross-correlation function between the loudspeaker signal $x(n)$ and system input signal $u(n)$ has small values for time lags larger than $L_{D_2} - 1$, the

remaining bias in $H(q,n)$ may be small or even negligible. The advantage is that the loudspeaker signal $x(n)$ is not delayed, thereby keeping the fidelity of the PA system. In practice, this approach can be useful because $\mathbf{f}(n)$, as a room impulse response, is theoretically characterized by an initial delay determined by the distance between microphone and loudspeaker. So, if this initial delay is known *a priori*, the corresponding first coefficients of the adaptive filter can be forced to 0.

3. Whitening prefilters [2, 3, 4, 19, 39, 40, 41, 49]: consider that the system input signal $u(n)$, which acts as interference to the adaptive filter $H(q,n)$, is modeled by a filter $M(q,n)$, the source model, whose input is white noise, which fits quite well for unvoiced segments of speech signals. Thus, the bias in $H(q,n)$ can be completely eliminated by prefiltering the loudspeaker signal $x(n)$ and the microphone signal $y(n)$ with inverse source model $M^{-1}(q,n)$ before feeding them to the adaptive filtering algorithm [2, 19]. In [39, 40, 41], a fixed source model was used for hearing aid (HA) application. In [4], the prediction error method based adaptive feedback canceller (PEM-AFC) used an adaptive filter to estimate the source model also for HA application. In [2, 3], the prediction error method based on adaptive filtering with row operations (PEM-AFROW) improved the PEM-AFC and extended it for long acoustic paths by replacing the adaptive filter with the well-known Levinson-Durbin algorithm in the estimation of the source model. Moreover, after applying the inverse source model, the PEM-AFROW also removes the pitch components in order to improve the method performance for voiced speech [2, 3]. It should be noted that, when using the Levinson-Durbin algorithm, the PEM-AFROW method became suitable mostly for speech signals. For other kinds of signals, other source models should be used [55]. In [36], the PEM-AFROW was combined with a generalized sidelobe canceller but its performance did not improve for long feedback paths, such as occur in PA systems, although its computational complexity was reduced.

The PEM-AFROW method is the state-of-art AFC method. Therefore, this chapter presents a careful review of the PEM-AFROW, including a brief discussion about the PEM framework and the PEM-AFC method, and simulation results.

3.3 The PEM-AFROW Method

The prediction error method (PEM) assumes that the system input signal $u(n)$ is defined as [4, 23, 40, 56]

$$u(n) = M(q,n)w(n), \tag{3.12}$$

where the excitation $w(n)$ is white noise and the source model $M(q,n)$ is monic and

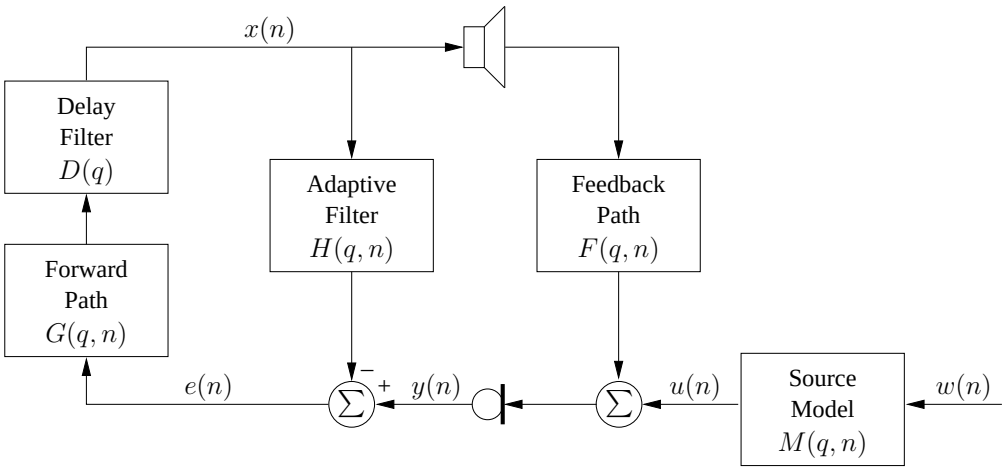


Figure 3.2: Acoustic feedback cancellation using source model.

inversely stable. Such a scheme is shown in Figure 3.2. Speech and audio signals can be closely approximated as a low-order autoregressive random process such that [57, 58]

$$M(q,n) = \frac{1}{A(q,n)} = \frac{1}{1 - a_1(n)q^{-1} - \ldots - a_{L_A-1}(n)q^{-L_A+1}}, \tag{3.13}$$

except for periodic signals such as voiced speech or pure sinusoids, where the excitation $w(n)$ is an impulse train [58].

Then, the open-loop system $\{F(q,n), M(q,n)\}$ to be identified is described by the microphone signal [4, 23, 40, 56]

$$y(n) = F(q,n)x(n) + M(q,n)w(n). \tag{3.14}$$

According to [4, 40, 56], the PEM produces estimates of $F(q,n)$ and $M(q,n)$, $H(q,n)$ and $\hat{M}(q,n)$ respectively, by minimizing the difference between the microphone signal $y(n)$ and its optimal one-step predictor with model $\{H(q,n), \hat{M}(q,n)\}$

$$\hat{y}(n) = \hat{M}^{-1}(q,n)H(q,n)x(n) + \left[1 - \hat{M}^{-1}(q,n)\right] y(n), \tag{3.15}$$

which is defined as the prediction error

$$e_p(n) = y(n) - \hat{y}(n) = M^{-1}(q,n)\left[y(n) - H(q,n)x(n)\right]. \tag{3.16}$$

As in any AFC method, $H(q,n)$ is estimated over time using an FIR adaptive filter. In [39, 40, 41], the estimate $\hat{M}(q,n)$ of the source model was a fixed low-pass filter that approximates the long-term average spectrum of speech. In practice, however, $M(q,n)$ is unknown and time-varying [4]. And the accuracy of the estimate $H(q,n)$ of the feedback path strongly depends on the accuracy of the estimate $\hat{M}(q,n)$ of the source model [4, 40]. Therefore, it is also desirable to estimate $M(q,n)$ over time [4].

However, in general, $F(q,n)$ and $M(q,n)$ are simultaneously identifiable using only measurements of the loudspeaker signal $x(n)$ and the microphone signal $y(n)$ if $G(q,n)D(q)$ has a delay $d_1 > L_A$, $M^{-1}(q,n)F(q,n)$ has a delay d_2 with $d_1+d_2 > L_A$, $G(q,n)D(q)$ and $H(q,n)$ are time-varying, $G(q,n)$ is nonlinear or a probe signal is added to $x(n)$ [4]. In the latter, the identifiability will depend on the level of the probe signal compared with the level of the loudspeaker signal $x(n)$ [4]. But, in all cases, the excitation $w(n)$ must be white noise. Otherwise, $F(q,n)$ and $M(q,n)$ are not identifiable which implies that, besides the desired solutions $H(q,n) = F(q,n)$ and $\hat{M}(q,n) = M(q,n)$, multiple solutions for $H(q,n)$ and $\hat{M}(q,n)$ may exist [4, 40]. This non-identifiability problem is due to the linear relationship between the loudspeaker signal $x(n)$ and the system input signal $u(n)$ caused by the cascade $G(q,n)D(q)$ [4].

In [4], the prediction error method based adaptive feedback canceller (PEM-AFC) exploits the delay filter $D(q)$ by making $L_D - 1 > L_A$ in order to overcome the non-identifiability problem and uses a second FIR adaptive filter to estimate $\hat{M}^{-1}(q,n)$ at each iteration [4]. However, the PEM-AFC considers that

$$M^{-1}(q,n-1) = M^{-1}(q,n-n_1),\ 1 \leq n_1 \leq L_H, \tag{3.17}$$

which represents the stationarity of the input signal $u(n)$ over frames of L_H samples. If $u(n)$ is speech, this approximation may be valid for short acoustic feedback paths $F(q,n)$ where ${}^{L_F}/_{fs} \leq 20$ ms, such as occur in HA applications, because speech is considered stationary during short frames with duration of about 20 ms [58]. But for long acoustic feedback paths $F(q,n)$, such as occur in PA systems where ${}^{L_F}/_{fs} \geq 100$ ms, this approximation is no longer valid because speech is highly nonstationary over long time periods.

Nevertheless, for speech signals, the estimation of the inverse source model $M^{-1}(q,n)$ is a very established technique in speech coding. It combines two prediction error filters in a cascade connection according to [59]

$$\begin{aligned} M^{-1}(q,n) &= A(q,n)B(q,n) \\ &= \left[1 - a_1(n)q^{-1} - \ldots - a_{L_A-1}(n)q^{-L_A+1}\right]\left[1 - b_{L_B-1}(n)q^{-L_B+1}\right]. \end{aligned} \tag{3.18}$$

The first filter $A(q,n)$ (called formant filter or short-time prediction filter) models the vocal tract and removes near-sample redundancies, and is computed using the well-known Levinson-Durbin algorithm [58, 59]. The second filter $B(q,n)$ (called pitch filter or long-time prediction filter) models the periodicity and acts on distant-sample waveform similarities, and is usually an one-tap filter with lag equal to the pitch period [59].

In [3], the prediction error method based on adaptive filtering with row operations (PEM-AFROW) uses the Levinson-Durbin algorithm, instead of an adaptive filter as in the PEM-AFC, to compute $A(q,n)$ over short frames. Hence, it considers that, inside a frame, the system input signal $u(n)$ is stationary and thus $A(q,n)$ is constant. In addition, the PEM-AFROW method also computes $B(q,n)$ to remove the pitch components in order to

improve the method performance for voiced speech, when $w(n)$ is periodic and not white noise as originally assumed by the PEM [2, 3].

Therewith, the PEM-AFROW method extended the PEM-AFC method for long acoustic feedback paths and also improved it for short ones [3]. However, because of using the Levinson-Durbin algorithm to compute $A(q,n)$, the PEM-AFROW became suitable mostly for speech signals. For other kinds of signals, other source models should be used [15].

In the following sections, the PEM-AFROW method will be described in detail. For a better understanding, the method will be divided in three parts: whitening of the system signals, update of the adaptive filter using whitened signals and feedback cancellation.

3.3.1 Part 1: Whitening of the System Signals

The first part of the method is responsible for estimating the inverse source model $M^{-1}(q,n)$ and whitening the system signals. It is the core of the PEM-AFROW method. The first prediction error filter $A(q,n)$, the short-time predictor, is estimated using a non-overlapping frame with length L_{stp} samples, which means that it is estimated every L_{stp} samples. Defining k as the short-time frame index such that it is the first integer higher than or equal to n/L_{stp}, the current frame (frame: k; sample indexes: $kL_{stp},\ldots,(k+1)L_{stp}-1$) of the loudspeaker signal $x(n)$ is filtered by $\mathbf{h}(kL_{stp}-1)$, the last estimate of the feedback path obtained in the previous frame $(k-1)$, and subtracted from the corresponding microphone samples resulting in

$$d(n) = y(n) - \mathbf{x}^T(n)\mathbf{h}(kL_{stp}-1),\ n = kL_{stp},\ldots,(k+1)L_{stp}-1, \tag{3.19}$$

where

$$\mathbf{x}(n) = [x(n)\ \ x(n-1)\ \ \ldots\ \ x(n-L_H+1)]^T. \tag{3.20}$$

Note that it was assumed that the current frame of the loudspeaker signal $x(n)$ does not depend on the current frame of the microphone signal $y(n)$. For that, the cascade $G(q,n)D(q)$ must have a delay $L_D - 1 > L_{stp}$. This assumption is not mentioned in [3]. Moreover, it is worth mentioning that if the last estimate of the feedback path is exact, i.e. $\mathbf{h}(kL_{stp}-1) = \mathbf{f}(n)$, then the current frame of the signal $d(n)$ will be equal to the current frame of the system input signal $u(n)$, i.e., $d(n) = u(n)$, $n = kL_{stp},\ldots,(k+1)L_{stp}-1$.

Hereupon, the short-time prediction filter for the frame k, $A_k(q)$, is computed by performing linear prediction on $d(n)$, $n = kL_{stp},\ldots,(k+1)L_{stp}-1$, using the Levinson-Durbin algorithm. And the short-time whitened loudspeaker and microphone signals are obtained as

$$\mathbf{x}_{sw}^T(n) = \mathbf{a}_k^T \begin{bmatrix} \mathbf{x}^T(n) \\ \mathbf{x}^T(n-1) \\ \vdots \\ \mathbf{x}^T(n-L_A) \end{bmatrix} \tag{3.21}$$

and

$$y_{sw}(n) = \mathbf{a}_k^T \begin{bmatrix} y(n) \\ y(n-1) \\ \vdots \\ y(n-L_A) \end{bmatrix}, \tag{3.22}$$

respectively, where $\mathbf{x}(n)$ is defined in (3.20) and

$$\mathbf{x}_{sw}(n) = [x_{sw}(n) \;\; x_{sw}(n-1) \;\; \ldots \;\; x_{sw}(n-L_H+1)]^T. \tag{3.23}$$

It should be noted that the short-time prediction filter $A_k(q)$ is used to update not only the samples of the frame k of the short-time whitened loudspeaker signal $x_{sw}(n)$ (the most recent L_{stp} samples) but, indeed, its last L_H samples. But, since $\mathbf{x}(n)$ is a shifted version of $\mathbf{x}(n-1)$ with one sample prepended and $\mathbf{a}_k$ remains constant during a frame of L_{stp} samples, $\mathbf{x}_{sw}(n)$ will be a shifted version of $\mathbf{x}_{sw}(n-1)$ with one sample prepended. Then, inside a frame, only one vector multiplication ($\mathbf{a}_k^T \mathbf{x}^T(n)$) has to be performed to calculate $\mathbf{x}_{sw}(n)$. However, at the beginning of each frame, one matrix multiplication should be performed according to (3.21) to calculate all the samples of $\mathbf{x}_{sw}(n)$.

In theory, if $w(n)$ is white noise, the short-time prediction filter $A(q,n)$ will remove all the correlation between the loudspeaker signal $x(n)$ and the system input signal $u(n)$, which is included in the microphone signal $y(n)$. In practice, however, although it will remove most of the correlation, the short-time whitened loudspeaker signal $x_{sw}(n)$ and the short-time whitened input signal, that is included in $y_{sw}(n)$, are still correlated mainly for voiced speech when $w(n)$ is periodic. Then, in order to improve the whitening performance, the short-time prediction filter $A(q,n)$ is followed by the long-time prediction filter $B(q,n)$.

The second prediction error filter $B(q,n)$, the long-time predictor, is estimated using frames with length L_{stp} samples and 50% overlap, which means that it is estimated every $L_{ltp} = {}^{L_{stp}}/_{2}$ samples. Defining j as the long-time frame index such that it is the first integer higher than or equal to n/L_{ltp}, the long-time prediction filter $B_j(q)$, for the frame j, is computed by minimizing [3]

$$\begin{aligned} \epsilon_j &= \min \mathrm{E}\left[\|x_{lw}(n)\|^2\right] \\ &= \min_{\{L_{B_j}, b_j\}} \mathrm{E}\left[\left\|x_{sw}(n) - b_j x_{sw}(n-L_{B_j}+1)\right\|^2\right], \end{aligned} \tag{3.24}$$

where $\mathrm{E}\{\cdot\}$ is the expected value operator. For a fixed L_{B_j}, the solution to (3.24) is [3]

$$b_j = \left[\tilde{\mathbf{x}}_{sw}^T(n-L_{B_j})\tilde{\mathbf{x}}_{sw}(n-L_{B_j})\right]^{-1} \tilde{\mathbf{x}}_{sw}^T(n)\tilde{\mathbf{x}}_{sw}(n-L_{B_j}), \tag{3.25}$$

where

$$\tilde{\mathbf{x}}_{sw}(n) = [x_{sw}(n) \;\; x_{sw}(n-1) \;\; \ldots \;\; x_{sw}(n-L_{stp}+1)]^T. \tag{3.26}$$

The variance of the long-time prediction residual is [3]

$$\epsilon_j = \tilde{\mathbf{x}}_{sw}^T(n)\tilde{\mathbf{x}}_{sw}(n) - \frac{\left[\tilde{\mathbf{x}}_{sw}^T(n)\tilde{\mathbf{x}}_{sw}(n-L_{B_j})\right]^2}{\tilde{\mathbf{x}}_{sw}^T(n-L_{B_j})\tilde{\mathbf{x}}_{sw}(n-L_{B_j})}. \tag{3.27}$$

This variance ϵ_j is evaluated for different values of L_{B_j}, where $L_{B_j} = L_{B\min}, L_{B\min}+1, \ldots, L_{B\max}$. The value of of L_{B_j} that results in the minimum ϵ_j and the corresponding b_j are chosen for the long-time predictor $B_j(q)$ of the frame j. Finally, the long-time whitened loudspeaker and microphone signals are obtained as

$$\mathbf{x}_{lw}^T(n) = \mathbf{b}_j^T \begin{bmatrix} \mathbf{x}_{sw}^T(n) \\ \mathbf{x}_{sw}^T(n-1) \\ \vdots \\ \mathbf{x}_{sw}^T(n-L_{B_j}) \end{bmatrix} \tag{3.28}$$

and

$$y_{lw}(n) = \mathbf{b}_j^T \begin{bmatrix} y_{sw}(n) \\ y_{sw}(n-1) \\ \vdots \\ y_{sw}(n-L_{B_j}) \end{bmatrix}, \tag{3.29}$$

respectively, where $\mathbf{x}_{sw}(n)$ is defined in (3.23) and

$$\mathbf{x}_{lw}(n) = \left[x_{lw}(n) \;\; x_{lw}(n-1) \;\; \ldots \;\; x_{lw}(n-L_H+1)\right]^T. \tag{3.30}$$

Similarly to $A_k(q)$, the long-time prediction filter $B_j(q)$ is used to update not only the samples of the frame j of the long-time whitened loudspeaker signal $x_{lw}(n)$ (the most recent L_{stp} samples) but, indeed, its last L_H samples. But, since $\mathbf{x}_{sw}(n)$ is a shifted version of $\mathbf{x}_{sw}(n-1)$ with one sample prepended and $\mathbf{b}_j$ remains constant during a frame of L_{stp} samples, $\mathbf{x}_{lw}(n)$ will also be a shifted version of $\mathbf{x}_{lw}(n-1)$ with one sample prepended. Then, inside a frame, only one vector multiplication ($\mathbf{b}_j^T\mathbf{x}_{sw}^T(n)$) has to be performed to calculate $\mathbf{x}_{lw}(n)$. However, at the beginning of each frame, one matrix multiplication should be performed according to (3.28) to calculate all the samples of $\mathbf{x}_{lw}(n)$.

It should be emphasized that, because of $B_j(q)$, the identifiability condition of the PEM-AFROW method is the existence of a delay $d_1 > L_A + L_{B\max}$ samples in the cascade $D(q)G(q,n)$ [3]. For the PEM-AFC, the condition is a delay $d_1 > L_A$ samples as previously discussed. The PEM-AFROW method exploits the delay filter $D(q)$ by making $L_D > L_A + L_{B\max}$ samples in order to overcome the non-identifiability problem.

Furthermore, for real-time implementation, the PEM-AFROW method involves a delay of one frame (L_{stp} samples) in updating the adaptive filter $H(q,n)$ because $\mathbf{a}_k$, the coefficients of the short-time prediction filter for the frame k, can only be calculated at time $n = (k+1)L_{stp} - 1$. This delay can be effectively implemented as a delay line for

the samples of the loudspeaker signal $x(n)$ before they are fed to (3.21) [3]. The practical influence of this latency will depend on the variations of $F(q,n)$ over time.

3.3.2 Part 2: Update of the Adaptive Filter using Whitened Signals

After obtaining the current frame of both long-time whitened loudspeaker signal $x_{lw}(n)$ and long-time whitened microphone signal $y_{lw}(n)$, the impulse response $\mathbf{h}(n)$ of the adaptive filter is update by solving

$$\min_{\mathbf{h}} \|e_{lw}(n)\| = \min_{\mathbf{h}} \left\| y_{lw}(n) - \mathbf{h}^T(n)\mathbf{x}_{lw}(n) \right\|, \; n = kL_{stp}, \ldots, (k+1)L_{stp} - 1, \tag{3.31}$$

using the NLMS adaptive filtering algorithm, where $\mathbf{x}_{lw}(n)$, which is defined in (3.30), is the input vector and $y_{lw}(n)$ is the desired sample. Consequently, an estimate $\mathbf{h}(n)$ of the feedback path is obtained for each of the L_{stp} samples of the current frame.

3.3.3 Part 3: Feedback Cancellation

After obtaining an estimate $\mathbf{h}(n)$ of the feedback path for each of the L_{stp} samples of the current frame, the actual values of the error signal $e(n)$ are obtained according to

$$e(n) = y(n) - \mathbf{x}^T(n)\mathbf{h}(n), \; n = kL_{stp}, \ldots, (k+1)L_{stp} - 1 \tag{3.32}$$

and then are fed to the forward path $G(q,n)$.

3.4 Improvements in PEM-AFROW

This section will present some improvements to the PEM-AFROW method that were proposed in [42]. However, these improvements are not specific changes of the PEM-AFROW. In fact, they are originated from concepts of adaptive filtering and the acoustic feedback problem, and can be applied to any AFC method.

3.4.1 Onset Detection

In AEC, if the source signal $v(n)$ and the ambient nosie $r(n)$ are zero and negligible, respectively, the adaptive filter $H(q,n)$ can converge to a good estimate of the path $F(q,n)$ and thus cancel the echo successfully. However, when the source signal $v(n)$ and the loudspeaker signal $x(n)$ are simultaneously different from zero, a situation known as double-talk in AEC, $v(n)$ acts as an uncorrelated noise to $H(q,n)$ and suddenly increases the amplitude of the microphone signal$y(n)$ and error signal $e(n)$. Since $e(n)$ is used by the traditional adaptive filtering algorithms to update the adaptive filter, this may disturb the filter update causing an excessive mismatch of $H(q,n)$ or, even, its divergence. The usual solution to this problem is to decrease or stop completely the filter update when the presence of $v(n)$ is detected. This is the rule of double-talk detectors (DTD). Besides, a

voice activity detector (VAD) is also used to stop the filter update when the energy of the loudspeaker signal $x(n)$ is below a pre-defined noise level.

In AFC, the VAD for the loudspeaker signal $x(n)$ is also required but the DTD is not anymore because the closed signal loop causes the PA system to be in a continuous double-talk situation. However, the system input signal $u(n)$ acts as estimation noise to the adaptive filter $H(q,n)$ and thus introduces a bias in its coefficients that is inversely related to the far-end to near-end ratio defined as

$$\mathrm{FNR} = \frac{E[x^2(n)]}{E[u^2(n)]}. \tag{3.33}$$

Fortunately, high values of FNR are obtained by means of high gains in the forward path $G(q,n)$, which is useful because it is the situation when the Larsen effect generally occurs. However, at an input signal onset (a sudden level increase of $u(n)$), the FNR is temporarily very small because the corresponding level increase in the loudspeaker signal $x(n)$ is delayed by the cascade $D(q)G(q,n)$ [42]. Hence, input signal onsets may cause an excessive mismatch of the adaptive filter $H(q,n)$ and instability of the whole system.

For this reason, an onset detection method based on the variance of the long-time whitened error signal $e_{lw}(n)$ was proposed in [42]. Note that $e_{lw}(n)$ is the whitened input signal $w(n)$ if $H(q,n)$ exactly models $F(q,n)$. The variance of $e_{lw}(n)$ is estimated, at every time, over an exponential window according to

$$\sigma^2_{e_{lw}}(n) = \lambda\sigma^2_{e_{lw}}(n-1) + (1-\lambda)e^2_{lw}(n), \tag{3.34}$$

where $0 \ll \lambda < 1$ is a forgetting factor. The onset detector rules similarly to a DTD: an onset is detected when $|e_{lw}(n)|$ is greater than a threshold T_{osd} and, if it occurs, the filter update is stopped during a time interval Δt_{osd}. A conservative value for Δt_{osd} is the sum of the forward delays and the number of filter coefficients.

In [42], the onset detection method and the PEM-AFROW method were combined and an evaluation was carried out in an simulation environment using white noise as the source signal $v(n)$. The results indicate an achievable increase in the step-size of the adaptive filter, which corresponds to a faster convergence and tracking speed, without audible instabilities.

3.4.2 Prior Knowledge of the Feedback Path

As previously discussed in Section 3.2, the adaptive filter $H(q,n)$ must be initialized such that (3.4) is not fulfilled and thus the closed-loop transfer function defined in (3.3) is stable. Commonly, $H(q,n)$ is initialized with zeros, i.e., $H(q,0) = 0$ in order to represent the lack of knowledge about the system to be identified, the feedback path $F(q,n)$. However, if a good estimate $\hat{F}(q,0)$ of the feedback path is known, it can be used as the initial guess of the adaptive filter, i.e., $H(q,0) = \hat{F}(q,0)$.

Moreover, in order to provide robustness, a change in the cost function of the adaptive filtering algorithm in the PEM-AFROW was proposed in [42] to incorporate the prior knowledge. Instead of (3.31), the impulse response $\mathbf{h}(n)$ of the adaptive filter can be updated by solving

$$\min_{\mathbf{h}} \left\{ \left\| y_{lw}(n) - \mathbf{h}^T(n)\mathbf{x}_{lw}(n) \right\| + \beta \left\| \mathbf{h}(n) - \hat{\mathbf{f}}(0) \right\| \right\} \tag{3.35}$$

using the NLMS adaptive filtering algorithm, where β is a parameter that controls the weight of the prior knowledge of the feedback path.

In practice, a time-varying $\beta(n)$ is suggested in [42] so that its value can be high at the start-up and then decrease gradually over time. According to [42], although no results are shown, experiments demonstrate that updating the adaptive filter using (3.35) can be especially useful at the start-up of the PEM-AFROW method.

3.4.3 Foreground and Background Filter

An adaptive filter with small step-size generally provides robustness against noise but slow convergence. On the other hand, an adaptive filter with large step-size usually presents fast convergence and track ability but it can suffer from instability. Then, in order to combine the strength of both cases, a twin adaptive filter structure is proposed in [42].

This idea was firstly proposed to AEC and consists in estimating the feedback path $F(q,n)$ through two adaptive filters with different convergence speeds. The foreground filter $H(q,n)$ has a small step-size and is responsible for the conservative solution of the system. The background filter $H_b(q,n)$ has a large step-size and is responsible for the fast tracking of variations in the feedback path impulse response.

The variance of the system input signal $u(n)$ is estimated to the foreground $H(q,n)$ and background $H_b(z,n)$ filters according to, respectively, (3.34) and

$$\sigma^2_{e_{lw,b}}(n) = \lambda\sigma^2_{e_{lw,b}}(n-1) + (1-\lambda)e^2_{lw,b}(n), \tag{3.36}$$

where $0 \ll \lambda < 1$ is a forgetting factor.

At time intervals these estimates are compared and if

$$\sigma^2_{e_{w,b}}(n) < \gamma_1\sigma^2_{e_w}(n), \tag{3.37}$$

the coefficients $\mathbf{h_b}(n)$ of the background filter are copied to the coefficients $\mathbf{h}(n)$ of the foreground filter in order to improve the system performance. On the other hand, if

$$\sigma^2_{e_{w,b}}(n) > \gamma_2\sigma^2_{e_w}(n), \tag{3.38}$$

the impulse response $\mathbf{h}(n)$ of the foreground filter is copied to the impulse response $\mathbf{h_b}(n)$ of the background filter aiming to avoid divergence of $\mathbf{h_b}(n)$. In this configuration, $0 <$

$\gamma_{1,2} < 1$ and $\gamma_1 \leq \gamma_2$.

In [42], an evaluation of this twin adaptive filter structure was carried out in the same simulated environment of the onset detection method. The results indicate that there is no audible distortion when the background filter $H_b(q,n)$ has $\mu = 0.9$ and the foreground filter $H(q,n)$ has $\mu = 0.09$. On the other hand, when a single adaptive filter $H(q,n)$ with $\mu = 0.09$ is used, an audible transient is clearly perceived.

3.4.4 Proactive Notch Filtering

An important characteristic of an AFC method is its ability to quickly track the variations in the acoustic feedback path $F(q,n)$. If an AFC method is not able to do it, the sound quality may be perceptually affected and the PA system may even become unstable.

In [42], it is stated that the PEM-AFROW method is very robust against variations in the feedback path caused by moving objects as speaker movements. However, when the position of the microphone or loudspeaker is changed, the feedback path impulse response shifts over the time axis and the PEM-AFROW cannot track the resulting variations quickly enough. This is due to fact that the difference between shifted room impulse responses has the same order of magnitude than the impulse responses themselves and, therefore, the PEM-AFROW maye need a considerable amount of time to compensate the difference. On the other hand, the frequency component of many peaks in the system open-loop frequency response does not change much when the feedback path impulse response is shifted [42], which may indicate that a notch-filtering-based approach can be more robust against displacement of the microphone or loudspeaker. As a consequence, in this case, the PEM-AFROW algorithm is not so robust as the NHS methods.

Hence, in order to provide robustness, a combination of the PEM-AFROW with a proactive notch filtering system was proposed in [42]. Considering the estimate $H(q,n)$ of the feedback path provided by the PEM-AFROW method and the knowledge of the cascade $D(q)G(q,n)$, the open-loop frequency response of the PA system is estimated as $G(e^{j\omega},n)D(e^{j\omega})H(e^{j\omega},n)$ and the set $P(n)$ of critical frequencies is computed from it.

Then, a notch filter with center frequency $\omega_c \in P(n)$ is designed if

$$\left|G(e^{j\omega_c},n)D(e^{j\omega_c})H(e^{j\omega_c},n)\right| > T_{max}, \tag{3.39}$$

or, if already exists, is removed if

$$\left|G(e^{j\omega_c},n)D(e^{j\omega_c})H(e^{j\omega_c},n)\right| < T_{min}, \tag{3.40}$$

where the thresholds $0 < T_{min} < T_{max} < 1$. Since $T_{max} < 1$, the described procedure leads to a proactive notch filtering because the notch filters are designed before the system becomes unstable at the corresponding frequencies. The notch filter design is repeated few

times per second and the maximum number of notch filters is limited. Finally, the bank of notch filters is inserted in the system open-loop immediately after the microphone [42].

It is worth mentioning that, when inserting the notch filters in this position, the PEM-AFROW method should model not only the feedback path $F(q,n)$ but also the notch filters. At first sight, this might seem unfavorable. However, if the notch filers are inserted after the subtraction of the estimate of the feedback signal, $\mathbf{h}(n) * x(n)$, from the microphone signal $y(n)$ so that the PEM-AFROW should not model them, the FNR will become very low at the center frequencies of the notch filters because the loudspeaker signal $x(n)$ will be attenuated at these frequencies while the system input signal $u(n)$ will not. This leads to wrong decisions about the notch filters according to [42].

In [42], the combination of the PEM-AFROW with the described proactive notch-filtering was evaluated using the same simulated environment of the previous improvements. The results indicate that fast movements of the microphone over a distance of 300 mm may not cause instability when the combined system is used, while some instability may occur when only the PEM-AFROW method is used.

3.5 Results of the PEM-AFROW Method in the Literature

In this section, results available in the literature about the use of the PEM-AFROW method will be presented. In [3], the PEM-AFROW method was evaluated in simulated environment of HA and PA systems where feedback paths $F(q,n)$ with $L_F = 50$ and $L_F = 1000$ were used, respectively. For performance comparison, the PEM-AFC method was used. The evaluation was performed by measuring the difference between the impulse responses of the feedback path $F(q,n)$ and the adaptive filter $H(q,n)$ according to

$$\varepsilon(n) = \|\mathbf{f}(n) - \mathbf{h}(n)\|. \tag{3.41}$$

The source signal $v(n)$ consisted of 7 speech signals with little more than 30 s of duration and $f_s = 8$ kHz. The PEM-AFROW parameters were $L_A = 10$, $L_{stp} = 160$, $L_{ltp} = 80$, $D = 200$, $L_{B_{\min}} = 20$, $L_{B_{\max}} = 160$. The NLMS adaptive algorithm was used in both PEM-AFC and PEM-AFROW methods. In the HA system, the PEM-AFROW and PEM-AFC methods presented similar results achieving $\varepsilon \approx 0.1$. But in the PA system, the PEM-AFROW method outperformed the PEM-AFC achieving $\varepsilon \approx 0.025$ at the end of the simulation time while the PEM-AFC obtained $\varepsilon \approx 0.05$.

A more complete evaluation of AFC methods in a simulated environment was presented in [49]. It included the PEM-AFROW and AFC methods based on noise injection (AFC-NI), frequency shifting (AFC-FS), half-wave rectifier (AFC-HWR), forward path delay (AFC-FD) and cancellation path delay (AFC-CD). The source signal $v(n)$ was speech signals with duration of 30 s and $f_s = 16$ kHz. The feedback path $F(q,n)$ was a measured impulse response of a room with reverberation time of 125 ms until $t = 22.5$ s and then it

was changed for other measured impulse response of the same room. The broadband gain $K(n)$ was raised to 7 dB above the MSG of the PA system after the initial convergence of the AFC systems. It is worth mentioning that nothing was said about the initial value of $K(n)$, the way it was raised (if abruptly or slowly), and whether its increase occurred in the same way and at the same time instant for all AFC methods. The evaluation was performed by analysing the mean values of MSG(n) and sound quality, which was measured through the frequency-weighted log-spectral signal distortion (SD) metric defined in Section 3.6.4, considering only the last 15 s of simulation.

The AFC-NI method were evaluated with different values of the signal-to-noise ratio (SNR), where $-2.5 \leq \text{SNR} \leq 10$ dB. Among all the AFC methods, it presented the largest increase in the MSG, $\Delta\text{MSG}(n)$, and the worst sound quality. The sound quality and $\Delta\text{MSG}(n)$ increased and decreased monotonically as the SNR increased, respectively. When SNR = 10 dB, $\Delta\text{MSG}(n) \approx 10$ dB. The AFC-FS method was evaluated with different values of the frequency shift f_0, where $1 \leq f_0 \leq 20$ Hz. As f_0 increased, the sound quality increased monotonically while $\Delta\text{MSG}(n)$ did not vary too much. In fact, $\Delta\text{MSG}(n) < 7$ dB. The AFC-HWR method was evaluated with different values of α, the parameter that controls the amount of added nonlinearity as defined in (3.11), where $0.001 \leq \alpha \leq 0.5$. The increase in the MSG was extremely poor and did not reach 2 dB.

The AFC-FD and AFC-CD methods were evaluated with different values of the forward path delay d_1 and cancellation path delay d_2, respectively, where $0.3125 \leq d_{1,2} \leq 10$ ms. The increase in the MSG achieved by both methods did not reach 7 dB. The PEM-AFROW method was evaluated with different values of the short-time prediction filter length L_A, where $5 \leq L_A \leq 30$ samples. It achieved $\Delta\text{MSG} \approx 9.5$ dB when $L_A = 20$ and the best sound quality among all the AFC methods, thereby confirming that the PEM-AFROW is the AFC method that achieves the best overall performance.

A similar evaluation of AFC methods in a simulated environment was presented in [2]. It included the PEM-AFROW, AFC-NI and AFC-FS methods. The AFC-NI method added white noise to the loudspeaker signal $x(n)$ with SNR = 10 dB. The AFC-FS method shifted the spectrum of the loudspeaker signal $x(n)$ by $f_0 = 5$ Hz. The PEM-AFROW method used in [2] differs from the original proposed in [3] by estimating the short-time prediction filter $A(q, n)$ using a 50% frame overlap, instead of a non-overlapping frame, and considering the long-time prediction filter $B(q, n)$ as a three-tap filter, instead of only one-tap. The PEM-AFROW parameters were $L_A = 20$, $L_{stp} = 320$, $L_{ltp} = 320$, $D = 160$, $L_{B\min} = 16$, $L_{B\max} = 160$.

The source signal $v(n)$ was a speech signal with duration of 30 s and $f_s = 16$ kHz. The feedback path $F(q, n)$ was a measured room impulse response until $t = 22.5$ s and then it was changed for other measured impulse response of the same room. The broadband gain $K(n)$ was initialized to a value such that the PA system had an initial gain margin of 3 dB and remained at this value until $t = 7.5$ s. During the next 7.5 s, it was increased linearly (in dB scale) by 10 dB and remained at this value during the last 15 s of simulation. It

is noteworthy that, for this configuration of the broadband gain $K(n)$, the increase in $K(n)$ achieved by the AFC methods (10 dB) was higher than those by the FS (3 dB) and NHS (5 dB) methods, which indicates a superior performance of the AFC methods. The evaluation was performed by analysing the mean and maximum values of the MSG(n) and sound quality, which was measured through the SD metric. And the values of MSG(n) over time were displayed.

Table 3.1 summarizes the results obtained by each AFC method considering only the last 15 s of simulation. The AFC-NI presented the highest ΔMSG but the worst sound quality while the PEM-AFROW achieved a similar ΔMSG and the best mean sound quality. However, it is worth emphasizing that only one speech signal was used which is statistically insufficient to accurately conclude about the efficiency of any method. Moreover, it is possible to observe from the curves MSG(n) that the initial value of MSG(n) obtained by the three evaluated AFC methods has different values: MSG(0) ≈ 10 dB for AFC-NI, MSG(0) ≈ 4.5 dB for AFC-FS and MSG(0) ≈ 3 dB for PEM-AFROW; and these values are not equal to the MSG of the PA system with no AFC method. This suggests that different initializations of the adaptive filter $H(q, n)$ may have been used for each AFC method and $H(q, 0) \neq 0$ in all cases.

Table 3.1: Performance comparison of AFC systems.

Method	Parameter values	ΔMSG (mean/max)	SD (mean/max)
AFC-NI	SNR = 10 dB	9.8 dB / 13.7 dB	15.1 dB / 31.7 dB
AFC-FS	$f_0 = 5$ Hz	6.6 dB / 11.1 dB	6.0 dB / 10.6 dB
PEM-AFROW	$L_A = 20$, $L = 320$, $D = 160$ $L_{B\min} = 16$, $L_{B\max} = 160$	9.6 dB / 12.8 dB	3.9 dB / 16.2 dB

3.6 Simulation Configurations

With the aim to assess the performance of the PEM-AFROW method, an experiment was carried out in a simulated environment to measure its ability to estimate the feedback path impulse response and increase the MSG of a PA system. Moreover, the spectral distortion in the resulting error signal $e(n)$ was also measured. To this purpose, the following configuration was used.

3.6.1 Simulated Environment

The impulse response $\mathbf{f}(n)$ of the acoustic feedback path was a measured room impulse response, from [60], and thus $\mathbf{f}(n) = \mathbf{f}$. The impulse response was downsampled to $f_s =$ 16 kHz and then truncated to length $L_F = 4000$ samples, and is illustrated in Figure 3.3.

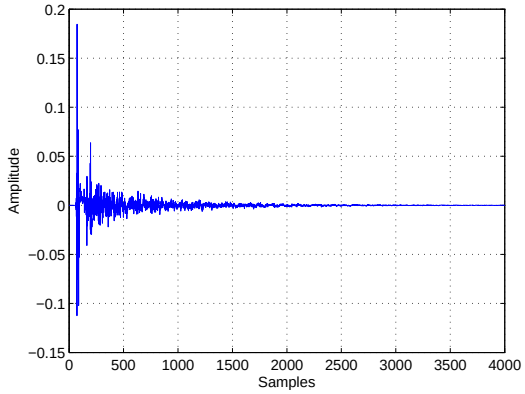


Figure 3.3: Impulse response $\mathbf{f}(n)$ of the feedback path.

The forward path $G(q)$, which is typically the amplifier of the PA system, was defined according to (2.2) as an unit delay and a gain, leading to

$$G(q,n) = g_1(n)q^{-1} \tag{3.42}$$

with length $L_G = 2$. Then, according to (2.9), $K(n) = g_1(n)$ and $J(q,n) = q^{-1}$.

Denoting the MSG of the PA system defined in (2.10) as $\text{MSG}_0 = 20\log_{10} K_0$, the broadband gain $K(n)$ of the forward path was initialized to a value K_1 such that $20\log_{10} K_1 < \text{MSG}_0$ in order to allow the AFC method to operate in a stable condition and thus the adaptive filter $H(q,n)$ to converge. As suggested in [2], it was defined that $20\log_{10} K_1 = \text{MSG}_0 - 3$, i.e., a 3 dB initial gain margin.

In a first configuration, $K(n) = K_1$ during all the simulation time $T = 20$ s to verify the method performance for a time-invariant $G(q,n)$. Afterwards, in a more practical configuration, $K(n) = K_1$ until 5 s and then $20\log_{10} K(n)$ was increased at the rate of $1\,^{\text{dB}}/_{\text{s}}$ up to $20\log_{10} K_2$ such that $20\log_{10} K_2 = 20\log_{10} K_1 + \Delta K$. Finally, $K(n) = K_2$ during 10 s totaling a simulation time $T = 15 + \Delta K$ s. This configuration of the broadband gain $K(n)$ is depicted in Figure 3.4.

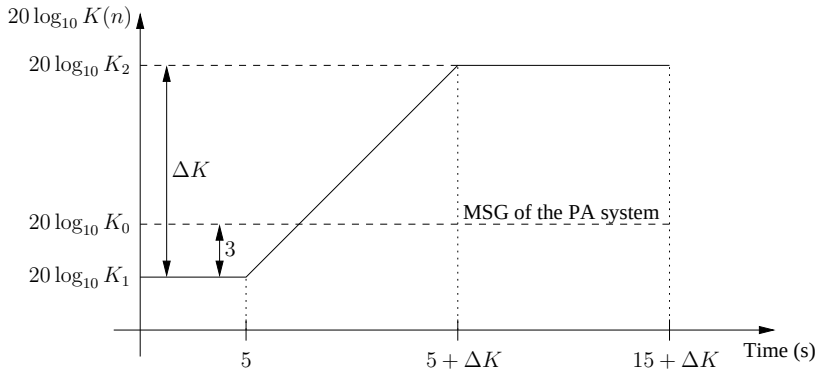


Figure 3.4: Practical configuration of the broadband gain $K(n)$ of the forward path.

The delay filter $D(q)$ was a delay line given by (2.3) with $L_D = 401$, equivalent to 25 ms as in [2, 3]. The use of the delay filter $D(q)$ is a common practice in the AFC because it helps reduce the correlation between the loudspeaker signal $x(n)$ and the system input signal $u(n)$ and, consequently, improve the performance of the adaptive filter. In the PEM-AFROW method, a delay filter $D(q)$ with L_D higher than the source model length, i.e., $L_D > L_A + L_{B\max}$, is strictly necessary to fulfill identifiability conditions [3].

3.6.2 Maximum Stable Gain

The main goal of any AFC method is to increase the MSG of the PA system that has an upper limit due to the acoustic feedback. Therefore, the MSG is the most important metric in evaluating AFC methods.

The PEM-AFROW method does not apply any processing to the signals that travel in the system other than the adaptive filter $H(q, n)$. Thus, for a AFC system using the PEM-AFROW method, the MSG of the AFC system and the increase in MSG, ΔMSG, were measured according to (3.6) and (3.8), respectively. Their optimum values are $\text{MSG}(n) = \Delta\text{MSG}(n) = \infty$ and they are achieved when $H(e^{j\omega}, n) = F(e^{j\omega}, n)$, $\omega \in P_H(n)$. In general, $\text{MSG}(n) \to \infty$ and $\Delta\text{MSG}(n) \to \infty$ as $H(e^{j\omega}, n) \to F(e^{j\omega}, n)$, $\omega \in P_H(n)$.

The frequency responses in (3.6) and (3.8) were computed using an N_{FFT_e}-point FFT. In obtaining the sets of critical frequencies $P(n)$ and $P_H(n)$, the phase of their respective functions was unwrapped and a search for each crossing by integer multiples of 2π was performed. For each crossing, the frequency component ω closer to the corresponding integer multiple of 2π was defined as critical frequency.

Considering $L_D = 1601$, an experiment was carried out to verify the number of detectable critical frequencies and, mainly, the accuracy of the measured $\text{MSG}(n)$ as a function of N_{FFT_e}. In order to cover a wide range of scenarios, 16 different impulse responses $\mathbf{h}$ of the adaptive filter (including $\mathbf{h} = 0$), such that $-30 \leq \text{MIS} \leq 0$ dB, were used. For each $\mathbf{h}$, the real value of MSG, MSG_r, of the AFC system was manually obtained by varying the broadband gain $K(n)$ of the forward path and observing the waveform of the loudspeaker signal $x(n)$ as in Figure 2.3, resulting in $0 \leq \text{MSG}_r \leq 25$ dB. The source signal $v(n)$ was one white noise.

Then, for each $\mathbf{h}$, the measured $\text{MSG}(n)$ and the number of critical frequencies, N_{cf}, of the AFC system were obtained with several values of N_{FFT_e}. The absolute error (in linear scale) between MSG_r and $\text{MSG}(n)$ was defined as the measurement error MSG_e. In addition, the variation in the number of detected critical frequencies, ΔN_{cf}, by increasing N_{FFT_e} was also obtained.

Table 3.2 shows the mean values of MSG_e and ΔN_{cf} for the evaluated N_{FFT_e} values. It can be observed that N_{FFT_e} has a great influence on the number of detectable critical frequencies and consequently on the accuracy of the measured $\text{MSG}(n)$. For a given system, increasing N_{FFT_e} generally increased the number of detected critical frequencies

Table 3.2: Summary of the results obtained by the PEM-AFROW method using speech as source signal.

N_{FFT_e}	$\overline{\text{MSG}_e}$ (dB)	$\overline{\Delta N_{cf}}$
2^{12}	2.9	
2^{13}	-10.3	734.56
2^{14}	-16.0	118.12
2^{15}	-19.2	19.31
2^{16}	-20.8	3.56
2^{17}	-24.0	0.62
2^{18}	-24.2	0.44
2^{19}	-24.3	0.19
2^{20}	-24.4	0.06

and decreased the measured MSG(n) value. However, both values saturate from $N_{FFT_e} = 2^{17}$ on and, therefore, this value was used in the following simulations.

With concerns about computational complexity, the MSG(n) measurement was only performed every 1000 samples (equivalent to 62.5 ms). In the meantime, the MSG(n) retained the last measured value.

3.6.3 Misalignment

A very common metric in evaluating adaptive filters when they are applied in system identification is the misalignment (MIS) . The MIS measures the mismatch between the adaptive filter and the system to be identified. In this work, the performance of the AFC methods was evaluated through the normalized MIS defined as [6]

$$\text{MIS}(n) = \frac{\|\mathbf{f}(n) - \mathbf{h}(n)\|}{\|\mathbf{f}(n)\|} = \frac{\|F(e^{j\omega}, n) - H(e^{j\omega}, n)\|}{\|F(e^{j\omega}, n)\|}. \tag{3.43}$$

Its optimum value is MIS(n) = 0 and is achieved when $\mathbf{h}(n) = \mathbf{f}(n)$ ($F(e^{j\omega}, n) = H(e^{j\omega}, n)$). In general, MIS($n$) $\to$ 0 as $\mathbf{h}(n) \to \mathbf{f}(n)$ ($F(e^{j\omega}, n) \to H(e^{j\omega}, n)$).

The MIS(n) has been used to evaluate and compare the performance of AFC methods as in [3, 4]. The MIS(n) and MSG(n) metrics are related, which means that an improvement in one of them usually results in an improvement in the other. However, this may not occur because the MSG(n) depends on the accuracy of $H(e^{j\omega}, n)$ in only one frequency component while the MIS(n) depends on its average accuracy over all frequency components.

3.6.4 Frequency-weighted Log-spectral Signal Distortion

The sound quality was measured through the frequency-weighted log-spectral signal distortion defined as [2]

$$\mathrm{SD}(n) = \sqrt{\int_{\omega_l}^{\omega_u} w(\omega) \left[10 \log_{10} \frac{S_e(e^{j\omega}, n)}{S_u(e^{j\omega}, n)}\right]^2 d\omega}, \tag{3.44}$$

where $S_e(e^{j\omega}, n)$ and $S_u(e^{j\omega}, n)$ are the short-term power spectral densities of the error signal $e(n)$ and system input signal $u(n)$, respectively, and $w(\omega)$ is a weighting function that gives equal weight to each auditory critical band between $\omega_l = 0.0375\pi$ (equivalent to 300 Hz) and $\omega_u = 0.8\pi$ (equivalent to 6400 Hz) [61]. The short-term power spectral densities were computed using non-overlapping frames with length of 20 ms.

Indeed, $\mathrm{SD}(n)$ measures the spectral distance (in dB scale) between the error signal $e(n)$ and the system input signal $u(n)$. Its optimum value is $\mathrm{SD}(n) = 0$ and is achieved when $\mathbf{h}(n) = \mathbf{f}(n)$ and thus $e(n) = u(n)$. In general, $\mathrm{SD}(n) \to 0$ as $\mathbf{h}(n) \to \mathbf{f}(n)$.

3.6.5 Wideband Perceptual Evaluation of Speech Quality

Objective measures of speech quality have evolved from those based on purely mathematical criteria, such as the SD previously described, towards perceptually salient metrics. The W-PESQ is a standard algorithm for objective quality evaluation of wideband (sampled at 16 kHz) speech signals [62, 63, 64, 65]. It employs reference (original) and degraded (processed) versions of a speech signal to evaluate the perceptible degradation of the latter, which can be quantified in the 1-5 mean opinion score (MOS) scale. The correspondence between the MOS scale and the degradation category rating (DCR) is shown in Table 3.3. However, the maximum MOS given by the W-PESQ algorithm is 4.644.

Table 3.3: MOS Scale.

Score	DCR Listening Quality
5	Inaudible
4	Audible but not annoying
3	Slightly annoying
2	Annoying
1	Very annoying

The W-PESQ achieves a correlation of 80% with MOS when assessing speech impairment by reverberation although it was not designed for this purpose [66, 67]. Hence, in this work, the W-PESQ algorithm was used to perceptually evaluate the resulting distortion in the error signal $e(n)$ due to the acoustic feedback. For that, the system input signal $u(n)$ was considered the reference signal.

The W-PESQ was originally validated with signals that mostly have $8 - 12$ s of duration but shorter signals can be used if they have at least 3.2 s of speech [68]. Thus, the error signal $e(n)$ and the system input signal $u(n)$ were divided in non-overlapping segments with duration of 4 s in order to evaluate the sound quality over time through the W-PESQ algorithm. Experiments proved that, for the AFC methods under evaluation, the maximum difference between the MOS given by W-PESQ when using segments with duration of 4 and 10 s was only 0.03.

3.6.6 Speech Database

The signal database used in the simulations consisted of 10 speech signals. Each speech signal was composed of several basic signals from a speech database. Each basic signal contains one short sentence recorded in a time slot of 4 s and with $f_s = 48$ kHz, but downsampled to $f_s = 16$ kHz. All sentences were spoken by native speakers, which had the following nationalities and genders:

- 4 Americans (2 males and 2 females)
- 2 British (1 male and 1 female)
- 2 French (1 male and 1 female)
- 2 Germans (1 male and 1 female)

Since the performance assessment of adaptive filters needs longer signals, several basic signals from the same speaker were concatenated and had their silence parts removed through a voice activity detector (VAD), resulting in the mentioned 10 speech signals (1 signal per speaker). The length of the speech signals varied with the simulation time.

3.7 Simulation Results

This section presents the performance of the PEM-AFROW method using the configuration of the PA system, the evaluation metrics and the signals described in Section 3.6. The parameters of the PEM-AFROW, except those of the adaptive filter, had the values originally proposed in [3] adjusted to $f_s = 16$ kHz resulting in $L_A = 20$, $L_{stp} = 320$, $L_{ltp} = 160$, $L_D = 401$, $L_{B\min} = 40$, $L_{B\max} = 320$ samples.

The parameters of the NLMS adaptive filtering algorithm (stepsize μ, normalization parameter δ and L_H) were optimized for each signal. From pre-defined ranges, the values of μ, δ and L_H were chosen empirically in order to optimize the curve $\mathrm{MSG}(n)$, and consequently $\Delta\mathrm{MSG}(n)$, with regard to minimum area of instability and, secondarily, maximum mean value within the simulation time. The optimal curves for the kth signal were denoted as $\mathrm{MSG}_k(n)$ and $\Delta\mathrm{MSG}_k(n)$ while the curves $\mathrm{MIS}(n)$, $\mathrm{SD}(n)$ and $\mathrm{WPESQ}(n)$

obtained with the same values of μ, δ and L_H were denoted as $\text{MIS}_k(n)$, $\text{SD}_k(n)$ and $\text{WPESQ}_k(n)$, respectively.

Then, the mean curves $\text{MSG}(n)$, $\Delta\text{MSG}(n)$, $\text{MIS}(n)$, $\text{SD}(n)$ and $\text{WPESQ}(n)$ were obtained by averaging the curves of each signal according to

$$\begin{aligned} \text{MSG}(n) &= \frac{1}{10}\sum_{k=1}^{10}\text{MSG}_k(n) & \Delta\text{MSG}(n) &= \frac{1}{10}\sum_{k=1}^{10}\Delta\text{MSG}_k(n) \\ \text{MIS}(n) &= \frac{1}{10}\sum_{k=1}^{10}\text{MIS}_k(n) & \text{SD}(n) &= \frac{1}{10}\sum_{k=1}^{10}\text{SD}_k(n) \\ \text{SD}(n) &= \frac{1}{10}\sum_{k=1}^{10}\text{SD}_k(n) & \text{WPESQ}(n) &= \frac{1}{10}\sum_{k=1}^{10}\text{WPESQ}_k(n). \end{aligned} \tag{3.45}$$

And their respective mean values were defined as

$$\begin{aligned} \overline{\text{MSG}} &= \frac{1}{N_T}\sum_{n=1}^{N_T}\text{MSG}(n) & \overline{\Delta\text{MSG}} &= \frac{1}{N_T}\sum_{n=1}^{N_T}\Delta\text{MSG}(n) \\ \overline{\Delta\text{MSG}} &= \frac{1}{N_T}\sum_{n=1}^{N_T}\Delta\text{MSG}(n) & \overline{\text{MIS}} &= \frac{1}{N_T}\sum_{n=1}^{N_T}\text{MIS}(n) \\ \overline{\text{SD}} &= \frac{1}{N_T}\sum_{n=1}^{N_T}\text{SD}(n) & \overline{\text{WPESQ}} &= \frac{1}{N_T}\sum_{n=1}^{N_T}\text{WPESQ}(n). \end{aligned} \tag{3.46}$$

where N_T is the number of samples related to the simulation time. Moreover, the asymptotic values of $\text{MIS}(n)$, $\Delta\text{MSG}(n)$, $\text{SD}(n)$ and $\text{WPESQ}(n)$ were denoted by $\overrightarrow{\text{MIS}}$, $\overrightarrow{\Delta\text{MSG}}$, $\overrightarrow{\text{SD}}$ and $\overrightarrow{\text{WPESQ}}$, respectively, and were estimated by graphically inspecting the curves.

The evaluation was done in two ambient noise conditions. The first was an ideal condition where the ambient noise signal $r(n) = 0$ and thus the source-signal-to-ambient-noise ratio $\text{SNR} = \infty$. The second was close to real-world conditions where $r(n) \neq 0$ such that $\text{SNR} = 30$ dB. The ambient noise $r(n)$ reduces the cross-correlation between the system input signal $u(n)$ and the loudspeaker signal $x(n)$, thereby improving the performance of any gradient-based or least-squares-based AFC method as the PEM-AFROW.

In the first configuration, the broadband gain $K(n)$ remained constant, i.e. $\Delta K = 0$, and coincidentally $\text{MSG}_0 \approx 0$ dB, which results in $\Delta\text{MSG}(n) \approx \text{MSG}(n)$ and $K_1 \approx -3$ dB. Figure 3.5 shows the results obtained by the PEM-AFROW method for $\Delta K = 0$. The bias problem in AFC is illustrated through the results obtained by the NLMS adaptive filtering algorithm (the same used by the PEM-AFROW) with no decorrelation method when $\text{SNR} = 30$ dB. The PEM-AFROW method achieved $\overrightarrow{\Delta\text{MSG}} \approx 7.5$ dB and $\overrightarrow{\text{MIS}} \approx -8.7$ dB when $\text{SNR} = \infty$, and $\overrightarrow{\Delta\text{MSG}} \approx 8$ dB and $\overrightarrow{\text{MIS}} \approx -9.3$ dB when $\text{SNR} = 30$ dB. With no decorrelation method, the NLMS algorithm achieved only $\overrightarrow{\Delta\text{MSG}} \approx 2.5$ dB and $\overrightarrow{\text{MIS}} \approx -1.4$ dB when $\text{SNR} = \infty$ or 30 dB. Regarding sound quality, the PEM-AFROW achieved $\overrightarrow{\text{SD}} \approx 1.7$ and $\overrightarrow{\text{WPESQ}} \approx 2.43$ when $\text{SNR} = \infty$, and $\overrightarrow{\text{SD}} \approx 1.5$ and

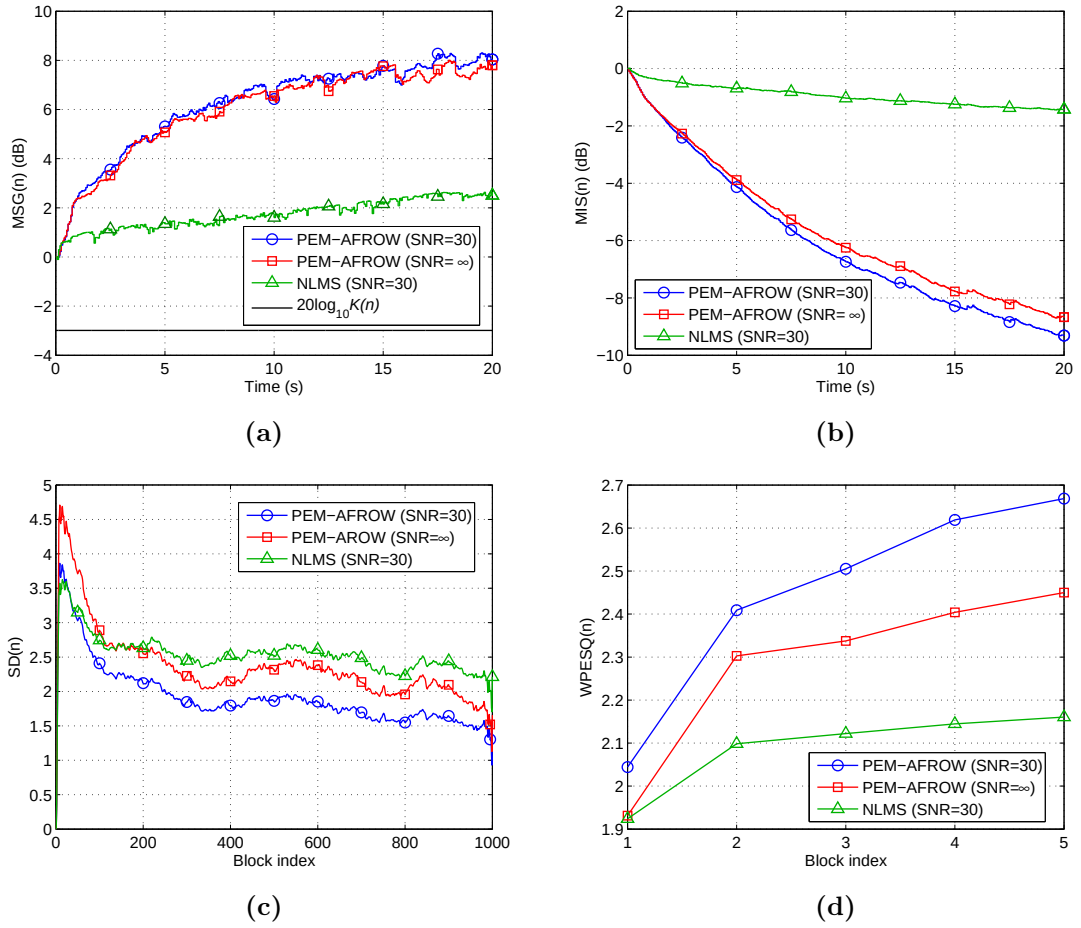


Figure 3.5: Average results of the PEM-AFROW method for speech signals and $\Delta K = 0$: (a) MSG(n); (b) MIS(n); (c) SD(n); (b) WPESQ(n).

$\overrightarrow{\text{WPESQ}} \approx 2.64$ when SNR $= 30$ dB. The NLMS obtained $\overrightarrow{\text{SD}} \approx 2.6$ and $\overrightarrow{\text{WPESQ}} \approx 2.01$ when SNR $= \infty$, and $\overrightarrow{\text{SD}} \approx 2.2$ and $\overrightarrow{\text{WPESQ}} \approx 2.15$ when SNR $= 30$ dB. The effectiveness of the PEM-AFROW becomes clear when comparing its results with those of the NLMS.

In the second configuration, $K(n)$ was increased, as explained in Section 3.6.1, in order to determine the maximum stable broadband gain (MSBG) achievable by the PEM-AFROW method for both ambient noise conditions. The MSBG was defined as the maximum value of K_2 with which an AFC method achieves a MSG(n) completely stable. Such situation occurred firstly with $\Delta K = 14$ dB for SNR $= \infty$. Figure 3.6 shows the results obtained by the PEM-AFROW method in this case. The PEM-AFROW method achieved $\overrightarrow{\Delta\text{MSG}} \approx 13.3$ dB and $\overrightarrow{\text{MIS}} \approx -14.3$ dB when SNR $= \infty$, and $\overrightarrow{\Delta\text{MSG}} \approx 13.4$ dB and $\overrightarrow{\text{MIS}} \approx -14.9$ dB when SNR $= 30$ dB. With respect to sound quality, the PEM-AFROW achieved $\overrightarrow{\text{SD}} \approx 5.0$ and $\overrightarrow{\text{WPESQ}} \approx 1.46$ when SNR $= \infty$, and $\overrightarrow{\text{SD}} \approx 3.9$ and $\overrightarrow{\text{WPESQ}} \approx 1.63$ when SNR $= 30$ dB.

Finally, $K(n)$ was increased further to determine the MSBG of the PEM-AFROW method when SNR $= 30$ dB. This situation occurred with $\Delta K = 16$ dB and Figure 3.7

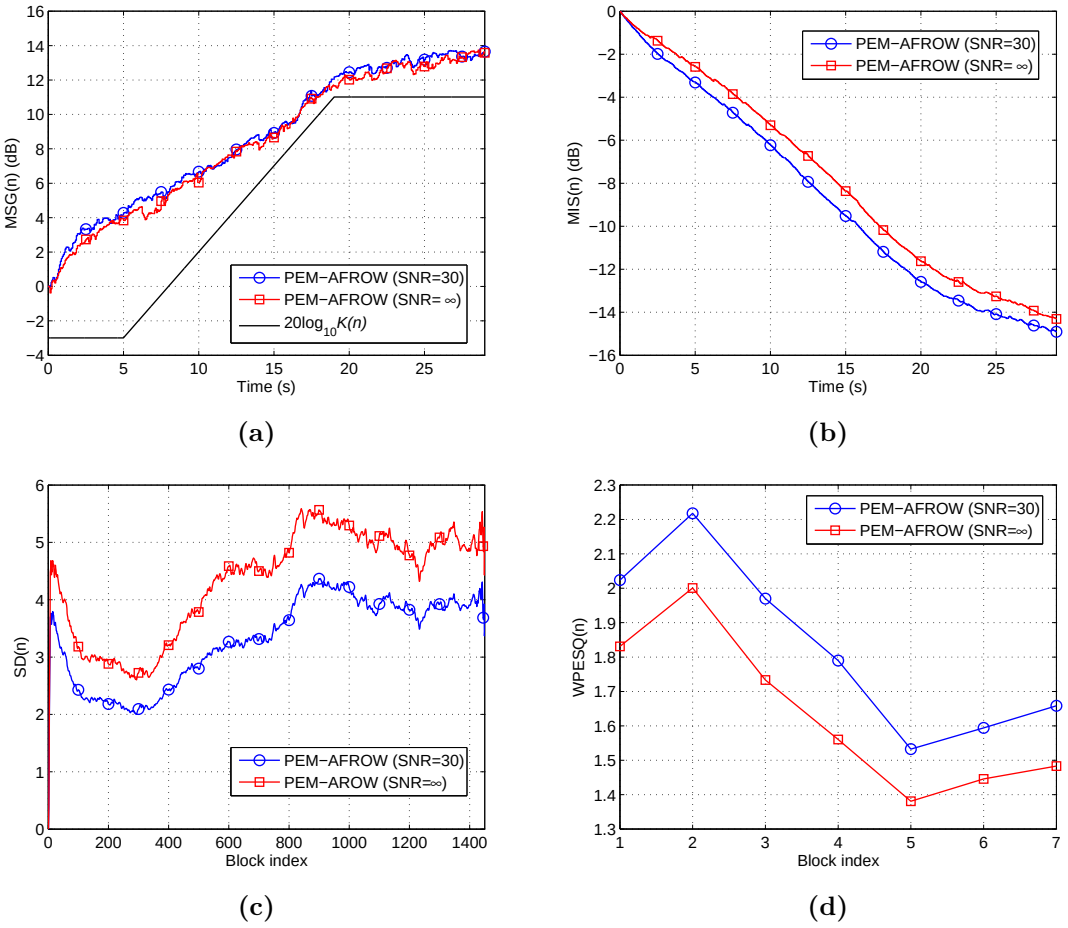


Figure 3.6: Average results of the PEM-AFROW method for speech signals and $\Delta K = 14$ dB: (a) MSG(n); (b) MIS(n); (c) SD(n); (b) WPESQ(n).

shows the results obtained by the PEM-AFROW method in this case. The PEM-AFROW method achieved $\overrightarrow{\Delta\text{MSG}} \approx 14$ dB and $\overrightarrow{\text{MIS}} \approx -15.4$ dB when SNR $= \infty$, and $\overrightarrow{\Delta\text{MSG}} \approx 15$ dB and $\overrightarrow{\text{MIS}} \approx -16.4$ dB when SNR $= 30$ dB. With respect to sound quality, the PEM-AFROW achieved $\overrightarrow{\text{SD}} \approx 5.1$ and $\overrightarrow{\text{WPESQ}} \approx 1.42$ when SNR $= \infty$ and $\overrightarrow{\text{SD}} \approx 3.9$ and $\overrightarrow{\text{WPESQ}} \approx 1.58$ when SNR $= 30$ dB. Table 3.4 summarizes the results obtained by the PEM-AFROW method using speech as source signal $v(n)$.

It can be observed that the results of MSG(n) and MIS(n) improve as ΔK increases. This can be explained by the fact that, when the broadband gain $K(n)$ of the forward path is increased, the energy of the feedback signal (desired signal to the adaptive filter) is increased while the energy of the system input signal $u(n)$ (noise signal to the adaptive filter) remains fixed. Then, the ratio between the energies of the feedback and input signals is increased which improves the performance of the traditional adaptive filtering algorithms and, consequently, of the PEM-AFROW method.

On the other hand, the results of SD(n) and WPESQ(n) worsen as ΔK increases. This is because, despite the improvement in the estimates of the feedback path provided by the

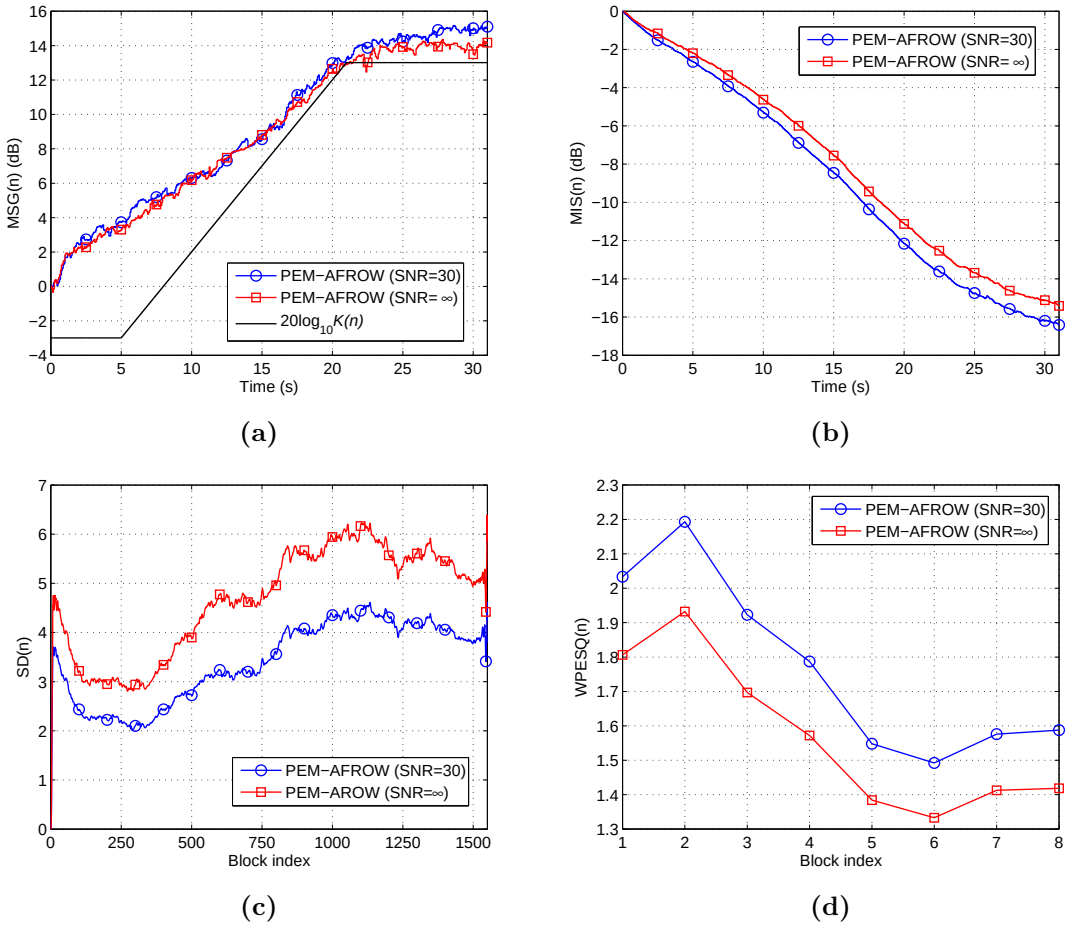


Figure 3.7: Average results of the PEM-AFROW method for speech signals and $\Delta K = 16$ dB: (a) MSG(n); (b) MIS(n); (c) SD(n); (b) WPESQ(n).

adaptive filters, the increase in the gain of $G(q, n)$ ultimately results in an increase in the energy of the uncancelled feedback signal $[\mathbf{f}(n) - \mathbf{h}(n)] * x(n)$. From an MSG point of view, this can be concluded by observing that the stability margin of the systems decreases. For $\Delta K = 14$ and mainly 16 dB, the stability margin became very low which resulted in an excessive reverberation or even in some howlings in the error signal $e(n)$.

Furthermore, the W-PESQ algorithm proved to be sensitive to the distortions caused by the uncancelled feedback signals because only mean values lower than 3, which is the middle of the MOS scale, were obtained. And this occurs even with a stability margin of approximately 8 dB as achieved by the PEM-AFROW for $\Delta K = 0$. This high sensitivity may be due to the W-PESQ algorithm not being designed to evaluate speech impairment by reverberation. However, from Figures 3.5, 3.6 and 3.7, it can be concluded that the SD metric and W-PESQ algorithm had a consistent behavior because they indicated that the sound quality improves as the energy of the uncancelled feedback signal decreases.

With respect to the ambient noise conditions, the results obtained with SNR = 30 dB are slightly better than those obtained with SNR = ∞ because, as already explained, the

ambient noise $r(n)$ reduces the cross-correlation between the system input signal $u(n)$ and the loudspeaker signal $x(n)$. This improves the performance of any AFC method that uses the traditional gradient-based or least-squares-based adaptive filtering algorithm, as the PEM-AFROW. Moreover, $r(n)$ helps to overcome a numeric issue of the SD metric, which will be explained in Section 4.7.1, and probably to perceptually mask the distortions inserted in $e(n)$. Both facts tend to improve the results of SD(n) and WPESQ(n).

Table 3.4: Summary of the results obtained by the PEM-AFROW method for speech signals.

			$\overline{\Delta\text{MSG}}$	$\overrightarrow{\Delta\text{MSG}}$	$\overline{\text{MIS}}$	$\overrightarrow{\text{MIS}}$	$\overline{\text{SD}}$	$\overrightarrow{\text{SD}}$	$\overline{\text{WPESQ}}$	$\overrightarrow{\text{WPESQ}}$
NLMS	$\Delta K = 0$	SNR $= 30$	1.7	2.5	-0.9	-1.4	2.6	2.2	2.09	2.15
		SNR $= \infty$	1.7	2.5	-0.9	-1.4	3.0	2.6	1.95	2.01
PEM-AFROW	$\Delta K = 0$	SNR $= 30$	6.2	8.0	-6.0	-9.3	1.9	1.5	2.45	2.64
		SNR $= \infty$	6.0	7.5	-5.6	-8.7	2.4	1.7	2.29	2.43
	$\Delta K = 14$	SNR $= 30$	8.7	13.4	-8.7	-14.9	3.3	3.9	1.83	1.63
		SNR $= \infty$	8.5	13.3	-7.9	-14.3	4.3	5.0	1.63	1.46
	$\Delta K = 16$	SNR $= 30$	9.2	15.0	-8.8	-16.4	3.4	3.9	1.77	1.58
		SNR $= \infty$	8.9	14.0	-8.0	-15.4	4.7	5.1	1.57	1.42

3.8 Conclusion

This chapter addressed the topic of acoustic feedback cancellation. The AFC approach uses an adaptive filter to identify the acoustic feedback path and remove its influence from the system. Nevertheless, due to the electro-acoustic path, the system input and loudspeaker signals are highly correlated, mainly when the source signal is colored as speech. Then, if the traditional gradient-based or least-squares-based adaptive filtering algorithms are used, a bias is introduced in adaptive filter coefficients.

The main solutions available in the literature to overcome the bias in the estimate of the feedback path were described. Mostly, they attempt to decorrelate the loudspeaker and system input signals but still using the traditional adaptive filtering algorithms. They can be divided in two groups. The first group contains the methods that insert a processing device in the system open-loop in order to change the waveform of the loudspeaker signal. This implies a fidelity loss of the PA system, even if the feedback signal is totally cancelled, that, however, may be neglected if the added processing device does not perceptually affect the sound quality of the system, which is particularly difficult to achieve. The second group is formed by the methods that do not apply any processing to the signals that travel in the system other than the adaptive filter and thereby keep the fidelity of the PA system as high as possible.

Among all, the PEM-AFROW method stood out for producing the best overall performance and, for this reason, was described in detail. The PEM-based methods consider that the system input signal, which acts as noise to the estimation of feedback path, is modeled by a filter whose input is white noise. Then, the idea consists on prefiltering the loudspeaker and microphone signals with the inverse source model, in order to whiten them, before feeding them to the adaptive filtering algorithm. The PEM-AFROW defines the source model as a cascade of short-time and long-time prediction filters that model the vocal tract and the periodicity, respectively.

An evaluation of the state-of-art PEM-AFROW method was carried out in a simulated environment using a measured room impulse response as the feedback path impulse response, a time-varying forward path broadband gain and two ambient noise conditions. Its ability to estimate the feedback path impulse response and increase the MSG of a PA system were measured as well as the spectral distortion in the resulting error signal.

Simulations demonstrated that, when the source signal is speech, the state-of-art PEM-AFROW method is able to estimate the feedback path impulse response with a MIS of -15.4 dB when SNR $= \infty$ and -16.4 dB when SNR $= 30$ dB. And it is able to increase the MSG of the PA system by 14 dB when SNR $= \infty$ and 15 dB when SNR $= 30$. With regard to sound quality when achieving these results, the PEM-AFROW method obtained a SD of 5.1 when SNR $= \infty$ and 3.9 when SNR $= 30$ dB, and a WPESQ grade of 1.42 when SNR $= \infty$ and 1.58 when SNR $= 30$ dB.

Chapter 4

Acoustic Feedback Cancellation Based on Cepstral Analysis

4.1 Introduction

As discussed in Chapter 3, AFC methods use an adaptive filter to identify the feedback path impulse response and then remove its influence from the system. However, due to the strong correlation between the system input and loudspeaker signals, a bias is introduced in the adaptive filter coefficients if the gradient-based or least-square-based adaptive filtering algorithms are used. To overcome the bias problem, the state-of-art PEM-AFROW method generates uncorrelated versions of the system input and loudspeaker signals to update the adaptive filter using the gradient-based NLMS adaptive filtering algorithm.

Another possible solution is to overcome the bias problem in AFC would be to not update the adaptive filter using the traditional gradient-based or least-square-based adaptive filtering algorithms. Following this approach, a method that updates the adaptive filter using information contained in the cepstrum of the microphone signal $y(n)$ was proposed in [69]. However, a detailed cepstral analysis of the system as a function of $G(q,n)$, $D(q)$, $F(q,n)$ and $H(q,n)$ was not considered, which most probably limited the results obtained at the time. Furthermore, the evaluation of the method performance was unclear and no comparison with other AFC methods was presented.

Cepstral analysis is a technique of signal analysis based on an homomorphic transformation that results in the so-called cepstrum. The cesptral representation enables that a convolution of two signals in the time domain, thus nonlinear in the frequency domain, is represented as a linear combination in the cesptral domain [58, 70, 71]. The cepstrum was proposed in 1963 as a better alternative to the autocorrelation function to detect echoes in seismic signals [70]. Due to the property of transforming a convolution into a linear combination, the cepstral analysis is quite suitable for deconvolution and has been widely applied in speech processing for pitch detection [58].

This chapter reformulates the cepstral analysis of PA and AFC systems. It proves that the cepstra of the microphone signal $y(n)$ and the error signal $e(n)$ may contain well-defined time domain information about the system through $G(q,n)$, $D(q)$, $F(q,n)$ and $H(q,n)$ if some gain conditions are fulfilled. Then, new AFC methods that compute estimates of the feedback path impulse response from cepstra of the microphone signal $y(n)$ and error signal $e(n)$ to update the adaptive filter are developed and their performances are compared with the state-of-art PEM-AFROW method.

4.2 Cepstral Analysis of PA Systems

The PA system depicted in Figure 2.1 is described by the following time domain equations

$$\begin{cases} y(n) = u(n) + \mathbf{f}(n) * x(n) \\ x(n) = \mathbf{g}(n) * \mathbf{d} * y(n) \end{cases} \tag{4.1}$$

and their corresponding representations in the frequency domain

$$\begin{cases} Y(e^{j\omega},n) = U(e^{j\omega},n) + F(e^{j\omega},n)X(e^{j\omega},n) \\ X(e^{j\omega},n) = G(e^{j\omega},n)D(e^{j\omega})Y(e^{j\omega},n) \end{cases}. \tag{4.2}$$

From (4.2), the frequency-domain relationship between the system input signal $u(n)$ and the microphone signal $y(n)$ is obtained as

$$Y(e^{j\omega},n) = \frac{1}{1 - G(e^{j\omega},n)D(e^{j\omega})F(e^{j\omega},n)} U(e^{j\omega},n), \tag{4.3}$$

which by applying the natural logarithm becomes

$$\ln\left[Y(e^{j\omega},n)\right] = \ln\left[U(e^{j\omega},n)\right] - \ln\left[1 - G(e^{j\omega},n)D(e^{j\omega})F(e^{j\omega},n)\right]. \tag{4.4}$$

If $\left|G(e^{j\omega},n)D(e^{j\omega})F(e^{j\omega},n)\right| < 1$, a sufficient condition to ensure the stability of the PA system, the second term on the right-hand side of (4.4) can be expanded in Taylor's series as

$$\ln\left[1 - G(e^{j\omega},n)D(e^{j\omega})F(e^{j\omega},n)\right] = -\sum_{k=1}^{\infty} \frac{\left[G(e^{j\omega},n)D(e^{j\omega})F(e^{j\omega},n)\right]^k}{k}. \tag{4.5}$$

Replacing (4.5) in (4.4) and applying the inverse Fourier transform as follows

$$\begin{aligned} \mathcal{F}^{-1}\left\{\ln\left[Y(e^{j\omega},n)\right]\right\} &= \mathcal{F}^{-1}\left\{\ln\left[U(e^{j\omega},n)\right]\right\} \\ &\quad + \mathcal{F}^{-1}\left\{\sum_{k=1}^{\infty} \frac{\left[G(e^{j\omega},n)D(e^{j\omega})F(e^{j\omega},n)\right]^k}{k}\right\}, \end{aligned} \tag{4.6}$$

the cepstral domain relationship between the system input signal $u(n)$ and the microphone signal $y(n)$ is obtained as

$$\mathbf{c_y}(n) = \mathbf{c_u}(n) + \sum_{k=1}^{\infty} \frac{[\mathbf{g}(n) * \mathbf{d} * \mathbf{f}(n)]^{*k}}{k}, \tag{4.7}$$

where $\{\cdot\}^{*k}$ denotes the kth convolution power which, in the case of an impulse response, is hereafter called k-fold impulse response.

In a PA system, the cepstrum $\mathbf{c_y}(n)$ of the microphone signal is the cepstrum $\mathbf{c_u}(n)$ of the system input signal added to a time domain series as a function of $\mathbf{g}(n)$, $\mathbf{d}$ and $\mathbf{f}(n)$. The presence of this time domain series is due to the disappearance of the logarithm operator in the rightmost term of (4.6). This series is formed by impulse responses that are k-fold convolutions of $\mathbf{g}(n) * \mathbf{d} * \mathbf{f}(n)$, the open-loop impulse response of the PA system, and they can be physically interpreted as impulse responses of k consecutive loops through the system. Therefore, it is crucial to understand that the cepstrum $\mathbf{c_y}(n)$ of the microphone signal contains time-domain information about the PA system through the impulse responses $\mathbf{g}(n)$, $\mathbf{d}$ and $\mathbf{f}(n)$.

In fact, the cepstral analysis modified the representation of the components of the PA system in relation to the system input signal $u(n)$. In (4.3), the system input signal $u(n)$ and the components of the PA system are represented in the frequency domain. But in (4.7), the system input signal $u(n)$ is represented in the cepstral domain while the components of the PA system are actually represented in the time domain.

It should be reminded that the cepstrum $\mathbf{c_y}(n)$ of the microphone signal in a PA system is defined by (4.7) if and only if the condition $\left|G(e^{j\omega}, n)D(e^{jw})F(e^{j\omega}, n)\right| < 1$ for the expansions in Taylor's series in (4.5) is fulfilled. Otherwise, nothing can be inferred about the mathematical definition of $\mathbf{c_y}(n)$ as a function of $\mathbf{g}(n)$, $\mathbf{d}$ and $\mathbf{f}(n)$. The condition $\left|G(e^{j\omega}, n)D(e^{jw})F(e^{j\omega}, n)\right| < 1$ is the gain condition of the Nyquist's stability criterion and therefore is hereafter called Nyquist's gain condition (NGC) of the PA system. The NGC of the PA system is sufficient to ensure system stability because it considers all the frequency components while the Nyquist's stability criterion considers only those that satisfy the phase condition defined in (2.5). As a consequence, the broadband gain $K(n)$ of the forward path, defined in (2.8), must be, in general, lower than the MSG of the PA system to fulfill it. And even though $\mathbf{c_y}(n)$ is mathematically defined by (4.7), the practical existence of these impulse responses in $\mathbf{c_y}(n)$ depends on whether the size of the time domain observation window is large enough to include their effects.

With the aim to illustrate the modification caused by the cepstral analysis on the representation of the components of the PA system, consider a PA system with the time-invariant open-loop impulse response $\mathbf{g}(n) * \mathbf{d} * \mathbf{f}(n)$ depicted in Figure 4.1b, a white noise with duration of 100 s as the source signal $v(n)$ and $r(n) = 0$. The NGC of the PA system is fulfilled in this case and Figure 4.1 shows the first 5000 samples of $\mathbf{c_y}(n)$ as well as the

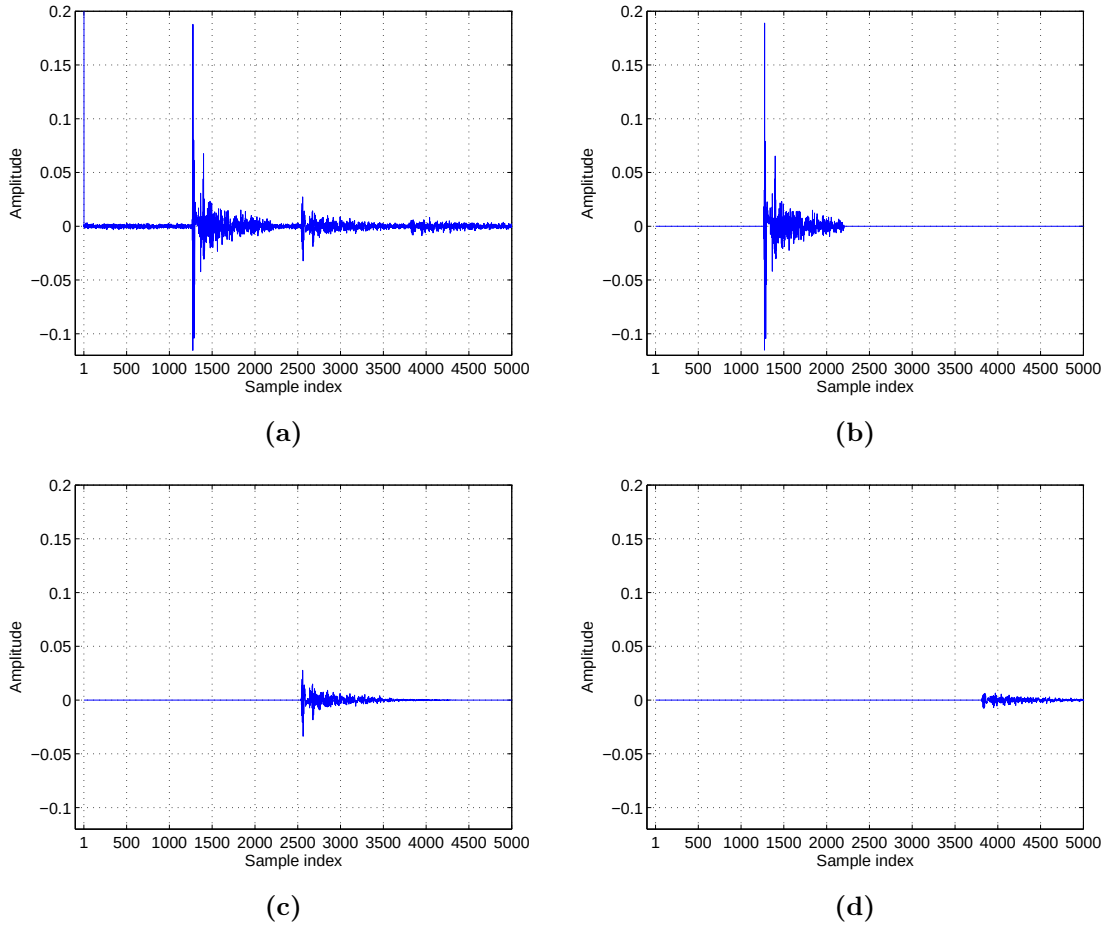


Figure 4.1: Cepstrum of the microphone signal $y(n)$ in a PA system when $v(n)$ is a white noise: (a) $\mathbf{c_y}(n)$; (b) $\mathbf{g}(n) * \mathbf{d} * \mathbf{f}(n)$; (c) $\frac{[\mathbf{g}(n)*\mathbf{d}*\mathbf{f}(n)]^{*2}}{2}$; (d) $\frac{[\mathbf{g}(n)*\mathbf{d}*\mathbf{f}(n)]^{*3}}{3}$

1, 2 and 3-fold convolutions of the open-loop impulse response. The cepstrum $\mathbf{c_y}(n)$ was computed using the entire content of the microphone signal and thus $\mathbf{c_u}(n)$ approached its theoretical impulse-like waveform. It can be concluded that, in $\mathbf{c_y}(n)$, the components of the PA system are really represented in the time domain.

Moreover, two characteristics of the k-fold impulse responses can be observed in Figure 4.1: decrease in magnitude with increasing fold k; and increasing sliding to the right on the sample axis of their non-zero values with increasing fold k. The former is explained by the fact that the absolute values of the open-loop impulse response $\mathbf{g}(n) * \mathbf{d} * \mathbf{f}(n)$ are generally much smaller than 1 so that the PA system is stable, as can be observed in Figure 4.1b, and the weight factor $1/k$ in the series penalizes the increase in the fold. The latter is due to the open-loop impulse response has a time delay, as can be seen in Figure 4.1b, because of $D(q)$ and $F(q, n)$ (which has a time delay determined by the distance between microphone and loudspeaker).

Along with the fact that $\mathbf{f}(n)$, as a room impulse response, typically has several prominent peaks associated with the early reflections [66], the first characteristic causes the 1-fold

Table 4.1: MSE between system input and microphone signals after removing consecutively the weighted k-fold impulse responses from the cepstrum of the microphone signal.

Number of removed impulse responses	MSE	ΔMSE
0	5.7e-1	-
1	6.0e-2	5.1e-1
2	1.9e-2	4.1e-2
5	3.5e-3	1.6e-3
10	8.8e-4	2.6e-3
20	2.1e-4	6.7e-4
50	3.0e-5	1.8e-4
100	6.7e-6	2.3e-5

impulse response, the open-loop impulse response, to be easily noticeable in $\mathbf{c_y}(n)$. The 2-fold impulse response is also noticeable but not as much as the 1-fold one. The 3-fold impulse response is hardly distinguishable from $\mathbf{c_u}(n)$. However, the ease of viewing the k-fold impulse responses in $\mathbf{c_y}(n)$ depends on the waveform of $\mathbf{c_u}(n)$, which, as a cepstrum, decays at least as fast as $1/m$ where m is its sample index [70].

In order to completely remove the acoustic feedback, it is necessary to remove all the time domain information about the PA system from the cepstrum of the microphone signal, i.e. in order to obtain $y(n) = u(n)$ it is necessary to make $\mathbf{c_y}(n) = \mathbf{c_u}(n)$. With $r(n) = 0$, Table 4.1 presents the mean square error (MSE) between the system input signal $u(n)$ and the microphone signal $y(n)$ after the removal of the weighted impulse responses from $\mathbf{c_y}(n)$ in a simulated environment. The removal process was performed by subtracting consecutively the weighted impulse responses from $\mathbf{c_y}(n)$, starting always by the 1-fold impulse response (open-loop impulse response $\mathbf{g}(n) * \mathbf{d} * \mathbf{f}(n)$). That is, to remove N impulse responses means to remove up to the N-fold impulse response ($k = 1, 2, \ldots, N$).

It can be observed from Table 4.1 that the greater the number of consecutively removed weighted impulse responses, the more the microphone signal $y(n)$ approaches the system input signal $u(n)$. However, the variation in MSE, ΔMSE, that is obtained by removing one impulse response decreases with increasing fold. This is due to the fact that the impulse responses with higher folds have a lower contribution to the distortion of the system input signal $u(n)$ because of, as already explained, their lower absolute values.

A process to remove the acoustic feedback can be developed similarly to the simulated experiment. It would be possible to detect or, at least, to estimate the region of $\mathbf{c_y}(n)$ where each weighted impulse response in (4.7) is located. This could be performed, for instance, by searching for the highest peak of the 1-fold impulse response in $\mathbf{c_y}(n)$ and using this knowledge to estimate the position of the other impulse responses. Hence, the impulse responses could be removed from $\mathbf{c_y}(n)$ through cepstral processing, i.e., by

processing directly in $\mathbf{c_y}(n)$. In this process, the lower fold impulse responses should be prioritized because of their larger contribution to the distortion of $u(n)$.

It is also possible to exploit the modification applied by the cepstral analysis on the representation of the components of the PA system in relation to the system input signal in order to develop an AFC method, where an adaptive filter $H(q,n)$ estimates the feedback path $F(q,n)$ and removes its influence from the system. But here, instead of the traditional gradient-based or least-squares-based adaptive filtering algorithms, the adaptive filter $H(q,n)$ will be updated based on time domain information about the PA system estimated from $\mathbf{c_y}(n)$.

4.3 Cepstral Analysis of AFC Systems

An AFC system is a PA system with an AFC method, i.e., that uses an adaptive filter $H(q,n)$ to remove the influence of the feedback path $F(q,n)$ from the system, as shown in Figure 3.1. The insertion of $H(q,n)$ changes the relationships between the system signals with respect to (4.1) and (4.2), in the PA system, and generates the error signal $e(n)$ from the microphone signal $y(n)$.

Regardless of how the adaptive filter $H(q,n)$ is updated, which allows to disregard the adaptive algorithm block with no loss of generality, the AFC system depicted in Figure 3.1 is described by the following time domain equations

$$\begin{cases} y(n) = u(n) + \mathbf{f}(n) * x(n) \\ e(n) = y(n) - \mathbf{h}(n) * x(n) \\ x(n) = \mathbf{g}(n) * \mathbf{d} * e(n) \end{cases} \tag{4.8}$$

and their corresponding representations in the frequency domain

$$\begin{cases} Y(e^{j\omega},n) = U(e^{jw},n) + F(e^{j\omega},n)X(e^{j\omega},n) \\ E(e^{j\omega},n) = Y(e^{j\omega},n) - H(e^{j\omega},n)X(e^{j\omega},n) \\ X(e^{j\omega},n) = G(e^{j\omega},n)D(e^{j\omega})E(e^{j\omega},n) \end{cases}. \tag{4.9}$$

4.3.1 Cepstral Analysis of the Microphone Signal

From (4.9), the frequency-domain relationship between the system input signal $u(n)$ and the microphone signal $y(n)$ is given by

$$Y(e^{j\omega},n) = \frac{1 + G(e^{j\omega},n)D(e^{j\omega})H(e^{j\omega},n)}{1 - G(e^{j\omega},n)D(e^{j\omega})\left[F(e^{j\omega},n) - H(e^{j\omega},n)\right]}U(e^{j\omega},n), \tag{4.10}$$

which by applying the natural logarithm becomes

$$\begin{aligned}\ln\left[Y(e^{j\omega},n)\right] = \ln\left[U(e^{j\omega},n)\right] + \ln\left[1+G(e^{j\omega},n)D(e^{j\omega})H(e^{j\omega},n)\right] \\ - \ln\left\{1-G(e^{j\omega},n)D(e^{j\omega})\left[F(e^{j\omega},n)-H(e^{j\omega},n)\right]\right\}.\end{aligned} \tag{4.11}$$

If $\left|G(e^{j\omega},n)D(e^{j\omega})H(e^{j\omega},n)\right| < 1$, the second term on the right-hand side of (4.11) can be expanded in Taylor's series as

$$\ln\left[1+G(e^{j\omega},n)D(e^{j\omega})H(e^{j\omega},n)\right] = \sum_{k=1}^{\infty}(-1)^{k+1}\frac{\left[G(e^{j\omega},n)D(e^{j\omega})H(e^{j\omega},n)\right]^k}{k}. \tag{4.12}$$

And if $\left|G(e^{j\omega},n)D(e^{j\omega})\left[F(e^{j\omega},n)-H(e^{j\omega},n)\right]\right| < 1$, a sufficient condition to ensure the stability of the AFC system, the third term on the right-hand side of (4.11) can be expanded in Taylor's series as

$$\begin{aligned}\ln\left\{1-G(e^{j\omega},n)D(e^{j\omega})\left[F(e^{j\omega},n)-H(e^{j\omega},n)\right]\right\} = \\ -\sum_{k=1}^{\infty}\frac{\left[G(e^{j\omega},n)D(e^{j\omega})\left[F(e^{j\omega},n)-H(e^{j\omega},n)\right]\right]^k}{k}.\end{aligned} \tag{4.13}$$

Replacing (4.12) and (4.13) in (4.11), and applying the inverse Fourier transform as follows

$$\begin{aligned}\mathcal{F}^{-1}\left\{\ln\left[Y(e^{j\omega},n)\right]\right\} = \mathcal{F}^{-1}\left\{\ln\left[U(e^{j\omega},n)\right]\right\} \\ + \mathcal{F}^{-1}\left\{\sum_{k=1}^{\infty}(-1)^{k+1}\frac{\left[G(e^{j\omega},n)D(e^{j\omega})H(e^{j\omega},n)\right]^k}{k}\right\} \\ + \mathcal{F}^{-1}\left\{\sum_{k=1}^{\infty}\frac{\left\{G(e^{j\omega},n)D(e^{j\omega})\left[F(e^{j\omega},n)-H(e^{j\omega},n)\right]\right\}^k}{k}\right\},\end{aligned} \tag{4.14}$$

the cepstral domain relationship between the system input signal $u(n)$ and the microphone signal $y(n)$ is obtained as

$$\begin{aligned}\mathbf{c_y}(n) = \mathbf{c_u}(n) + \sum_{k=1}^{\infty}(-1)^{k+1}\frac{[\mathbf{g}(n)*\mathbf{d}*\mathbf{h}(n)]^{*k}}{k} \\ + \sum_{k=1}^{\infty}\frac{\{\mathbf{g}(n)*\mathbf{d}*[\mathbf{f}(n)-\mathbf{h}(n)]\}^{*k}}{k}.\end{aligned} \tag{4.15}$$

In an AFC system, the cepstrum $\mathbf{c_y}(n)$ of the microphone signal is the cepstrum $\mathbf{c_u}(n)$ of the system input signal added to two time-domain series as functions of $\mathbf{g}(n)$, $\mathbf{d}$, $\mathbf{f}(n)$ and $\mathbf{h}(n)$. Similarly to (4.7), the presence of these time-domain series is due to the disappearance of the logarithm operator in the last two terms of (4.14). These series are formed by k-fold convolutions of $\mathbf{g}(n)*\mathbf{d}*[\mathbf{f}(n)-\mathbf{h}(n)]$, the open-loop impulse

response of the AFC system, and $\mathbf{g}(n) * \mathbf{d} * \mathbf{h}(n)$. Therefore, the cepstrum $\mathbf{c_y}(n)$ of the microphone signal contains time domain information about the AFC system through the impulse responses $\mathbf{g}(n)$, $\mathbf{d}$, $\mathbf{f}(n)$ and $\mathbf{h}(n)$.

The cepstral domain relationship in (4.15) can be re-written as

$$\mathbf{c_y}(n) = \mathbf{c_u}(n) + \sum_{k=1}^{\infty} \frac{[\mathbf{g}(n) * \mathbf{d}]^{*k}}{k} * \left\{ [\mathbf{f}(n) - \mathbf{h}(n)]^{*k} + (-1)^{k+1} \mathbf{h}^{*k}(n) \right\}. \tag{4.16}$$

The resulting 1-fold ($k = 1$) impulse response is $\mathbf{g}(n) * \mathbf{d} * \mathbf{f}(n)$, the open-loop impulse response, and is identical to the one in (4.7). It is crucial to understand that, regardless of $\mathbf{h}(n)$, the open-loop impulse response $\mathbf{g}(n) * \mathbf{d} * \mathbf{f}(n)$ is always the 1-fold impulse response present in $\mathbf{c_y}(n)$. On the other hand, the resulting higher fold ($k > 1$) impulse responses present in (4.16) are different from those in (4.7) due to the insertion of the adaptive filter $H(q, n)$. It is noticeable that (4.15) and (4.16) differ from (4.7) except when $\mathbf{h}(n) = 0$, condition that makes the two systems equivalent.

Ideally, if the adaptive filter exactly matches the feedback path, i.e., $H(q, n) = F(q, n)$, the frequency domain relationship between the system input signal $u(n)$ and the microphone signal $y(n)$ defined in (4.10) will become

$$Y(e^{j\omega}, n) = \left[1 + G(e^{j\omega}, n) D(e^{j\omega}) F(e^{j\omega}, n)\right] U(e^{j\omega}, n), \tag{4.17}$$

which will imply the following time domain relationship

$$y(n) = [1 + \mathbf{g}(n) * \mathbf{d} * \mathbf{f}(n)] * u(n). \tag{4.18}$$

This means that the microphone signal $y(n)$ will continue to have acoustic feedback even in the ideal situation where $H(q, n) = F(q, n)$. This is explained by the fact that the influence of the open-loop impulse response, $\mathbf{g}(n) * \mathbf{d} * \mathbf{f}(n)$, is unavoidable because the AFC method is applied only after the feedback signal is picked-up by the microphone. This is the reason why $\mathbf{g}(n) * \mathbf{d} * \mathbf{f}(n)$ is always present in $\mathbf{c_y}(n)$ regardless of $\mathbf{h}(n)$. In the cepstral domain, the relationship in (4.16) will become

$$\mathbf{c_y}(n) = \mathbf{c_u}(n) + \sum_{k=1}^{\infty} (-1)^{k+1} \frac{[\mathbf{g}(n) * \mathbf{d} * \mathbf{f}(n)]^{*k}}{k}, \tag{4.19}$$

which proves that the peaks of $\mathbf{c_y}(n)$ caused by the acoustic feedback will exist even if $H(q, n) = F(q, n)$. The difference to (4.7), in the PA system, is that the even k-fold weighed impulse responses have mirrored amplitudes.

Note that the cesptrum $\mathbf{c_y}(n)$ of the microphone signal in an AFC system is defined by (4.16) if and only if the conditions $\left|G(e^{j\omega}, n) D(e^{j\omega}) H(e^{j\omega}, n)\right| < 1$ and $\left|G(e^{j\omega}, n) D(e^{j\omega}) \left[F(e^{j\omega}, n) - H(e^{j\omega}, n)\right]\right| < 1$ for the expansions in Taylor's series in (4.12) and (4.13), respectively, are fulfilled. Otherwise, nothing can be inferred about the mathematical

definition of $\mathbf{c_y}(n)$ as a function of $\mathbf{g}(n)$, $\mathbf{d}$, $\mathbf{f}(n)$ and $\mathbf{h}(n)$.

Similarly to the condition $\left|G(e^{j\omega},n)D(e^{j\omega})F(e^{j\omega},n)\right| < 1$ in the PA system, the condition $\left|G(e^{j\omega},n)D(e^{j\omega})\left[F(e^{j\omega},n) - H(e^{j\omega},n)\right]\right| < 1$ is the NGC of the AFC system. But, while the fulfillment of the NGC of the PA system is the only requirement to define $\mathbf{c_y}(n)$ according to (4.7), the fulfillment of the NGC of the AFC system is not sufficient to define $\mathbf{c_y}(n)$ according to (4.16). In addition to it, the condition $\left|G(e^{j\omega},n)D(e^{j\omega})\right.$ $\left.H(e^{j\omega},n)\right| < 1$ must also be fulfilled.

In a practical AFC system, $H(q,0) = 0$ and $H(q,n) \to F(q,n)$ as $n \to \infty$. When $n = 0$, the additional condition $\left|G(e^{j\omega},n)D(e^{j\omega})H(e^{j\omega},n)\right| < 1$ is fulfilled and $K(0)$ can be infinite. But as $H(q,n)$ converges to $F(q,n)$, the maximum value of the broadband gain $K(n)$ of the forward path that fulfills the condition decreases. Finally, when $n \to \infty$, the condition becomes $\left|G(e^{j\omega},n)D(e^{j\omega})F(e^{j\omega},n)\right| < 1$, the NGC of the PA system, and the broadband gain $K(n)$ must be lower than the MSG of the PA system to fulfill it.

Therefore, in an AFC system, the cepstrum $\mathbf{c_y}(n)$ of the microphone signal is ultimately defined by (4.16) if the NGC of both AFC and PA systems are fulfilled. This restricts the use of $\mathbf{c_y}(n)$ in AFC systems because if the broadband gain $K(n)$ of the forward path is increased above the MSG of the PA system, as intended in AFC systems, the condition $\left|G(e^{j\omega},n)D(e^{j\omega})H(e^{j\omega},n)\right| < 1$ may no longer be fulfilled and thereby $\mathbf{c_y}(n)$ may not be defined by (4.16). This is the critical issue of the cepstral analysis of the microphone signal in AFC systems that limits the performance of any AFC method solely based on $\mathbf{c_y}(n)$.

In addition to the above theoretical discussion about the critical issue of $\mathbf{c_y}(n)$ in AFC systems, the present work will demonstrate it in practice. In Section 4.4.1, an AFC method based on the cepstrum $\mathbf{c_y}(n)$ of the microphone signal will be proposed. The method will use the fact that $\mathbf{g}(n) * \mathbf{d} * \mathbf{f}(n)$ is always the 1-fold impulse response present in $\mathbf{c_y}(n)$, as proved in (4.16), and will estimate it from $\mathbf{c_y}(n)$ to update $H(q,n)$. It will be demonstrated in Section 4.7 that the AFC method based on $\mathbf{c_y}(n)$ will still work properly even if the broadband gain $K(n)$ of the forward path exceeds the MSG of the PA system by around 10 dB. However, above a certain value, $K(n)$ causes (4.16) to become inaccurate to the point of disrupting the estimate of the feedback path provided by the method. As a consequence, the method performance is limited by the broadband gain $K(n)$ of the forward path because of the need to fulfill the condition $\left|G(e^{j\omega},n)D(e^{j\omega})H(e^{j\omega},n)\right| < 1$. In general, this need may limit the use of $\mathbf{c_y}(n)$ in AFC systems.

4.3.2 Cepstral Analysis of the Error Signal

The cepstral analysis can provide time domain information about the AFC system in such a way that, as in a PA system, the only requirement is the fulfilment of its NGC. It should be understood that the need to fulfill the condition $\left|G(e^{j\omega},n)D(e^{j\omega})H(e^{j\omega},n)\right| < 1$ in order to mathematically define the cepstrum $\mathbf{c_y}(n)$ of the microphone signal by (4.16) is due to the numerator of (4.10). And this condition can be avoided by realizing, from (4.9), that the frequency domain relationship between the system input signal $u(n)$ and the error

signal $e(n)$, which was generated from the microphone signal $y(n)$, is given by

$$E(e^{j\omega},n)=\frac{1}{1-G(e^{j\omega},n)D(e^{j\omega})\left[F(e^{j\omega},n)-H(e^{j\omega},n)\right]}U(e^{j\omega},n), \tag{4.20}$$

which by applying the natural logarithm becomes

$$\ln\left[E(e^{j\omega},n)\right]=\ln\left[U(e^{j\omega},n)\right]-\ln\left\{1-G(e^{j\omega},n)D(e^{j\omega})\left[F(e^{j\omega},n)-H(e^{j\omega},n)\right]\right\}. \tag{4.21}$$

If $\left|G(e^{j\omega},n)D(e^{j\omega})\left[F(e^{j\omega},n)-H(e^{j\omega},n)\right]\right|<1$, a sufficient condition to ensure the stability of the AFC system, the second term on the right-hand side of (4.21) can be expanded in Taylor's series according to (4.13). Replacing (4.13) in (4.21), and applying the inverse Fourier transform as follows

$$\begin{aligned}\mathcal{F}^{-1}\left\{\ln\left[E(e^{j\omega},n)\right]\right\}&=\mathcal{F}^{-1}\left\{\ln\left[U(e^{j\omega},n)\right]\right\}\\&\quad+\mathcal{F}^{-1}\left\{\sum_{k=1}^{\infty}\frac{\left\{G(e^{j\omega},n)D(e^{j\omega})\left[F(e^{j\omega},n)-H(e^{j\omega},n)\right]\right\}^{k}}{k}\right\},\end{aligned} \tag{4.22}$$

the cepstral domain relationship between the system input signal $u(n)$ and the error signal $e(n)$ is obtained as

$$\mathbf{c_e}(n)=\mathbf{c_u}(n)+\sum_{k=1}^{\infty}\frac{\{\mathbf{g}(n)*\mathbf{d}*[\mathbf{f}(n)-\mathbf{h}(n)]\}^{*k}}{k}. \tag{4.23}$$

In an AFC system, the cepstrum $\mathbf{c_e}(n)$ of the error signal is the cepstrum $\mathbf{c_u}(n)$ of the system input signal added to a time domain series as a function of $\mathbf{g}(n)$, $\mathbf{d}$, $\mathbf{f}(n)$ and $\mathbf{h}(n)$. Similarly to (4.7) and (4.15), the presence of the time domain series is due to the disappearance of the logarithm operator in the rightmost term of (4.22). This series is formed by impulse responses that are k-fold convolutions of $\mathbf{g}(n)*\mathbf{d}*[\mathbf{f}(n)-\mathbf{h}(n)]$, the open-loop impulse response of the AFC system, and they can be physically interpreted as impulse responses of k consecutive loops through the system. Therefore, the cepstrum $\mathbf{c_e}(n)$ of the error signal also contains time domain information about the AFC system through $\mathbf{g}(n)$, $\mathbf{d}$, $\mathbf{f}(n)$ and $\mathbf{h}(n)$.

Contrary to $\mathbf{c_y}(n)$, all the k-fold impulse responses present in $\mathbf{c_e}(n)$ depend on $\mathbf{h}(n)$. It is noticeable that (4.23) differs from (4.16) except when $\mathbf{h}(n)=0$, condition that makes $e(n)=y(n)$. And most importantly, unlike $\mathbf{c_y}(n)$, the only requirement to define $\mathbf{c_e}(n)$ according to (4.23) is the fulfillment of $\left|G(e^{j\omega},n)D(e^{j\omega})\left[F(e^{j\omega},n)-H(e^{j\omega},n)\right]\right|<1$, the NGC of the AFC system. If the NGC of the AFC system is not fulfilled at all frequency components, the inaccuracy of (4.23) will depend on its deviation. However, experiments showed that (4.23) may remain accurate even with a deviation of a few dB in the NGC of the AFC system or even in the MSG of the AFC system.

Ideally, if the adaptive filter exactly matches the feedback path, i.e., $H(q,n) = F(q,n)$, (4.20) and (4.23) will become

$$E(e^{j\omega}, n) = U(e^{j\omega}, n) \tag{4.24}$$

and

$$\mathbf{c_e}(n) = \mathbf{c_u}(n), \tag{4.25}$$

respectively. In the time domain it will lead to

$$e(n) = u(n), \tag{4.26}$$

which means that the acoustic feedback will be completely cancelled. Generally, in a more realistic situation where $H(q,n) \approx F(q,n)$, the better the adaptive filter $H(q,n)$ matches the feedback path $F(q,n)$, the more the error signal $e(n)$ approaches the system input signal $u(n)$.

The present work demonstrated that, in an AFC system, the cepstrum $\mathbf{c_e}(n)$ of the error signal is mathematically defined as a function of $\mathbf{g}(n)$, $\mathbf{d}$, $\mathbf{f}(n)$ and $\mathbf{h}(n)$ if the NGC of the AFC system is fulfilled. This clearly represents an advantage over the cepstrum $\mathbf{c_y}(n)$ of the microphone signal because, besides the fulfillment of the NGC of the AFC system, it also requires the fulfillment of the condition $\left|G(e^{j\omega}, n)D(e^{j\omega})H(e^{j\omega}, n)\right| < 1$, which ultimately becomes the NGC of the PA system, to be similarly defined.

In addition to the above theoretical discussion, the present work will demonstrate it in practice. In Section 4.4.2, an AFC method based on the cepstrum $\mathbf{c_e}(n)$ of the error signal will be proposed. The method will use the fact that $\mathbf{g}(n) * \mathbf{d} * [\mathbf{f}(n) - \mathbf{h}(n)]$ is always the 1-fold impulse response present in $\mathbf{c_e}(n)$, as proved in (4.23), and will estimate it from $\mathbf{c_e}(n)$ to update $H(q,n)$. It is expected that the AFC method based on $\mathbf{c_e}(n)$ works properly regardless of the broadband gain $K(n)$ of the forward path if the NGC of the AFC system is fulfilled and, therefore, can further increase the MSG of the PA system compared with the AFC method based on $\mathbf{c_y}(n)$.

4.4 AFC Based on Cepstral Analysis

The only known method that uses cepstral analysis to eliminate or control the Larsen effect was proposed in [69]. The method uses the cepstrum $\mathbf{c_y}(n)$ of the microphone signal and an adaptive filter in a configuration similar to that shown in Figure 4.2 but using the error signal $e(n)$ as the adaptive filter input. As a consequence, the acoustic feedback would be completely removed if $H(q,n) = G(q,n)D(q)F(q,n)$, which means that the adaptive filter $H(q,n)$ must track variations not only in $F(q,n)$ but also in $G(q,n)$. This configuration is not commonly used because a change in the forward path $G(q,n)$ will increase the mismatch between the adaptive filter $H(q,n)$ and feedback path $F(q,n)$, thereby worsening the cancellation of the feedback signal and the stability condition.

However, a detailed cepstral analysis of the AFC system was not carried out in [69]. The expansions in Taylor's series as functions of $F(q,n)$, $G(q,n)$, $D(q)$ and $H(q,n)$ of the natural logarithms intrinsic to the cesptral analysis of the AFC system, in (4.12) and (4.13), were not considered. Hence, the need to fulfill de condition $\left|G(e^{j\omega},n)D(e^{j\omega})H(e^{j\omega},n)\right| < 1$ and the resulting consequences were not discussed. Lastly, $\mathbf{c_y}(n)$ was not defined as in (4.16), where it is clear that it comprises weighted impulse responses added to $\mathbf{c_u}(n)$. Instead, it was only stated that peaks are inserted in $\mathbf{c_y}(n)$ due to the acoustic feedback.

In [69], for each block of the microphone signal $y(n)$, the update of the adaptive filter was performed by, starting from a pre-defined sample index, selecting the peaks of $\mathbf{c_y}(n)$ above a pre-defined threshold, multiplying their values by a small pre-defined constant and adding them to $\mathbf{h}(n)$ at the sample indexes of the selected peaks. However, in the same way that it was proven in (4.16) that the system open-loop impulse response $\mathbf{g}(n)*\mathbf{d}*\mathbf{f}(n)$ is always the 1-fold impulse response present in $\mathbf{c_y}(n)$ regardless of $\mathbf{h}(n)$, the same occurs in the system configuration used in [69], although that was not observed. Then, if the peaks of $\mathbf{g}(n)*\mathbf{d}*\mathbf{f}(n)$ remain, over time, in the same sample indexes, only their values will be detected and added to $\mathbf{h}(n)$. As a consequence, the method may update the adaptive filter only in these sample indexes. This probably was the reason why the adaptive filter $H(q,n)$ did not have more than 50 coefficients in [69], which limits the performance of any AFC method because only a very small part of feedback path $F(q,n)$ is modeled. It is evident that this characteristic of the method proposed in [69] is not beneficial to the AFC system and should be avoided.

4.4.1 AFC Method Based on the Cepstrum of the Microphone Signal

The present work proposes a new AFC method based on the cepstrum $\mathbf{c_y}(n)$ of the microphone signal (AFC-CM) and its scheme is shown in Figure 4.2. As any AFC method, the AFC-CM method identifies and tracks the acoustic feedback path using an adaptive FIR filter. But, instead of the traditional gradient-based or least-squares-based adaptive filtering algorithms, the proposed AFC-CM method updates the adaptive filter using estimates of the impulse response $\mathbf{f}(n)$ of the feedback path computed from $\mathbf{c_y}(n)$.

As discussed in Section 4.2 and illustrated in Figure 4.1, among all the k-fold impulse responses present in the cepstrum $\mathbf{c_y}(n)$ of the microphone signal defined in (4.16), the open-loop impulse response $\mathbf{g}(n)*\mathbf{d}*\mathbf{f}(n)$ is the one with highest absolute values because it is the 1-fold impulse response. Therefore, although dependent on the waveform of $\mathbf{c_u}(n)$, it tends to be the impulse response more accurately estimated from $\mathbf{c_y}(n)$.

Hence, the proposed method starts by calculating $\{\mathbf{g}(n)*\mathbf{d}*\mathbf{f}(n)\}\hat{}$, an estimate of $\mathbf{g}(n)*\mathbf{d}*\mathbf{f}(n)$, the open-loop impulse response of the PA system, from $\mathbf{c_y}(n)$. This is performed by selecting the first $L_G+L_D+L_H-2$ samples of the cepstrum $\mathbf{c_y}(n)$ of the microphone signal, resulting in

$$\{\mathbf{g}(n)*\mathbf{d}*\mathbf{f}(n)\}\hat{} = \mathbf{g}(n)*\mathbf{d}*\mathbf{f}(n)+\mathbf{c_{u_0}}(n), \tag{4.27}$$

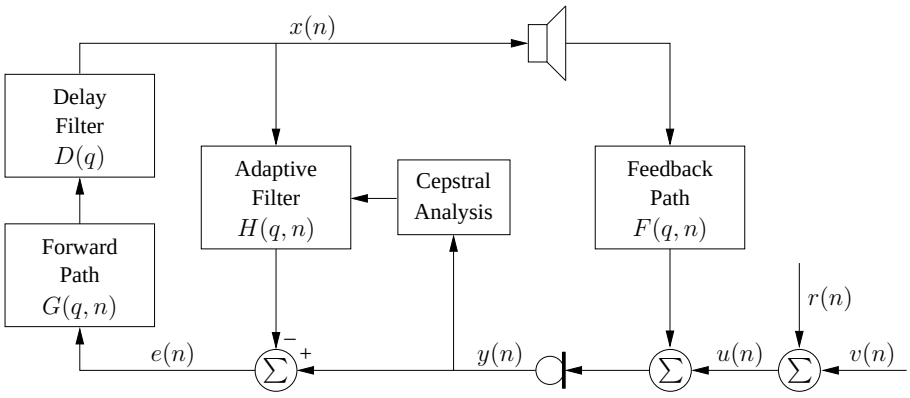


Figure 4.2: Acoustic feedback cancellation based on cepstral analysis of the microphone signal.

where

$$\mathbf{c}_{\mathbf{u}_0}(n) = \left[c_{u_0}(n) \;\; c_{u_1}(n) \;\; \ldots \;\; c_{u_{L_G+L_D+L_H-3}}(n)\right]^T. \tag{4.28}$$

Thereafter, the presented method calculates $\{\mathbf{g}(n) * \mathbf{f}(n)\}\hat{}$, an estimate of $\mathbf{g}(n) * \mathbf{f}(n)$, from (4.27) according to

$$\{\mathbf{g}(n) * \mathbf{f}(n)\}\hat{} = \{\mathbf{g}(n) * \mathbf{d} * \mathbf{f}(n)\}\hat{} * \mathbf{d}^{-1}. \tag{4.29}$$

Note that the convolution with $\mathbf{d}^{-1}$ is performed by sliding on the sample axis. This procedure results in

$$\{\mathbf{g}(n) * \mathbf{f}(n)\}\hat{} = \mathbf{g}(n) * \mathbf{f}(n) + \mathbf{c}_{\mathbf{u}_{L_D-1}}(n), \tag{4.30}$$

where

$$\mathbf{c}_{\mathbf{u}_{L_D-1}}(n) = \left[c_{u_{L_D-1}}(n) \;\; c_{u_{L_D}}(n) \;\; \ldots \;\; c_{u_{L_G+L_D+L_H-3}}(n)\right]^T. \tag{4.31}$$

The segment $\mathbf{c}_{\mathbf{u}_{L_D-1}}(n)$ from the cepstrum of the input signal acts as noise in the estimation of $\mathbf{g}(n) * \mathbf{f}(n)$ and it would prevent the proposed AFC-CM method from reaching the optimal solution $\{\mathbf{g}(n) * \mathbf{f}(n)\}\hat{} = \mathbf{g}(n) * \mathbf{f}(n)$. However, it will be proved in Section 4.4.3.1 that this estimation will be asymptotically consistent for the samples of $\mathbf{g}(n) * \mathbf{f}(n)$ with the highest absolute values, which are the most important ones, because it tends to reach the optimal solution.

The forward path $G(q, n)$ can be accurately estimated from its input (error $e(n)$) and output (loudspeaker $x(n)$) signals through any open-loop system identification method. Then, assuming prior knowledge of the forward path $G(q, n)$, the proposed method computes $\hat{\mathbf{f}}(n)$, an estimate of the impulse response $\mathbf{f}(n)$ of the feedback path, from (4.30) as follows

$$\hat{\mathbf{f}}(n) = \{\mathbf{g}(n) * \mathbf{f}(n)\}\hat{} * \mathbf{g}^{-1}(n). \tag{4.32}$$

Although the adaptive filter may be updated directly as $\mathbf{h}(n) = \hat{\mathbf{f}}(n)$, in order to increase robustness to short-burst disturbances, the update of the adaptive filter is performed

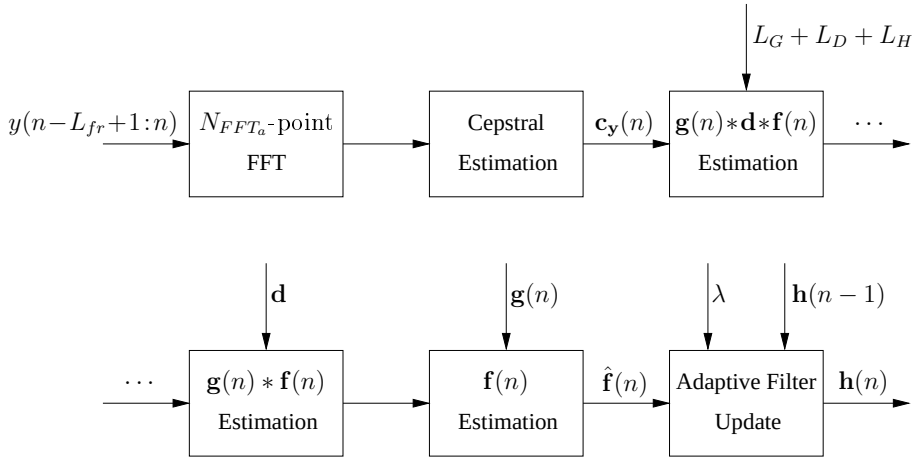


Figure 4.3: Block diagram of the proposed AFC-CM method.

according to

$$\mathbf{h}(n) = \lambda \mathbf{h}(n-1) + (1-\lambda)\hat{\mathbf{f}}(n), \tag{4.33}$$

where $0 \leq \lambda < 1$ is the factor that controls the trade-off between robustness and tracking rate of the adaptive filter.

In conclusion, the proposed AFC-CM method calculates an estimate of $\mathbf{f}(n)$ from $\mathbf{c_y}(n)$ to update $H(q,n)$. Depending on the variations of $F(q,n)$ over time, it can be deduced that this computational effort may not be worth it, regarding performance, if the method is applied to each new sample of the microphone signal $y(n)$. Therefore, the AFC-CM will be applied every N_{fr} samples, where N_{fr} is a parameter that controls the trade-off between performance (latency and tracking capability) and computational complexity.

The block diagram of the proposed AFC-CM method is depicted in Figure 4.3. Every N_{fr} samples, a frame of the microphone signal $y(n)$ containing its newest L_{fr} samples is selected; the frame has its spectrum $Y(e^{j\omega}, n)$ and power cepstrum $\mathbf{c_y}(n)$ calculated through an N_{FFT_a}-point Fast Fourier Transform (FFT); $\{\mathbf{g}(n) * \mathbf{d} * \mathbf{f}(n)\}\hat{}$ is computed from $\mathbf{c_y}(n)$; with the knowledge of $\mathbf{d}$, $\{\mathbf{g}(n) * \mathbf{f}(n)\}\hat{}$ is calculated; with an estimate of $\mathbf{g}(n)$, $\hat{\mathbf{f}}(n)$ is computed; finally, $\mathbf{h}(n)$ is updated.

4.4.2 AFC Method Based on the Cepstrum of the Error Signal

The present work also proposes an AFC method based on the cepstrum $\mathbf{c_e}(n)$ of the error signal (AFC-CE) and its scheme is shown in Figure 4.4. As any AFC method, the AFC-CE method identifies and tracks the acoustic feedback path using an adaptive FIR filter. But, instead of the traditional gradient-based or least-squares-based adaptive filtering algorithms, the proposed AFC-CE method updates the adaptive filter using estimates of the impulse response $\mathbf{f}(n)$ of the feedback path computed from $\mathbf{c_e}(n)$.

The concepts of the AFC-CE and AFC-CM are similar. They differ in the signal to which the cepstral analysis is applied and, consequently, in the time domain information

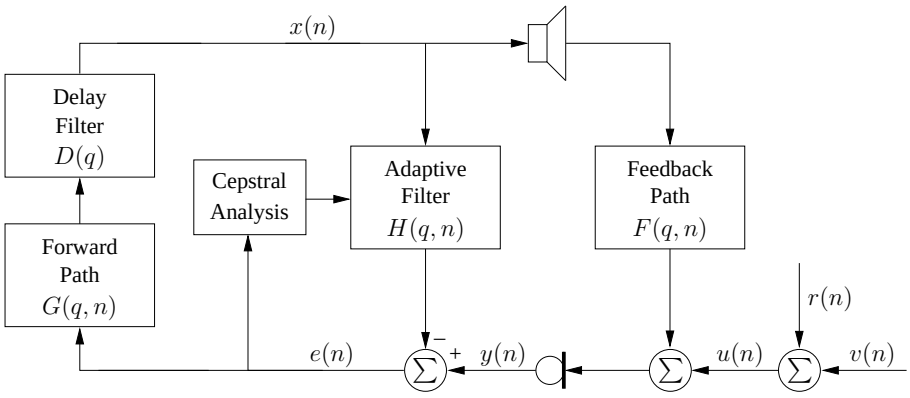


Figure 4.4: Acoustic feedback cancellation based on cepstral analysis of the error signal.

that is estimated from the cepstra. And most importantly, as discussed in detail in Sections 4.3.1 and 4.3.2, the only requirement in order for $\mathbf{c_e}(n)$ to be defined according to (4.23) is the fulfillment of the NGC of the AFC system. In contrast, in order for $\mathbf{c_y}(n)$ to be defined according to (4.16), the broadband gain $K(n)$ of the forward path must also fulfill the condition $\left|G(e^{j\omega},n)D(e^{j\omega})H(e^{j\omega},n)\right| < 1$, which ultimately becomes the NGC of the PA system. Therefore, it is expected that the AFC-CE method works properly regardless of $K(n)$ if the NGC of the AFC system is fulfilled, unlike the AFC-CM, and thus further increases the MSG of the PA system compared with the AFC-CM method.

The proposed AFC-CE method starts by calculating $\{\mathbf{g}(n) * \mathbf{d} * [\mathbf{f}(n) - \mathbf{h}(n)]\}\hat{}$, an estimate of $\mathbf{g}(n) * \mathbf{d} * [\mathbf{f}(n) - \mathbf{h}(n)]$, the open-loop impulse response of the AFC system, from $\mathbf{c_e}(n)$. This is performed by selecting the first $L_G + L_D + L_H - 2$ samples of the cepstrum $\mathbf{c_e}(n)$ of the error signal, resulting in

$$\{\mathbf{g}(n) * \mathbf{d} * [\mathbf{f}(n) - \mathbf{h}(n)]\}\hat{} = \mathbf{g}(n) * \mathbf{d} * [\mathbf{f}(n) - \mathbf{h}(n)] + \mathbf{c}_{\mathbf{u}_0}(n), \tag{4.34}$$

with $\mathbf{c}_{\mathbf{u}_0}(n)$ as defined in (4.28).

Thereafter, the presented method calculates $\{\mathbf{g}(n) * [\mathbf{f}(n) - \mathbf{h}(n)]\}\hat{}$, an estimate of $\mathbf{g}(n) * [\mathbf{f}(n) - \mathbf{h}(n)]$, from (4.34) according to

$$\{\mathbf{g}(n) * [\mathbf{f}(n) - \mathbf{h}(n)]\}\hat{} = \{\mathbf{g}(n) * \mathbf{d} * [\mathbf{f}(n) - \mathbf{h}(n)]\}\hat{} * \mathbf{d}^{-1}. \tag{4.35}$$

Note that the convolution with $\mathbf{d}^{-1}$ is performed by sliding on the sample axis. This procedure results in

$$\{\mathbf{g}(n) * [\mathbf{f}(n) - \mathbf{h}(n)]\}\hat{} = \mathbf{g}(n) * [\mathbf{f}(n) - \mathbf{h}(n)] + \mathbf{c}_{\mathbf{u}_{L_D-1}}(n), \tag{4.36}$$

with $\mathbf{c}_{\mathbf{u}_{L_D-1}}(n)$ as defined in (4.31).

Similarly to (4.30), the segment $\mathbf{c}_{\mathbf{u}_{L_D-1}}(n)$ from the cepstrum of the input signal acts as noise in the estimation of $\mathbf{g}(n) * [\mathbf{f}(n) - \mathbf{h}(n)]$ and it would prevent the proposed AFC-CE method from reaching the optimal solution $\{\mathbf{g}(n) * [\mathbf{f}(n) - \mathbf{h}(n)]\}\hat{} = \mathbf{g}(n) * [\mathbf{f}(n) - \mathbf{h}(n)]$.

However, it will be proved in Section 4.4.3.1 that this estimation will be asymptotically consistent for the samples of $\mathbf{g}(n) * [\mathbf{f}(n) - \mathbf{h}(n)]$ with the highest absolute values, which are the most important ones, because it tends to reach the optimal solution.

Assuming prior knowledge of the forward path $G(q, n)$, as in the AFC-CM method, the AFC-CE method computes $[\mathbf{f}(n) - \mathbf{h}(n)]\hat{}$, an estimate of $\mathbf{f}(n) - \mathbf{h}(n)$, from (4.36) according to

$$[\mathbf{f}(n) - \mathbf{h}(n)]\hat{} = \{\mathbf{g}(n) * [\mathbf{f}(n) - \mathbf{h}(n)]\}\hat{} * \mathbf{g}^{-1}(n). \tag{4.37}$$

Then, the proposed method calculates $\hat{\mathbf{f}}(n)$, an estimate of the impulse response $\mathbf{f}(n)$ of the feedback path, from (4.37) as follows

$$\hat{\mathbf{f}}(n) = [\mathbf{f}(n) - \mathbf{h}(n)]\hat{} + \mathbf{h}(n-1). \tag{4.38}$$

Although the adaptive filter may be updated directly as $\mathbf{h}(n) = \hat{\mathbf{f}}(n)$, in order to increase robustness to short-burst disturbances, the proposed AFC-CE method updates the adaptive filter according to

$$\mathbf{h}(n) = \lambda \mathbf{h}(n-1) + (1-\lambda)\hat{\mathbf{f}}(n), \tag{4.39}$$

where $0 \leq \lambda < 1$ is the factor that controls the trade-off between robustness and tracking rate of the adaptive filter.

In conclusion, the presented AFC-CE method calculates an estimate of $\mathbf{f}(n)$ from $\mathbf{c_e}(n)$ to update $H(q, n)$. Depending on the variations of $F(q, n)$ over time, it can be deduced that this computational effort may not be worth it, regarding performance, if the method is applied to each new sample of the microphone signal $y(n)$. Therefore, the AFC-CE will be applied every N_{fr} samples, where N_{fr} is a parameter that controls the trade-off between performance (latency and tracking capability) and computational complexity.

The block diagram of the proposed AFC-CE is depicted in Figure 4.5. Every N_{fr} samples, a frame of the error signal $e(n)$ containing its newest L_{fr} samples is selected; the frame has its spectrum $E(e^{j\omega}, n)$ and power cepstrum $\mathbf{c_e}(n)$ calculated using an N_{FFT_a}-point FFT; $\{\mathbf{g}(n) * \mathbf{d}\,[\mathbf{f}(n) - \mathbf{h}(n)]\}\hat{}$ is computed from $\mathbf{c_e}(n)$; with the knowledge of $\mathbf{d}$, $\{\mathbf{g}(n) * [\mathbf{f}(n) - \mathbf{h}(n)]\}\hat{}$ is calculated; with an estimate of $\mathbf{g}(n)$, $[\mathbf{f}(n) - \mathbf{h}(n)]\hat{}$ is computed; using $\mathbf{h}(n-1)$, $\hat{\mathbf{f}}(n)$ is calculated; finally, $\mathbf{h}(n)$ is updated.

4.4.3 Influence of Some Parameters and Improvements

This section analyzes the influence of some parameters on the performance of the proposed AFC-CM and AFC-CE methods. The assessed parameters were the cepstrum $\mathbf{c_u}(n)$ of the source input, the length L_D of the delay filter, the frame length L_{fr} , the use of smoothing windows (non-rectangular) in the frame selection of the microphone signal $y(n)$ and error signal $e(n)$, and the length L_F of the feedback path.

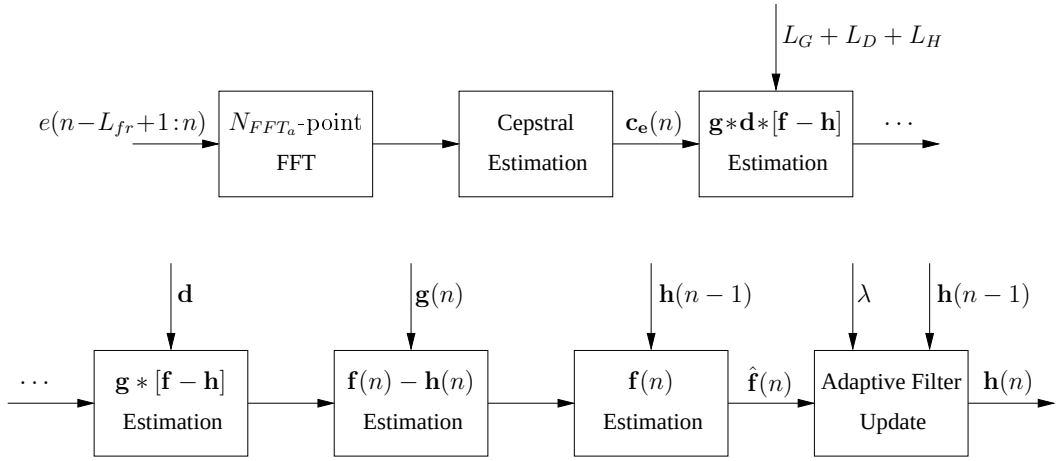


Figure 4.5: Block diagram of the proposed AFC-CE method.

4.4.3.1 Cepstrum of the System Input Signal and Delay Filter Length

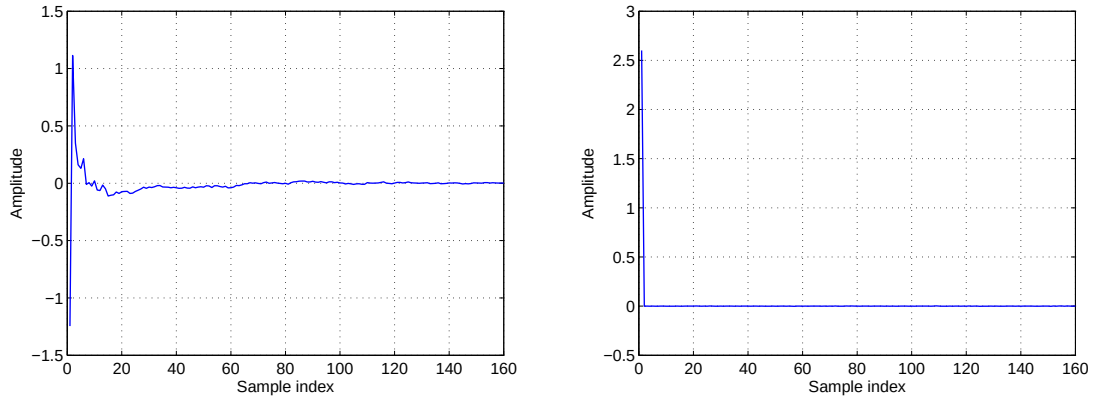
As explained in Sections 4.4.1 and 4.4.2, the AFC-CM and AFC-CE methods compute the estimates $\{\mathbf{g}(n) * \mathbf{f}(n)\}\hat{}$ and $\{\mathbf{g}(n) * [\mathbf{f}(n) - \mathbf{h}(n)]\}\hat{}$ from $\mathbf{c_y}(n)$ and $\mathbf{c_e}(n)$, respectively, resulting in (4.30) and (4.36). In both cases, the segment $\mathbf{c}_{\mathbf{u}_{L_D-1}}(n)$ from the cepstrum of the system input signal acts as estimation noise and would prevent these estimates from reaching their optimal solutions. Therefore, this section analyzes the influence that $\mathbf{c}_{\mathbf{u}_{L_D-1}}(n)$ may have in these estimates and, consequently, in the final performance of the AFC-CM and AFC-CE methods.

The estimates $\{\mathbf{g}(n) * \mathbf{f}(n)\}\hat{}$ and $\{\mathbf{g}(n) * [\mathbf{f}(n) - \mathbf{h}(n)]\}\hat{}$ are calculated by selecting the first $L_G + L_D + L_H - 2$ samples from $\mathbf{c_y}(n)$ and $\mathbf{c_e}(n)$, respectively. These procedures will be repeated for every frame of the microphone signal $y(n)$ and error signal $e(n)$. Considering $\mathbf{c_u}(n)$ as a random process with P realizations and $\mathbf{c_u}^{(p)}$ as the realization of the pth frame, the process mean value is defined as

$$\mathrm{E}\{\mathbf{c_u}(n)\} = \frac{1}{P}\sum_{p=1}^{P}\mathbf{c_u}^{(p)}, \tag{4.40}$$

where $\mathrm{E}\{\cdot\}$ is the statistical expectation operator.

The process mean value $\mathrm{E}\{\mathbf{c_u}(n)\}$ was computed according to (4.40) using frames with $L_{fr} = 8000$ and $N_{fr} = 1000$, and all the signals described in Section 4.6.6 as the system input signal u(n), resulting in $P \approx 3200$ realizations for both white noise and speech. Figure 4.6 shows the waveform of $\mathrm{E}\{\mathbf{c_u}(n)\}$ and it can be observed that its magnitude decreases with increasing sample index, in agreement with the cepstrum property of decaying at least as fast as $1/m$ where m is the sample index [70]. From Figure 4.6b, it follows that, when $u(n)$ is white noise, $\mathrm{E}\{\mathbf{c_u}(n)\}$ approaches its theoretical impulse-like waveform. When $u(n)$ is speech, it can be noticed from Figure 4.6a that the waveform of $\mathrm{E}\{\mathbf{c_u}(n)\}$ has a slower decay but $|\mathrm{E}\{\mathbf{c_u}(n)\}| < 1 \times 10^{-2}$ for $m > 80$.



(a) **(b)**

Figure 4.6: Waveform of $\mathrm{E}\{\mathbf{c}_{\mathbf{u}}(n)\}$ when $u(n)$ is: (a) speech; (b) white noise.

From (4.30) and (4.36), the estimate $\{\mathbf{g}(n) * \mathbf{f}(n)\}\hat{}$ in the AFC-CM and the estimate $\{\mathbf{g}(n) * [\mathbf{f}(n) - \mathbf{h}(n)]\}\hat{}$ in the AFC-CE are, on average, approximated as

$$\begin{aligned} \mathrm{E}\{\{\mathbf{g}(n) * \mathbf{f}(n)\}\hat{}\} &= \mathrm{E}\{\mathbf{g}(n) * \mathbf{f}(n) + \mathbf{c}_{\mathbf{u}_{L_D-1}}(n)\} \\ &= \mathrm{E}\{\mathbf{g}(n) * \mathbf{f}(n)\} + \mathrm{E}\{\mathbf{c}_{\mathbf{u}_{L_D-1}}(n)\} \end{aligned} \tag{4.41}$$

and

$$\begin{aligned} \mathrm{E}\{\{\mathbf{g}(n) * [\mathbf{f}(n) - \mathbf{h}(n)]\}\hat{}\} &= \mathrm{E}\{\mathbf{g}(n) * [\mathbf{f}(n) - \mathbf{h}(n)] + \mathbf{c}_{\mathbf{u}_{L_D-1}}(n)\} \\ &= \mathrm{E}\{\mathbf{g}(n) * [\mathbf{f}(n) - \mathbf{h}(n)]\} + \mathrm{E}\{\mathbf{c}_{\mathbf{u}_{L_D-1}}(n)\}, \end{aligned} \tag{4.42}$$

respectively.

In the literature, it is usually assumed that there is a delay of 25 ms in the cascade $D(q)G(q,n)$ [2, 3]. Considering only a delay of 10 ms caused by the delay filter $D(q)$ with $L_D = 161$ ($f_s = 16$ kHz), Figure 4.7 shows the waveform of $\mathrm{E}\{\mathbf{c}_{\mathbf{u}_{L_D-1}}(n)\}$ in the region where the values of $\mathbf{g}(n) * \mathbf{f}(n)$ and $\mathbf{g}(n) * [\mathbf{f}(n) - \mathbf{h}(n)]$ will be located. In this range, $|\mathrm{E}\{\mathbf{c}_{\mathbf{u}_{L_D-1}}(n)\}| < 6 \times 10^{-3}$ and $|\mathrm{E}\{\mathbf{c}_{\mathbf{u}_{L_D-1}}(n)\}| < 2 \times 10^{-3}$ when the source signal $u(n)$ is speech and white noise, respectively. These low values when $u(n)$ is white noise were expected because the cepstrum has, in theory, an impulse-like waveform. But, when the system input signal $u(n)$ is speech, the low values are quite interesting especially considering the diversity of 10 talkers and 4 languages used.

Although the values of $\mathrm{E}\{\mathbf{c}_{\mathbf{u}_{L_D-1}}(n)\}$ are small, their relative influence will depend on their ratio to the values of $\mathrm{E}\{\mathbf{g}(n) * \mathbf{f}(n)\}$ and $\mathrm{E}\{\mathbf{g}(n) * [\mathbf{f}(n) - \mathbf{h}(n)]\}$. For a better understanding, the influence of $\mathrm{E}\{\mathbf{c}_{\mathbf{u}_{L_D-1}}(n)\}$ will be analyzed separately for the AFC-CM and AFC-CE methods.

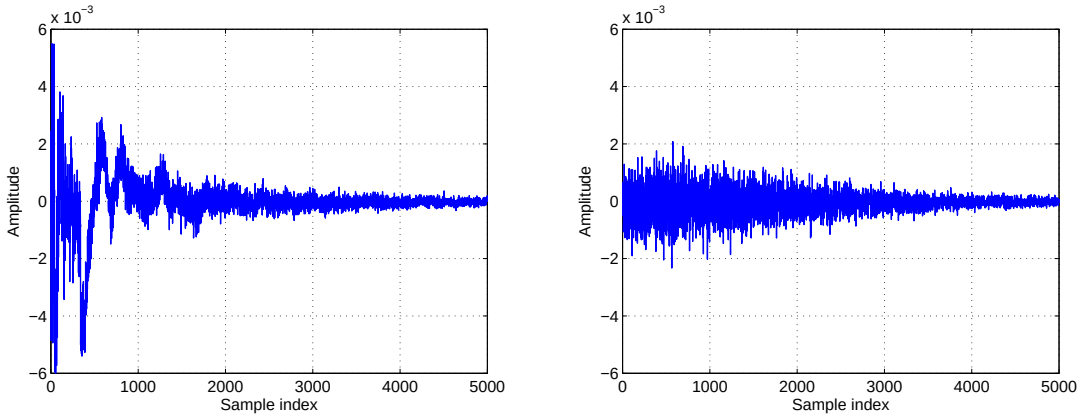


Figure 4.7: Waveform of $\mathrm{E}\{\mathbf{c}_{\mathbf{u}_{L_D-1}}(n)\}$ when $L_D = 161$ and $u(n)$ is: (a) speech; (b) white noise.

AFC-CM Method The influence of $\mathrm{E}\{\mathbf{c}_{\mathbf{u}_{L_D-1}}(n)\}$ on $\mathrm{E}\{\mathbf{g}(n)*\mathbf{f}(n)\}$ will be analyzed as a function of $\mathbf{f}(n)$ and $\mathbf{g}(n)$. The feedback path $F(q,n)$ will be the room impulse response shown in Figure 3.3 and the forward path $G(q,n)$ will be a gain such that $\left|G(e^{j\omega},n)\right| = \left[\max_w \left|D(e^{j\omega})F(e^{j\omega},n)\right|\right]^{-1}$.

Since the magnitude of $\mathbf{f}(n)$, as a room impulse response, typically decays exponentially with increasing sample index, the magnitude of $\mathbf{g}(n)*\mathbf{f}(n)$ also decays exponentially. And since the magnitude of $\mathrm{E}\{\mathbf{c}_{\mathbf{u}_{L_D-1}}(n)\}$ also decays with increasing sample index, as showed in Figure 4.6, the influence of $\mathrm{E}\{\mathbf{c}_{\mathbf{u}_{L_D-1}}(n)\}$ on $\mathbf{g}(n)*\mathbf{f}(n)$ will depend on the decay speed of both curves. The relative influence of $\mathrm{E}\{\mathbf{c}_{\mathbf{u}_{L_D-1}}(n)\}$ on $\mathbf{g}(n)*\mathbf{f}(n)$ can be measured by the ratio

$$\mathbf{r}_{L_D-1}(n) = 20\log_{10}\frac{|\mathrm{E}\{\mathbf{c}_{\mathbf{u}_{L_D-1}}(n)\}|}{|\mathbf{g}(n)*\mathbf{f}(n)|}. \tag{4.43}$$

Disregarding the samples related to the initial delay of $\mathbf{f}(n)$, $\mathbf{r}_{L_D-1}(n)$ is represented in Figure 4.8 for $L_D = 401$, the delay filter length that will be used in this work as in [2, 3], along with its linear approximation when the system input signal $u(n)$ is speech and white noise. It can be observed that $\mathbf{r}_{L_D-1}(n)$ increases with increasing sample index, which means that $\mathbf{f}(n)$ decays faster than $\mathrm{E}\{\mathbf{c}_{\mathbf{u}_{L_D-1}}(n)\}$.

In the initial samples, the influence of $\mathrm{E}\{\mathbf{c}_{\mathbf{u}_{L_D-1}}(n)\}$ can be considered negligible. This is an advantage characteristic of the AFC-CM method because they are the samples of $\mathbf{f}(n)$ with the highest absolute values and thus have the largest contribution to the acoustic feedback problem. But above a certain sample index, $|\mathrm{E}\{\mathbf{c}_{\mathbf{u}_{L_D-1}}(n)\}|$ becomes higher than $|\mathbf{g}(n)*\mathbf{f}(n)|$, which makes the estimation of $\mathbf{g}(n)*\mathbf{f}(n)$ from $\mathbf{c}_{\mathbf{y}}(n)$ impossible. In general, this increase in $\mathbf{r}_{L_D-1}(n)$ causes the AFC-CM method to have more difficulty, or even impossibility, in estimating the tail of $\mathbf{f}(n)$ as will be demonstrated in Section 4.4.3.4. But, fortunately, the lower absolute values of $\mathbf{f}(n)$ have a smaller contribution to the acoustic feedback and then this drawback of the AFC-CM is not so critical.

The linear approximation of $\mathbf{r}_{L_D-1}(n)$ for different values of L_D is shown in Figure 4.9.

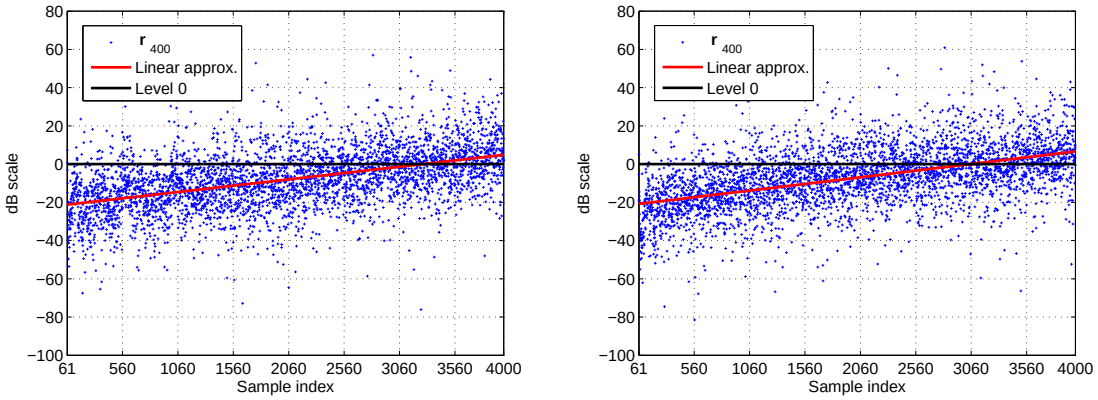


Figure 4.8: Ratio $\mathbf{r}_{L_D-1}(n)$ when $L_D = 401$ and $u(n)$ is: (a) speech; (b) white noise.

It can be observed that the values of $\mathbf{r}_{L_D-1}(n)$ decrease as L_D increases, more so for speech signals. The delay filter $D(q)$ shifts the values of $\mathbf{g}(n) * \mathbf{f}(n)$ to the right on the sample axis in (4.16) such that the amount of shifting increases with L_D. Because of the cepstrum decay, shifting $\mathbf{g}(n) * \mathbf{f}(n)$ to the right on the sample axis means shifting it towards the lower magnitudes of $\mathbf{c}_{\mathbf{u}}(n)$. Therefore, the estimation of $\mathbf{g}(n) * \mathbf{f}(n)$ may be improved by increasing L_D. Nevertheless, for all the evaluated values of L_D, the conclusions of the previous paragraph remain valid.

The gain $\mathbf{g}(n)$ determines the offset of $\mathbf{r}_{L_D-1}(n)$ and its linear approximation. An increase in $\mathbf{g}(n)$ causes both curves to slide downward and thus decreases the influence of $\mathrm{E}\{\mathbf{c}_{\mathbf{u}_{L_D-1}}(n)\}$ on $\mathbf{g}(n) * \mathbf{f}(n)$. However, as explained in detail in Section 4.3.1, the conditions $|G(e^{j\omega}, n)D(e^{j\omega})H(e^{j\omega}, n)| < 1$ and $|G(e^{j\omega}, n)D(e^{j\omega})\left[F(e^{j\omega}, n) - H(e^{j\omega}, n)\right]| < 1$ are required in order for $\mathbf{c}_{\mathbf{y}}(n)$ to be defined according to (4.16). The former condition is initially fulfilled and ultimately becomes $|G(e^{j\omega}, n)D(e^{j\omega})F(e^{j\omega}, n)| < 1$. The latter is the NGC of the AFC system and is sufficient to ensure system stability. Therefore, the influence of $\mathrm{E}\{\mathbf{c}_{\mathbf{u}_{L_D-1}}(n)\}$ will always limit the performance of the proposed AFC-CM method but is ultimately minimized when $\left|G(e^{j\omega}, n)\right| = \left|D(e^{j\omega})F(e^{j\omega}, n)\right|^{-1}$. When the forward path $G(q, n)$ is only a gain, this gain value is precisely the one that resulted in the influence of $\mathrm{E}\{\mathbf{c}_{\mathbf{u}_{L_D-1}}(n)\}$ shown in Figures 4.8 and 4.9.

Therefore, for the samples of $\mathbf{g}(n) * \mathbf{f}(n)$ with the highest absolute values, which are the most important ones, the influence of $\mathbf{c}_{\mathbf{u}}(n)$ can be made negligible over time by making $\left|G(e^{j\omega}, n)\right| = \left|D(e^{j\omega})F(e^{j\omega}, n)\right|^{-1}$. Consequently, for these samples, (4.41) can be approximated as

$$\mathrm{E}\{\{\mathbf{g}(n) * \mathbf{f}(n)\}\hat{}\} \approx \mathrm{E}\{\mathbf{g}(n) * \mathbf{f}(n)\}, \tag{4.44}$$

which means that the estimation of $\mathbf{g}(n)*\mathbf{f}(n)$ from $\mathbf{c}_{\mathbf{y}}(n)$ will be asymptotically consistent because it tends to reach the optimal solution.

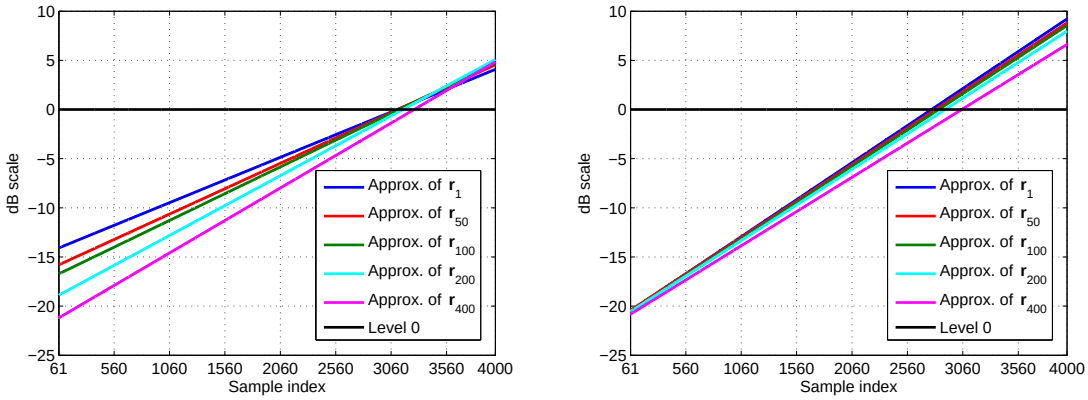


Figure 4.9: Linear approximations of the ratio $\mathbf{r}_{L_D-1}(n)$ for different values of L_D when $u(n)$ is: (a) speech; (b) white noise.

AFC-CE Method In the AFC-CE method, the influence of $\mathrm{E}\{\mathbf{c}_{\mathbf{u}_{L_D-1}}(n)\}$ on $\mathrm{E}\{\mathbf{g}(n) * [\mathbf{f}(n) - \mathbf{h}(n)]\}$ will be analyzed as a function of $\mathbf{f}(n) - \mathbf{h}(n)$ and $\mathbf{g}(n)$. Again, the feedback path $F(q,n)$ will be the room impulse response shown in Figure 3.3 and the forward path $G(q,n)$ will be a gain such that $\left|G(e^{j\omega},n)\right| = \left[\max_w \left|D(e^{j\omega})F(e^{j\omega},n)\right|\right]^{-1}$.

Although the magnitude of $\mathbf{f}(n)$ typically decays exponentially with increasing sample index, the magnitude behavior of $\mathbf{g}(n) * [\mathbf{f}(n) - \mathbf{h}(n)]$ depends on $\mathbf{h}(n)$. The relative influence of $\mathrm{E}\{\mathbf{c}_{\mathbf{u}_{L_D-1}}(n)\}$ on $\mathbf{g}(n) * [\mathbf{f}(n) - \mathbf{h}(n)]$ can be measured by the ratio

$$\mathbf{r}_{\mathbf{2}L_D-1}(n) = 20\log_{10}\frac{|\mathrm{E}\{\mathbf{c}_{\mathbf{u}_{L_D-1}}(n)\}|}{|\mathbf{g}(n) * [\mathbf{f}(n) - \mathbf{h}(n)]\,|}. \tag{4.45}$$

Consider that the adaptive filter $H(q,n)$ is initialized with zeros and converges to $F(q,n)$ over time, i.e., $H(q,0) = 0$ and $H(q,n) \to F(q,n)$ as $n \to \infty$. When $n = 0$, $\mathbf{r}_{\mathbf{2}L_D-1}(n) = \mathbf{r}_{L_D-1}(n)$ and the relative influence of $\mathrm{E}\{\mathbf{c}_{\mathbf{u}_{L_D-1}}(n)\}$ on the AFC-CE method will be the same as on the AFC-CM method, which was discussed in detail in the previous section. In the first few seconds of operation of the AFC-CE method, $\mathbf{r}_{\mathbf{2}L_D-1}(n) \approx \mathbf{r}_{L_D-1}(n)$ because $\mathbf{h}(n)$ has very low values.

In the same way as with $\mathbf{r}_{L_D-1}(n)$, the gain $\mathbf{g}(n)$ determines the offset of $\mathbf{r}_{\mathbf{2}L_D-1}(n)$ and its linear approximation. Consider now that, in the course of time n, $\mathbf{g}(n)$ can be increased and the samples of $\mathbf{h}(n)$ converge in proportion to the samples of $\mathbf{f}(n)$. The former situation shifts $\mathbf{r}_{\mathbf{2}L_D-1}(n)$ downward and the latter shifts it upward. If $\mathbf{g}(n)$ remains unchanged as $\mathbf{h}(n)$ converges to $\mathbf{f}(n)$, $\mathbf{r}_{\mathbf{2}L_D-1}(n)$ will be shifted upward and thus the influence of $\mathrm{E}\{\mathbf{c}_{\mathbf{u}_{L_D-1}}(n)\}$ will increase. But if $\mathbf{g}(n)$ increases as $\mathbf{h}(n)$ converges to $\mathbf{f}(n)$, $\mathbf{g}(n)$ may compensate the upward shifting, that would be caused by $\mathbf{h}(n)$, by making the samples of $\mathbf{g}(n) * [\mathbf{f}(n) - \mathbf{h}(n)]$ constant over time n. However, the condition $\left|G(e^{j\omega},n)D(e^{j\omega})\left[F(e^{j\omega},n) - H(e^{j\omega},n)\right]\right| < 1$ is required in order for $\mathbf{c}_{\mathbf{e}}(n)$ to be defined according to (4.23). Therefore, the influence of $\mathrm{E}\{\mathbf{c}_{\mathbf{u}_{L_D-1}}(n)\}$ will

always limit the performance of the proposed AFC-CE method but is minimized when $\left|G(e^{j\omega},n)\right| = \left|D(e^{j\omega})\left[F(e^{j\omega},n) - H(e^{j\omega},n)\right]\right|^{-1}$, i.e., the system is at the stability limit. When the forward path $G(q,n)$ is only a gain and the samples of $\mathbf{h}(n)$ converge in proportion to the samples of $\mathbf{f}(n)$, as it was assumed, this gain value results in the influence of $\mathrm{E}\{\mathbf{c}_{\mathbf{u}_{L_D-1}}(n)\}$ shown in Figures 4.8 and 4.9.

Therefore, for the samples of $\mathbf{g}(n) * [\mathbf{f}(n) - \mathbf{h}(n)]$ with the highest absolute values, which are the most important ones, the influence of $\mathbf{c}_{\mathbf{u}}(n)$ can be made negligible in the course of time n by increasing the gain of the forward path $G(q,n)$ as $H(q,n)$ converges to $F(q,n)$. Consequently, for these samples, (4.42) can be approximated as

$$\mathrm{E}\{\{\mathbf{g}(n) * [\mathbf{f}(n) - \mathbf{h}(n)]\}\hat{}\} \approx \mathrm{E}\{\mathbf{g}(n) * [\mathbf{f}(n) - \mathbf{h}(n)]\}, \tag{4.46}$$

which means that the estimation of $\mathbf{g}(n) * [\mathbf{f}(n) - \mathbf{h}(n)]$ from $\mathbf{c}_{\mathbf{e}}(n)$ will be asymptotically consistent because it tends to reach the optimal solution.

4.4.3.2 Frame Length

In linear system identification, the system impulse response is usually estimated through its input and output signals. It is precisely the case of the AFC methods based on the traditional adaptive filtering algorithms, as the PEM-AFROW method, that estimate the impulse response $\mathbf{f}(n)$ of the feedback path by considering the loudspeaker signal $x(n)$ as its input and the microphone signal $y(n)$ as its output. The bias problem in the identification occurs because $y(n)$, in addition to the real output signal $x(n) * \mathbf{f}(n)$, also contains the system input signal $u(n)$ that is strongly correlated to $x(n)$.

The cepstral analysis also estimates the impulse responses in (4.16) and (4.23) through their input and output signals. But, in this case, they are not considered separately. Instead, the cepstral analysis uses only the microphone signal $y(n)$ or error signal $e(n)$. Therefore, in order for the cepstral analysis to be able to estimate the impulse responses in (4.16) and (4.23), it is necessary that their input and output signals are jointly contained in the frame of the microphone signal $y(n)$ and error signal $e(n)$, respectively. It is worth mentioning that the input signal of an impulse response is not restricted to the system input signal $u(n)$ and can include feedback samples from previous cycles.

Figure 4.10 depicts the block processing of a filtering operation according to the overlap-and-add procedure. Consider, for illustration, the AFC-CM method and $\mathbf{b}$ as the k-fold impulse response present in (4.16) such that $L_B = k \times (L_G + L_D + L_F - 3) + 1$. Its input and output signals are defined as $y_i(n)$ and $y_o(n)$, respectively, such that the microphone signal $y(n) = y_i(n) + y_o(n)$. At the discrete-time n, the AFC-CM method selects the frame

$$\mathbf{y}_2 = [y(n - L_{fr} + 1) \;\; y(n - L_{fr} + 2) \;\; \ldots \;\; y(n)]^T \tag{4.47}$$

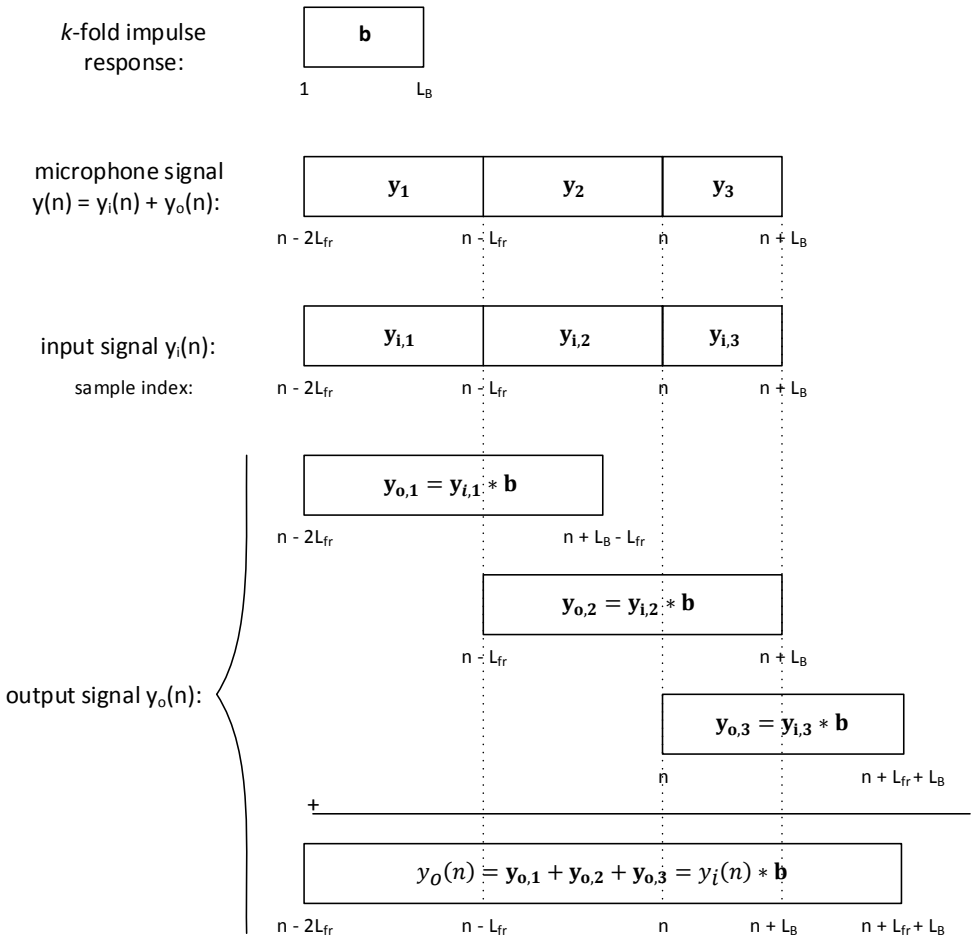


Figure 4.10: Block processing of a filtering operation according to the overlap-and-add procedure.

of the microphone signal $y(n)$.

It can be observed that the frame $\mathbf{y}_2$ does not contain all the output signal $\mathbf{y}_{o,2}$, generated from the input signal $\mathbf{y}_{i,2}$, because its convolution tail is disregarded. Then, the frame $\mathbf{y}_2$ contains the input signal $\mathbf{y}_{i,2}$ but not the last L_B samples of its output signal $\mathbf{y}_{o,2}$. On the other hand, the tail of the output signal $\mathbf{y}_{o,1}$, generated from the previous input signal $\mathbf{y}_{i,1}$, is included in the frame $\mathbf{y}_2$. Therefore, the frame $\mathbf{y}_2$ does not contain the input signal $\mathbf{y}_{i,1}$ but does the last L_B samples of its output signal $\mathbf{y}_{o,1}$. These facts degrade the estimate of the k-fold impulse response present in (4.16) provided by the cepstral analysis. The same occurs for the AFC-CE method considering the error signal $e(n)$.

As the goal of the proposed AFC-CM and AFC-CE methods is to estimate the 1-fold impulse response from (4.16) and (4.23), respectively, only the two sample blocks with length $L_B = L_G + L_D + L_F - 2$, one at the beginning and another at the end of the selected frame, can disturb the method performance. For fixed $L_G + L_D + L_F$, increasing L_{fr} increases the amount of useful samples of the frame and thus reduces the influence of

these two sample blocks until they become irrelevant. As a consequence, the estimates of the 1-fold impulse responses may improve.

However, the conclusion that increasing L_{fr} may improve the estimation of the 1-fold impulse responses from (4.16) and (4.23) can be ensured if the impulse responses were time-invariant throughout the frame length L_{fr}. If they are time-varying, the cepstral analysis will estimate an average of the 1-fold impulse responses over the frame length L_{fr}. Then, in this case, increasing L_{fr} may give a lower weight to the current values of the impulse responses and thus worsen their estimates. Therefore, for time-varying 1-fold impulse responses in (4.16) and (4.23), the frame length L_{fr} controls the trade-off between the amount of useful samples provided for the cepstral analysis and the weight given by the cepstral analysis to the current impulse responses.

4.4.3.3 Smoothing Window and High-Pass Filtering

Simulations showed that, when the source signal $v(n)$ is speech and the signal-to-noise ratio (SNR) is particularly high, the resulting $|H(e^{j\omega}, n)|$ computed by the proposed AFC-CM and AFC-CE methods may have values considerably higher than $|F(e^{j\omega}, n)|$ at the low-frequency components (below 100 Hz). These high values of $|H(e^{j\omega}, n)|$ at the low-frequency components may insert distortion in the system signals and adversely affect the stability of the AFC system.

This effect at low-frequency components of $|H(e^{j\omega}, n)|$ becomes even more severe if smoothing windows are used to select the frame of the microphone signal $y(n)$ and error signal $e(n)$ instead of a rectangular window. On the other hand, at the remaining frequency components, the use of smoothing windows usually improves $|H(e^{j\omega}, n)|$, provided by the proposed methods, with respect to $|F(e^{j\omega}, n)|$. Both issues are illustrated in Figure 4.11 for Blackman and rectangular windows.

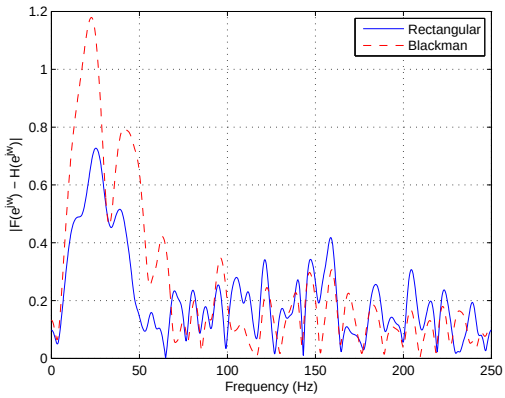


Figure 4.11: Illustration of the increase in the low-frequency components of $\left|F(e^{j\omega,n}) - H(e^{j\omega}, n)\right|$ due to the use of smoothing windows.

However, it is possible to overcome the undesirable effect at low-frequency components of $|H(e^{j\omega}, n)|$ and still benefit from the improvement in the remaining frequency components of $|H(e^{j\omega}, n)|$ caused by the use of smoothing windows. To this purpose, a Blackman window will be used to select the frames of $y(n)$ and $e(n)$, and the frequency components below 100 Hz in $G(q, n)D(q)\,[F(q, n) - H(q, n)]$, the open-loop transfer function of the AFC system, and $H(q, n)$ will be attenuated by a linear-phase highpass filter $B(q)$ designed with the Parks-McClellan algorithm.

It was verified that a highpass filter $B(q)$ with length $L_B = 801$ samples fulfills the necessary requirements of frequency response and, at the same time, generates a time delay similar to a delay filter $D(q)$ with $L_D = 401$ samples as used in [2, 3]. The specifications of the highpass filter $B(q)$ are $L_B = 801$ samples, stopband and passband edge frequencies of 90 and 120 Hz, stopband and passband ripples of 23.3 and 0.1 dB. Its impulse response and frequency response magnitude are shown in Figure 4.12.

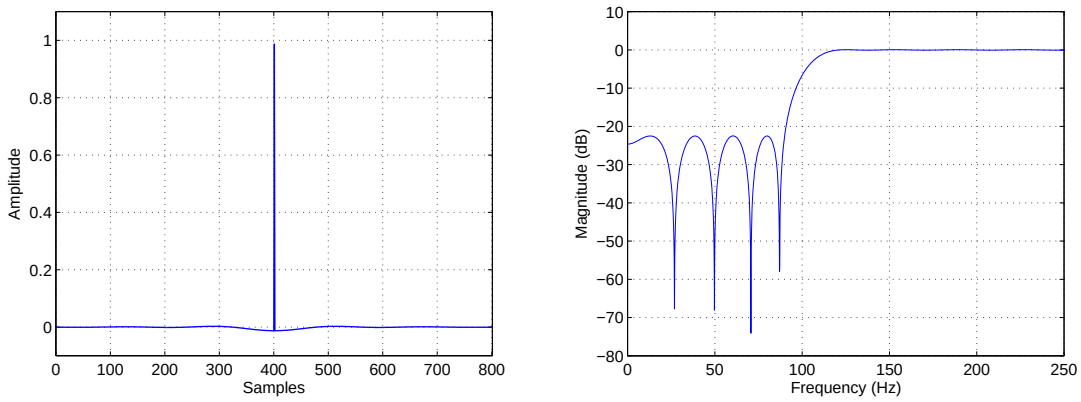


Figure 4.12: High-pass filter $B(q)$: (a) impulse response; (b) frequency response.

The frequency components below 100 Hz in the open-loop transfer function of the AFC system were attenuated by replacing the delay filter $D(q)$ with the highpass filter $B(q)$. Hence, the delay filter $D(q)$, its length L_D, its impulse response $\mathbf{d}$ and its frequency response $D(e^{j\omega})$ present in equations as well as in discussions of previous sections must be replaced, respectively, by $B(q)$, L_B, $\mathbf{b}$ and $B(e^{j\omega})$. It should be noted that L_B cannot be too small in order for $B(q)$ to fulfill the necessary requirements to adequately attenuate the low-frequency components. Therefore, by replacing $D(q)$ with $B(q)$, the time delay of the open-loop transfer function cannot be too small. For instance, a highpass filter $B(q)$ with same passband edge frequency and that generates a time delay of 10 ms ($L_B = 161$) has already been used without significant loss in performance.

It can be seen from Figure 4.12a that the impulse response $\mathbf{b}$ of the highpass filter has some very low values around its maximum absolute value. Since $\mathbf{c_u}(n)$ acts as estimation noise, the effect of these low values on $\mathbf{g}(n) * \mathbf{b} * \mathbf{f}(n)$ and $\mathbf{g}(n) * \mathbf{b} * [\mathbf{f}(n) - \mathbf{h}(n)]$ cannot

be accurately obtained from $\mathbf{c_y}(n)$ and $\mathbf{c_e}(n)$, respectively. As a consequence, in the AFC-CM mehod, the estimate $\{\mathbf{g}(n) * \mathbf{b} * \mathbf{f}(n)\}\hat{}$ calculated from $\mathbf{c_y}(n)$ according to (4.27) is really closer to $\mathbf{g}(n) * \mathbf{d} * \mathbf{f}(n)$ than to $\mathbf{g}(n) * \mathbf{b} * \mathbf{f}(n)$. Similarly, in the AFC-CE method, the estimate $\{\mathbf{g}(n) * \mathbf{b} * [\mathbf{f}(n) - \mathbf{h}(n)]\}\hat{}$ calculated from $\mathbf{c_e}(n)$ according to (4.34) is really closer to $\mathbf{g}(n) * \mathbf{d} * [\mathbf{f}(n) - \mathbf{h}(n)]$ than to $\mathbf{g}(n) * \mathbf{b} * [\mathbf{f}(n) - \mathbf{h}(n)]$. Hence, (4.29) and (4.35) are actually performed as

$$\{\mathbf{g}(n) * \mathbf{f}(n)\}\hat{} = \{\mathbf{g}(n) * \mathbf{b} * \mathbf{f}(n)\}\hat{} * \mathbf{d}^{-1} \tag{4.48}$$

and

$$\{\mathbf{g}(n) * [\mathbf{f}(n) - \mathbf{h}(n)]\}\hat{} = \{\mathbf{g}(n) * \mathbf{b} * [\mathbf{f}(n) - \mathbf{h}(n)]\}\hat{} * \mathbf{d}^{-1}, \tag{4.49}$$

respectively.

And the frequency components below 100 Hz in $H(q, n)$ were attenuated by performing

$$\{\mathbf{g}(n) * \mathbf{f}(n)\}\hat{} = \{\mathbf{g}(n) * \mathbf{f}(n)\}\hat{} * \mathbf{b} * \mathbf{d}^{-1} \tag{4.50}$$

before feeding it to (4.32) in the AFC-CM method and

$$\{\mathbf{g}(n) * [\mathbf{f}(n) - \mathbf{h}(n)]\}\hat{} = \{\mathbf{g}(n) * [\mathbf{f}(n) - \mathbf{h}(n)]\}\hat{} * \mathbf{b} * \mathbf{d}^{-1} \tag{4.51}$$

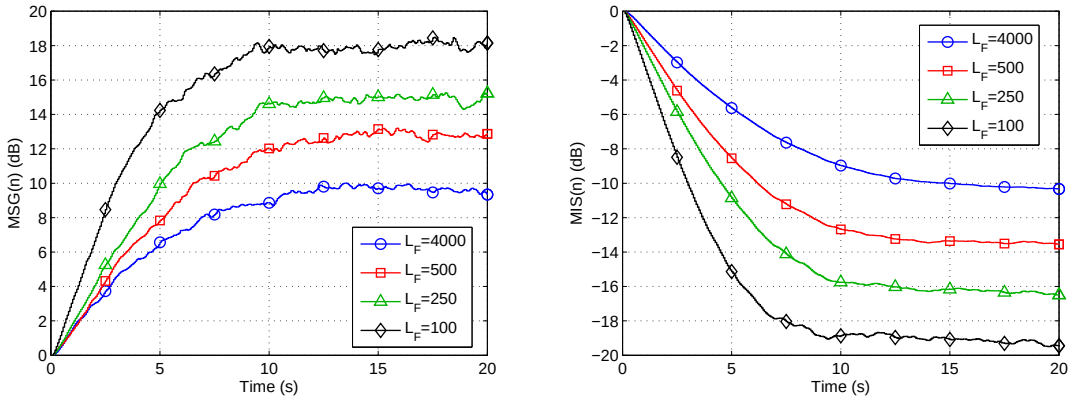
before feeding it to (4.37) in the AFC-CE method.

4.4.3.4 Length of the Feedback Path

As discussed in Section 4.4.3.1, the cepstrum $\mathbf{c_u}(n)$ of the system input signal acts as noise in the estimation of $\mathbf{g}(n) * \mathbf{f}(n)$ and $\mathbf{g}(n) * [\mathbf{f}(n) - \mathbf{h}(n)]$ from $\mathbf{c_y}(n)$ and $\mathbf{c_e}(n)$, respectively. And, on average, the influence of $\mathbf{c_u}(n)$ on these estimations increases with increasing sample index, which makes the proposed AFC-CM and AFC-CE methods have more difficulty in estimating the lower absolute values, mainly the tail, of the impulse response $\mathbf{f}(n)$ of the feedback path.

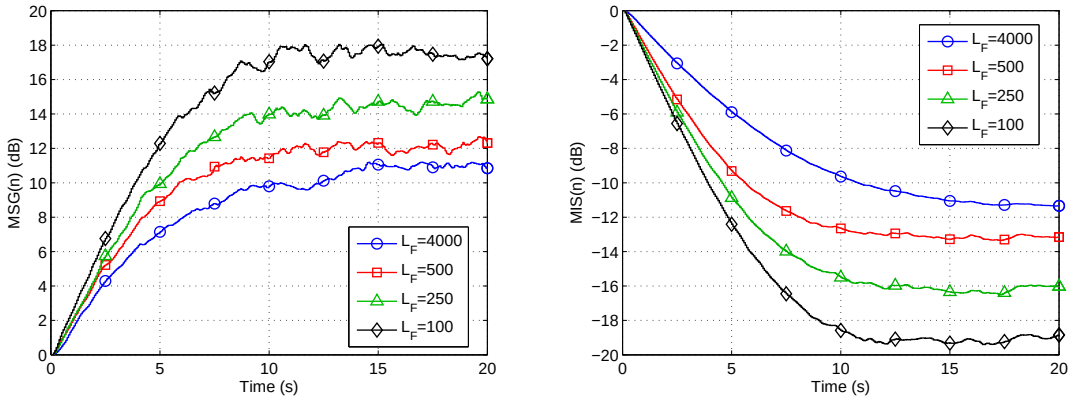
This section aims to illustrate this characteristic of the AFC-CM and AFC-CE methods with a practical example. To this end, simulations were carried out using the configuration of the PA system, the evaluation metrics and the signals described in Section 4.6. The broadband gain $K(n)$ of the forward path was constant over time ($\Delta K = 0$) and the impulse response $\mathbf{f}(n)$ of the feedback path was truncated to length $L_F = 100, 250, 500$ and 4000 samples.

For simplicity, only the results obtained by the AFC-CE method will be illustrated because the AFC-CM presents similar variations in performance as a function of L_F. The AFC-CE method started only after 125 ms of simulation to avoid initial inaccurate estimates, $L_{fr} = 8000$, $N_{fr} = 1000$, $N_{FFT_a} = 2^{15}$ and $N_{FFT_e} = 2^{17}$. The optimization of



(a) **(b)**

Figure 4.13: Influence of L_F on the performance of the AFC-CE method when $u(n)$ is white noise: (a) MSG(n); (b) MIS(n).



(a) **(b)**

Figure 4.14: Influence of L_F on the performance of the AFC-CE method when $u(n)$ is speech: (a) MSG(n); (b) MIS(n).

its adaptive filter parameters λ and L_H was performed identically to the PEM-AFROW method and is described in Section 3.7.

Figures 4.13 and 4.14 show the results obtained by the AFC-CE method when the system input signal $u(n)$ is white noise and speech, respectively. It can be observed that the performance of the AFC-CE method gets worse as the impulse response $\mathbf{f}(n)$ of the feedback path gets longer. This is due to its difficulty in estimating the lower absolute values of $\mathbf{f}(n)$ as explained in Section 4.4.3.1. And, since the magnitude of $\mathbf{f}(n)$ decays with increasing sample index, lower absolute values are included in $\mathbf{f}(n)$ as its length L_F increases, thereby worsening the performance of the AFC-CE method. These results confirm in practice the analysis presented in Section 4.4.3.1. This characteristic makes the proposed AFC-CM and AFC-CE methods even more suitable to deal with acoustic feedback paths with short tails such as occur in hearing aid applications.

4.5 Computational Complexity

In this section, the computational complexity of the proposed AFC-CM and AFC-CE methods is calculated considering one multiplication and one addition as two separate floating-point operations. As the computational complexity of both methods is very similar, its calculation will be based on the AFC-CE method. For the AFC-CM, a similar procedure can be performed and results in more L_H real multiplications.

The selection of the L_{fr}-length frame of the error signal $e(n)$ requires L_{fr} multiplications. In order to compute the power cepstrum $\mathbf{c_e}(n)$, it is necessary to compute the spectrum $E(e^{j\omega}, n)$ of the selected frame, $|E(e^{j\omega}, n)|^2$, its natural logarithm and convert the result to the time domain. The computation of $E(e^{j\omega}, n)$ is performed through an N_{FFT_a}-point FFT which, considering the radix-2 algorithm and a real signal, requires $\frac{N_{FFT_a}}{2}\log_2 N_{FFT_a} - \frac{3}{2}N_{FFT_a} + 2$ complex multiplications and $N_{FFT_a}\log_2 N_{FFT_a}$ complex additions [72]. This results in $2N_{FFT_a}\log_2 N_{FFT_a} - 6N_{FFT_a} + 8$ real multiplications and $3N_{FFT_a}\log_2 N_{FFT_a} - 3N_{FFT_a} + 4$ real additions. The computation of $|E(e^{j\omega}, n)|^2$ requires N_{FFT_a} real multiplications and $\frac{N_{FFT_a}}{2}$ real additions while its natural logarithm needs $\frac{N_{FFT_a}}{2}$ real multiplications and $\frac{N_{FFT_a}}{2}$ real additions when using lookup tables [73]. The conversion of the result to the time domain is performed through an N_{FFT_a}-point Inverse FFT (IFFT), which requires $2N_{FFT_a}\log_2 N_{FFT_a} - 6N_{FFT_a} + 8$ real multiplications and $3N_{FFT_a}\log_2 N_{FFT_a} - 3N_{FFT_a} + 4$ real additions.

The convolution with $\mathbf{d}^{-1}$ in (4.49) and (4.51) is simply performed by sliding on the time axis. Considering $M_1 = L_G + L_H - 1$, the convolution with $\mathbf{g}^{-1}(n)$ in (4.37) can be performed in the frequency domain using two M_1-point FFTs, M_1 complex divisions and one M_1-point IFFT, requiring $6M_1\log_2 M_1 - 10M_1 + 24$ real multiplications and $9M_1\log_2 M_1 - 6M_1 + 12$ real additions. Note that if $G(q, n)$ is only a gain and a delay, only 1 real multiplication is required.

Considering $M_2 = L_H + L_B - 1$ and an M_2-point FFT of $B(q)$ previously computed, the convolution with $\mathbf{b}$ in (4.51) can be performed in the frequency domain using one M_2-point FFT, M_2 complex multiplications and one M_2-point IFFT, requiring $4M_2\log_2 M_2 - 8M_2 + 16$ real multiplications and $6M_2\log_2 M_2 - 4M_2 + 8$ real additions. Finally, (4.38) and (4.39) can be effectively combined to need L_H real multiplications and L_H real additions.

In conclusion, the proposed AFC-CE method requires

$$\mathcal{O} = \frac{1}{N_{fr}} \times \left[\left(10N_{FFT_a}\log_2 N_{FFT_a} - \frac{31}{2}N_{FFT_a} + 24 \right) + (15M_1\log_2 M_1 - 16M_1 + 36) + (10M_2\log_2 M_2 - 12M_2 + 24) + L_{fr} + 2L_H \right] \quad (4.52)$$

floating-point operations per iteration. Since $L_H \ll \mathcal{O} \times N_{fr}$, it can be considered that the AFC-CM method has the same computational complexity. Considering $N_{FFT_a} = 2^{15}$, $G(q, n)$ defined as (3.42), $L_H = 4000$, $L_B = 801$, $L_{fr} = 8000$ and $N_{fr} = 1000$, the

AFC-CM and AFC-CE methods require approximately 4952 floating-point operations per iteration. In comparison, with the parameter values originally proposed in [3] adjusted to $f_s = 16$ kHz and $L_H = 4000$, the PEM-AFROW method requires approximately 34000 floating-point operations per iteration.

Keeping the values of the other parameters unchanged, the computational complexity of both methods is similar if the AFC-CE or AFC-CM are applied every $N_{fr} = 145$ samples (equivalent to 9 and 3.3 ms for $f_s = 16$ and 44.1 kHz, respectively). This possible latency should not have great influence on the performance of the AFC-CE and AFC-CM methods because the variations of $F(q, n)$ in the meantime should be small.

4.6 Simulation Configurations

With the aim to evaluate the performance of the proposed AFC-CM and AFC-CE methods, an experiment was carried out in a simulated environment to measure their ability to estimate the feedback path impulse response and increase the MSG of a PA system. The resulting distortion in the error signal $e(n)$ was also measured. To this purpose, the following configuration was used.

4.6.1 Simulated Environment

The simulated environment was the same as used for the PEM-AFROW method and included two different configurations of the forward path $G(q, n)$. In the first, the broadband gain $K(n)$ remained constant, i.e., $\Delta K = 0$ and the system had an initial gain margin of 3 dB. In the second, $K(n)$ was increased in order to determine the MSBG achievable by the AFC methods. A complete description can be found in Section 3.6.1.

4.6.2 Maximum Stable Gain

As discussed in Section 3.6.2, the main goal of any AFC method is to increase the MSG of the PA system that has an upper limit due to the acoustic feedback. Therefore, the MSG is the most important metric in evaluating AFC methods.

The proposed AFC-CM and AFC-CE methods, as the PEM-AFROW, do not apply any processing to the signals that travel in the system other than the adaptive filter $H(q, n)$. Then, the MSG of the AFC system and the increase in MSG achieved by the AFC-CM or AFC-CE, ΔMSG, were measured according to (3.6) and (3.8), respectively.

The frequency responses were also computed using an N_{FFT_e}-point FFT with $N_{FFT_e} = 2^{17}$. The sets of critical frequencies $P(n)$ and $P_H(n)$ were obtained by searching, in the corresponding unwrapped phase, each crossing by integer multiples of 2π. A detailed explanation can be found in Section 3.6.2.

4.6.3 Misalignment

In addition to the MSG, the performance of the proposed AFC-CM and AFC-CE methods were also evaluated through the normalized misalignment (MIS) metric. The MIS(n) measures the mismatch between the adaptive filter and the feedback path according to (3.43). A detailed description can be found in Section 3.6.3.

4.6.4 Frequency-weighted Log-spectral Signal Distortion

The sound quality of the AFC system using the proposed AFC-CM and AFC-CE methods was evaluated through the frequency-weighted log-spectral signal distortion (SD). The SD(n) measures the spectral distance between the error signal $e(n)$ and the system input signal $u(n)$ according to (3.44). A detailed description can be found in Section 3.6.4.

4.6.5 Wideband Perceptual Evaluation of Speech Quality

Moreover, the sound quality of the AFC system using the proposed AFC-CM and AFC-CE methods was perceptually evaluated through the standardized W-PESQ algorithm. The W-PESQ quantifies the perceptible distortion in the error signal $e(n)$ due to the acoustic feedback by comparing it with the system input signal $u(n)$ according to the degradation category rating. A detailed description can be found in Section 3.6.5.

4.6.6 Signal Database

The signal database used in the simulations was formed by 10 white noise and 10 speech signals. Each noise signal was a sequence of pseudorandom values drawn from the standard normal distribution. The speech signals were the same described in Section 3.6.6. The length of the signals varied with the simulation time.

4.7 Simulation Results

This section presents and discusses the performance of the proposed AFC-CM and AFC-CE methods using the configuration of the PA system, the evaluation metrics and the signals described in Section 4.6. The configuration of the proposed methods includes the highpass filter $B(q)$, instead of the delay filter $D(q)$, and the use of a Blackman window, as discussed in Section 4.4.3.3. Although it is not necessary to use a large L_D or even the highpass filter $B(q)$ when the source signal $v(n)$ is white noise as previously discussed, even in this case $B(q)$ and Blackman window were used to prove that such a configuration of the AFC-CM and AFC-CE methods is suitable for white noise and speech signals.

The proposed AFC-CM and AFC-CE started only after 125 ms of simulation to avoid initial inaccurate estimates, $L_{fr} = 8000$, $N_{fr} = 1000$, $N_{FFT_a} = 2^{15}$ and $N_{FFT_e} = 2^{17}$. The optimization of their adaptive filter parameters λ and L_H was performed identically

Table 4.2: Summary of the results obtained by the traditional NLMS algorithm and the proposed AFC-CM and AFC-CE methods for white noise.

		$\overline{\Delta\text{MSG}}$	$\overrightarrow{\Delta\text{MSG}}$	$\overline{\text{MIS}}$	$\overrightarrow{\text{MIS}}$	$\overline{\text{SD}}$	$\overrightarrow{\text{SD}}$
NLMS	$\Delta K = 0$	7.9	9.9	-7.7	-11.1	0.4	0.2
	$\Delta K = 13$	12.1	19.0	-12.5	-20.6	0.5	0.4
	$\Delta K = 30$	18.5	33.2	-19.2	-34.6	0.7	0.5
	$\Delta K = 38$	21.4	37.0	-22.8	-40.8	0.7	0.7
AFC-CM	$\Delta K = 0$	7.4	9.4	-7.5	-10.0	0.4	0.3
	$\Delta K = 13$	8.3	10.8	-7.8	-9.7	1.4	2.2
AFC-CE	$\Delta K = 0$	7.7	9.7	-7.5	-10.3	0.4	0.3
	$\Delta K = 13$	11.9	18.5	-11.9	-19.7	0.6	0.4
	$\Delta K = 30$	16.5	29.0	-17.1	-29.0	0.8	0.8

to the state-of-art PEM-AFROW method as described in Section 3.7, resulting in (3.45) and (3.46) as well as in the asymptotic values $\overrightarrow{\text{MIS}}$, $\overrightarrow{\Delta\text{MSG}}$, $\overrightarrow{\text{SD}}$ and $\overrightarrow{\text{WPESQ}}$.

4.7.1 Performance for White Noise

In general, new adaptive filtering algorithms are evaluated using white noise as their input. First, white noise excites consistently all frequencies of the system under identification which allows the adaptive filter to estimate its complete frequency response. Second, white noise eases any performance issues that may be caused by the existence of coloring in the input signal of the adaptive filter or its correlation with any other signal. In the specific case of AFC, if the source signal $v(n)$ is white noise, the correlation between the system input signal $u(n)$ and the loudspeaker signal $x(n)$ vanishes because of the delay inserted by $D(q)$ or $B(q)$, thereby resulting in an unbiased estimate of the feedback path.

Hence, the proposed AFC-CM and AFC-CE were first evaluated using white noise as the source signal $v(n)$. The ambient noise signal $r(n) = 0$. For performance comparison, the traditional NLMS adaptive filtering algorithm was used. The parameters of the NLMS, stepsize μ and L_H, were obtained following the same procedure of the proposed AFC-CM and AFC-CE methods. Table 4.2 summarizes the results obtained by the NLMS and the proposed AFC-CM and AFC-CE methods for white noise.

In the first configuration of the forward path $G(q, n)$, the broadband gain $K(n)$ remained constant, i.e., $\Delta K = 0$. Figure 4.15 compares the results obtained by the AFC methods under evaluation for $\Delta K = 0$. It can be observed that all the AFC methods presented similar performances with a slight advantage for the NLMS. The NLMS achieved $\overrightarrow{\Delta\text{MSG}} \approx 9.9$ dB and $\overrightarrow{\text{MIS}} \approx -11.1$ dB, outscoring respectively the AFC-CM by 0.5 dB

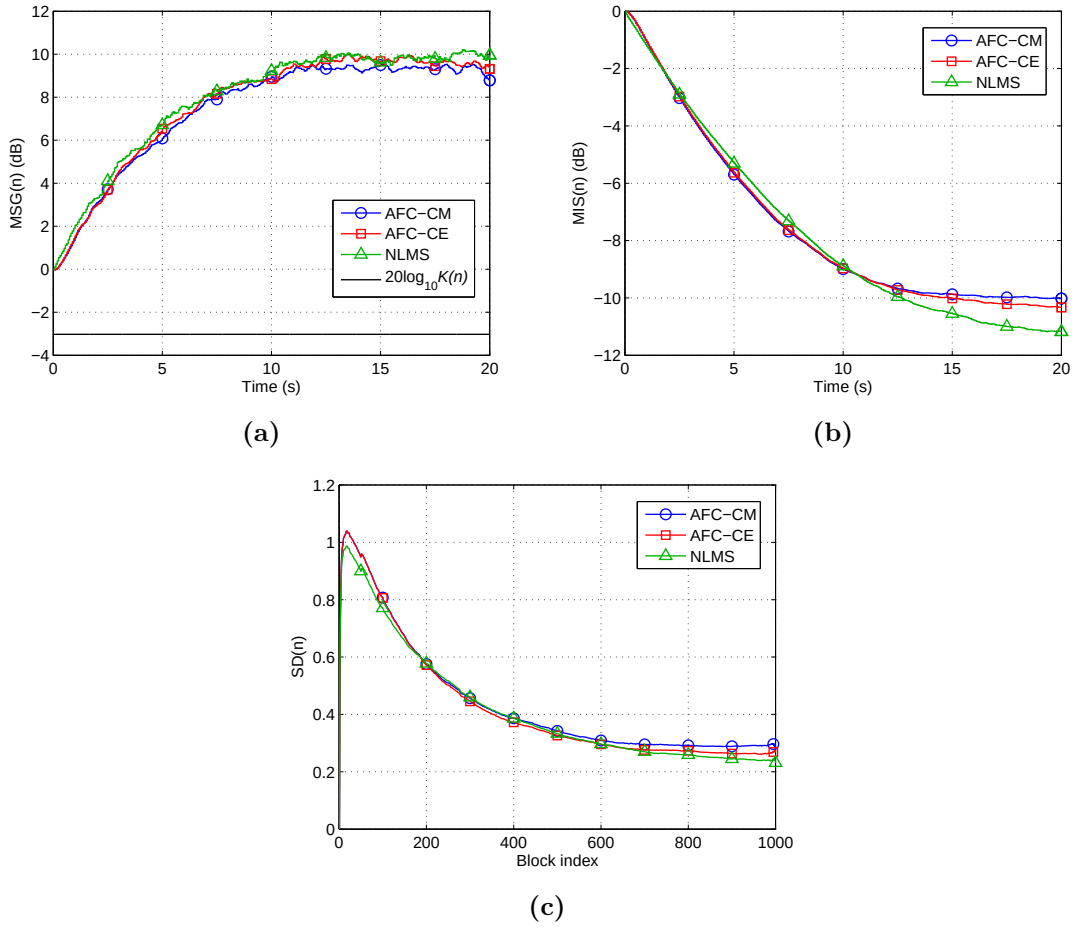


Figure 4.15: Performance comparison between the NLMS, AFC-CM and AFC-CE methods for white noise and $\Delta K = 0$: (a) MSG(n); (b) MIS(n); (c) SD(n).

and 1.1 dB and the AFC-CE by 0.2 dB and 0.8 dB. Regarding sound quality, the NLMS achieved $\overrightarrow{\text{SD}} \approx 0.2$ outscoring the AFC-CM and AFC-CE by only 0.1.

In the second configuration of the forward path $G(q, n)$, $K(n)$ was increased in order to determine the MSBG of each method, that is the maximum value of K_2 with which an AFC method achieves a MSG(n) completely stable. The first method to reach this situation was the AFC-CM method when $\Delta K = 13$ dB. Figure 4.16 compares the results obtained by the AFC methods under evaluation for $\Delta K = 13$ dB. It can be noticed that the AFC-CM performed well until 10 s of simulation. After this time, the performance of the AFC-CM method was limited by the inaccuracy of (4.16). A complete explanation about the performance of the proposed AFC-CM method will be presented in Section 4.7.2.1. The traditional NLMS and the proposed AFC-CE method presented, as the previous case, similar performances with a slight advantage for the NLMS. The NLMS achieved $\overrightarrow{\Delta\text{MSG}} \approx 19.0$ dB and $\overrightarrow{\text{MIS}} \approx -20.6$ dB, outscoring respectively the AFC-CM by 8.2 dB and 10.9 dB and the AFC-CE by 0.5 dB and 0.9 dB.

With respect to sound quality, the AFC-CM method presented the worst performance

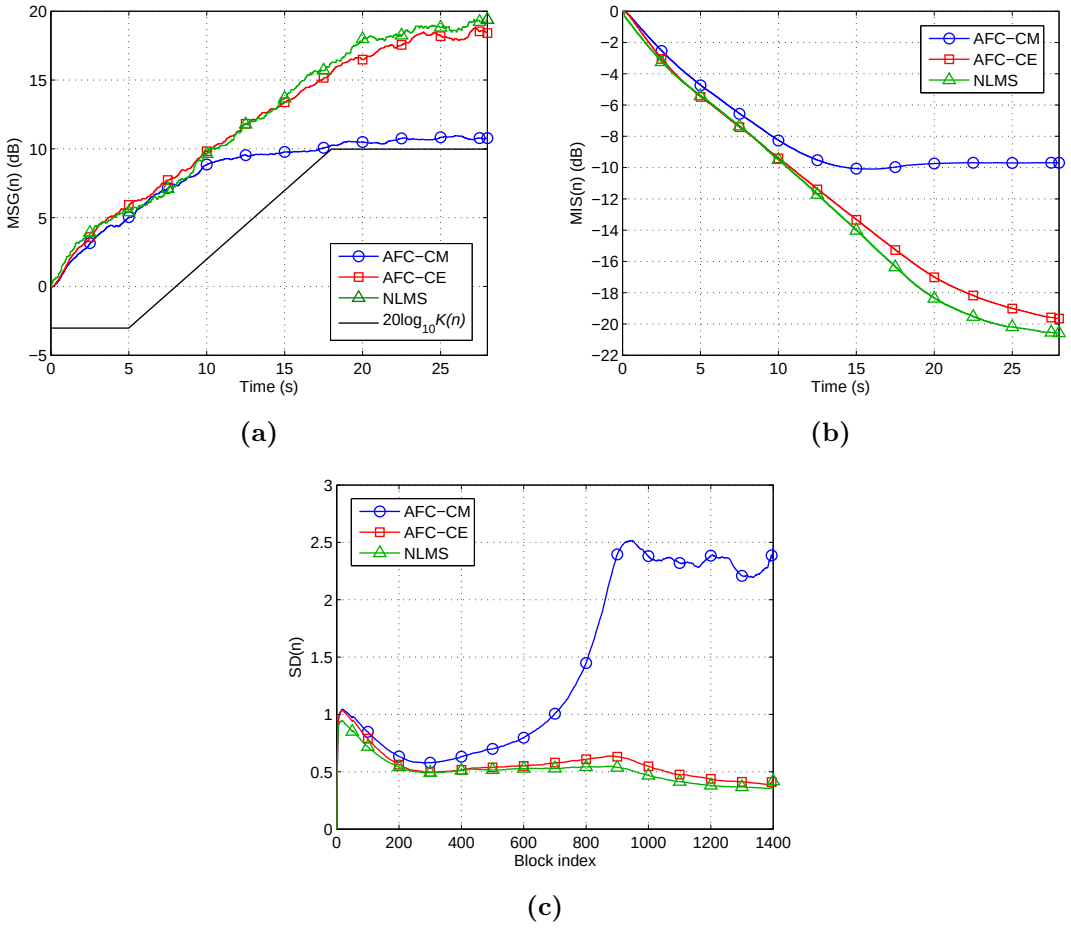


Figure 4.16: Performance comparison between the NLMS, AFC-CM and AFC-CE methods for white noise and $\Delta K = 13$ dB: (a) MSG(n); (b) MIS(n); (c) SD(n).

by obtaining $\overrightarrow{\text{SD}} = 2.2$ due to its less accurate estimate of the feedback path, as can be observed in Figure 4.16b. Hence, among all the methods, its uncancelled feedback signal $[\mathbf{f}(n) - \mathbf{h}(n)] * x(n)$ has the highest energy and, consequently, its error signal $e(n)$ has the largest distortion compared with the system input signal $u(n)$. From an MSG point of view, this can be concluded by observing in Figure 4.16a that the AFC-CM method has the lowest stability margin. Although its MSG(n) is completely stable, some instability occurred for a few signals which resulted in excessive reverberation or even in low-intensity howlings in the error signal $e(n)$. The NLMS and AFC-CE achieved $\overrightarrow{\text{SD}} = 0.4$ due to their more accurate estimates of the feedback path.

Hereupon, $K(n)$ continued to be increased to determine the MSBG of the other methods. The second method to reach this situation was the proposed AFC-CE when $\Delta K = 30$ dB. Figure 4.17 shows the results obtained by the AFC-CE and NLMS for $\Delta K = 30$ dB. It can be observed that the traditional NLMS outperformed slightly the proposed AFC-CE method. The NLMS achieved $\overrightarrow{\Delta\text{MSG}} \approx 33.2$ dB and $\overrightarrow{\text{MIS}} \approx -34.6$ dB while the AFC-CE obtained $\overrightarrow{\Delta\text{MSG}} \approx 29$ dB and $\overrightarrow{\text{MIS}} \approx -29$ dB. Regarding sound quality,

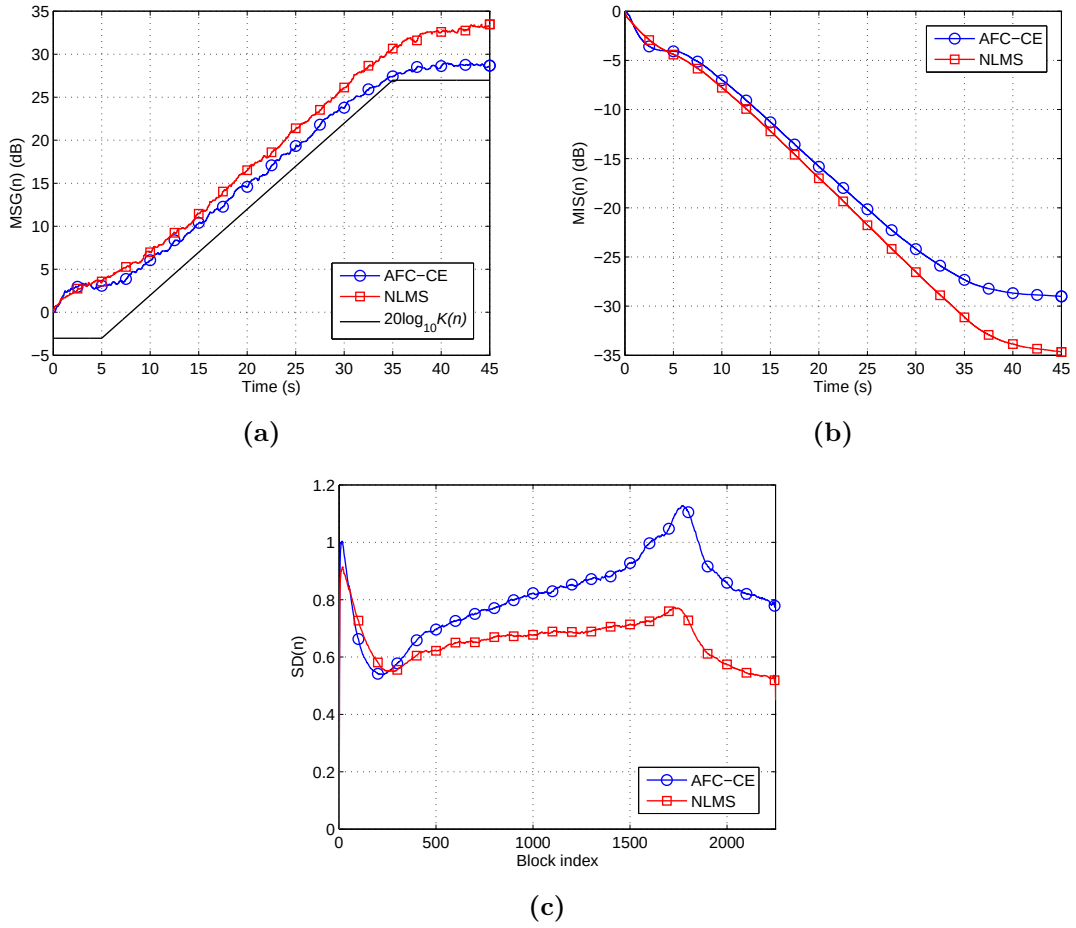


Figure 4.17: Performance comparison between the NLMS and AFC-CE methods for white noise and $\Delta K = 30$ dB: (a) MSG(n); (b) MIS(n); (c) SD(n).

the NLMS also presented the best performance by achieving $\overrightarrow{\text{SD}} = 0.5$ while the AFC-CE obtained $\overrightarrow{\text{SD}} = 0.8$.

Finally, $K(n)$ was increased further to determine the MSBG of the traditional NLMS algorithm. This situation occurred only when $\Delta K = 38$ dB. Figures 4.18a and 4.18b show the results obtained by the NLMS for $\Delta K = 38$ dB. The NLMS achieved $\overrightarrow{\Delta\text{MSG}} \approx 37.0$ dB and $\overrightarrow{\text{MIS}} \approx -40.8$ dB. With regard to sound quality, the NLMS achieved $\overrightarrow{\text{SD}} = 0.7$.

In conclusion, when the source signal $v(n)$ is white noise, the proposed AFC-CM and AFC-CE methods did not outperform the traditional NLMS algorithm. The NLMS increased by 37.0 dB the MSG of the PA system, outscoring the AFC-CM and AFC-CE by 26.2 and 8 dB, respectively. Moreover, the NLMS algorithm estimated the impulse response of the feedback path with an MIS of -33.9 dB, outscoring the AFC-CM and AFC-CE by 31.1 and 11.8 dB, respectively. And even with the same variation in the broadband gain of the forward path $G(q, n)$, ΔK, the NLMS always outperformed the other methods not only regarding MSG(n) and MIS(n) but also SD(n).

However, it is worth mentioning that, when the source signal $v(n)$ is white noise, the

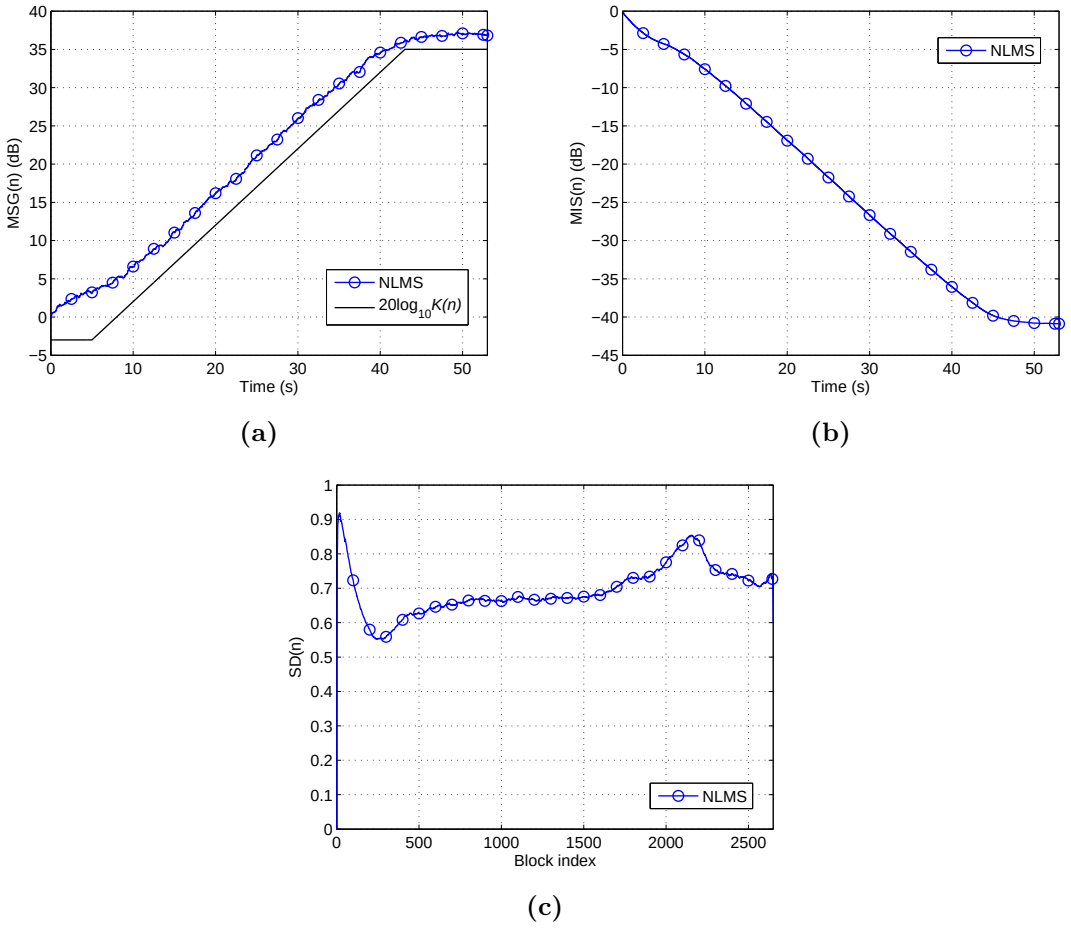


Figure 4.18: Average results of the NLMS for white noise and $\Delta K = 38$ dB: (a) MSG(n); (b) MIS(n); (c) SD(n).

system input signal $u(n)$ and the loudspeaker signal $x(n)$ are uncorrelated because of the delay applied by $G(q)D(q)$ (or $G(q,n)B(q)$). Then, the traditional gradient-based or least-squares-based adaptive filtering algorithms work properly and provide unbiased solutions. Moreover, white noise excitations guarantee the fastest convergence speed of the NLMS algorithm because the input autocorrelation matrix equals the identity matrix [72, 74]. This causes the NLMS to be equivalent to the LMS-Newton algorithm, which has a performance similar to the recursive least-squares (RLS) algorithm [72]. And, even in this situation so advantageous to the traditional NLMS algorithm, the proposed AFC-CE method performed well.

4.7.2 Performance for Speech Signals

For speech as source signal $v(n)$, the evaluation of the proposed AFC-CM and AFC-CE methods was done in two ambient noise conditions. The first was an ideal condition where the ambient noise signal $r(n) = 0$ and thus the source-signal-to-noise ratio was SNR $= \infty$. The second was closer to real-world conditions where $r(n) \neq 0$ such that SNR = 30 dB. The ambient white noise $r(n)$ contributes to approach the cepstrum $\mathbf{c_u}(n)$ of the system input signal to an impulse-like waveform, which may improve the estimate of the acoustic feedback path provided by the methods. Table 4.3 summarizes the results obtained by the AFC-CM and AFC-CE methods for speech signals.

4.7.2.1 AFC-CM Method

The performance of the AFC-CM method is shown in Figures 4.19 and 4.20. Figure 4.19 shows the results obtained for $\Delta K = 0$. In order to illustrate the bias problem in AFC, the results obtained by the NLMS adaptive filtering algorithm when SNR = 30 dB are also considered. The AFC-CM method achieved $\overrightarrow{\Delta\text{MSG}} \approx 9.6$ dB and $\overrightarrow{\text{MIS}} \approx -10.2$ dB when SNR $= \infty$, and $\overrightarrow{\Delta\text{MSG}} \approx 9.8$ dB and $\overrightarrow{\text{MIS}} \approx -10.2$ dB when SNR = 30 dB. The relative efficiency of the AFC-CM is clear when comparing its results with those of the NLMS. With respect to sound quality, the AFC-CM achieved $\overrightarrow{\text{SD}} \approx 1.7$ and $\overrightarrow{\text{WPESQ}} \approx 2.74$ when SNR $= \infty$, and $\overrightarrow{\text{SD}} \approx 1.4$ and $\overrightarrow{\text{WPESQ}} \approx 2.53$ when SNR = 30 dB.

Hereupon, $K(n)$ was increased in order to determine the MSBG achievable by the AFC-CM method. This situation occurred with $\Delta K = 14$ dB for both ambient noise conditions. Figure 4.20 shows the results obtained by the AFC-CM method for $\Delta K = 14$ dB. The AFC-CM method achieved $\overrightarrow{\Delta\text{MSG}} \approx 12.0$ dB and $\overrightarrow{\text{MIS}} \approx -9.8$ dB when SNR $= \infty$, and $\overrightarrow{\Delta\text{MSG}} \approx 12.0$ dB and $\overrightarrow{\text{MIS}} \approx -9.8$ dB when SNR = 30 dB. With respect to sound quality, the AFC-CM achieved $\overrightarrow{\text{SD}} \approx 9.0$ and $\overrightarrow{\text{WPESQ}} \approx 1.21$ when SNR $= \infty$, and $\overrightarrow{\text{SD}} \approx 8.1$ and $\overrightarrow{\text{WPESQ}} \approx 1.23$ when SNR = 30 dB.

It can be observed that the results of MSG(n) and MIS(n) improve as ΔK increases. The same occurred with the PEM-AFROW method as shown in Section 3.7. In the case of the AFC-CM method, as explained in Section 4.4.3.1, the improvement in MSG and MIS is due to the fact that, when the broadband gain $K(n)$ of the forward path increases, the absolute values of the system open-loop impulse response $\mathbf{g}(n) * \mathbf{f}(n)$ increase while the cepstrum $\mathbf{c_u}(n)$ of the system input signal is not affected. Then, the estimation of $\mathbf{g}(n) * \mathbf{f}(n)$ from the cepstrum $\mathbf{c_y}(n)$ of the microphone signal is improved which, consequently, improves the estimate of the acoustic feedback path provided by the AFC-CM method.

On the other hand, the results of SD(n) and WPESQ(n) worsen as ΔK increases. This is because, despite the improvement in the estimates of the feedback path provided by the adaptive filters, the increase in the gain of $G(q, n)$ ultimately results in an increase in the energy of the uncancelled feedback signal $[\mathbf{f}(n) - \mathbf{h}(n)] * x(n)$. From an MSG point of view, this can be concluded by observing that the stability margins of the systems

Table 4.3: Summary of the results obtained by the proposed AFC-CM and AFC-CE methods for speech signals.

			$\overline{\Delta\text{MSG}}$	$\overrightarrow{\Delta\text{MSG}}$	$\overline{\text{MIS}}$	$\overrightarrow{\text{MIS}}$	$\overline{\text{SD}}$	$\overrightarrow{\text{SD}}$	$\overline{\text{WPESQ}}$	$\overrightarrow{\text{WPESQ}}$
AFC-CM	$\Delta K = 0$	SNR = 30	7.8	9.8	-7.8	-10.2	1.8	1.4	2.58	2.74
		SNR = ∞	7.8	9.6	-7.8	-10.2	2.1	1.7	2.40	2.53
	$\Delta K = 14$	SNR = 30	9.2	12.0	-8.1	-9.8	5.0	8.1	1.69	1.23
		SNR = ∞	8.9	12.0	-7.7	-9.8	5.7	9.0	1.59	1.21
AFC-CE	$\Delta K = 0$	SNR = 30	8.3	10.7	-8.0	-11.0	1.7	1.2	2.64	2.90
		SNR = ∞	8.4	11.0	-8.1	-11.3	2.0	1.4	2.48	2.73
	$\Delta K = 14$	SNR = 30	13.3	20.0	-13.5	-20.9	2.2	2.0	2.34	2.32
		SNR = ∞	13.6	20.4	-13.7	-21	2.6	2.6	2.18	2.22
	$\Delta K = 16$	SNR = 30	13.8	21.0	-14.1	-22.4	2.2	2.1	2.32	2.29
		SNR = ∞	14.2	21.2	-14.5	-22.3	2.7	2.5	2.13	2.10
	$\Delta K = 30$	SNR = 30	17.1	30.0	-16.4	-25.0	3.3	4.0	1.85	1.54
		SNR = ∞	17.0	29.6	-15.1	-22	4.0	4.6	1.67	1.46

decreased. For $\Delta K = 14$ dB, the stability margin became very low, mainly after $t = 17$ s as can be observed in Figure 4.20a, and some instability occurred for a few signals, which resulted in excessive reverberation or even in some howlings in the error signal $e(n)$.

It is noteworthy that the values of SD(n) obtained when the source signal $v(n)$ is speech are higher than those obtained when $v(n)$ is white noise. As explained in Section 3.6.4, the SD(n) is a ratio between the short-term power spectral densities $S_e(e^{j\omega}, n)$ and $S_u(e^{j\omega}, n)$, which are computed using frames with duration of 20 ms of the system input signal $u(n)$

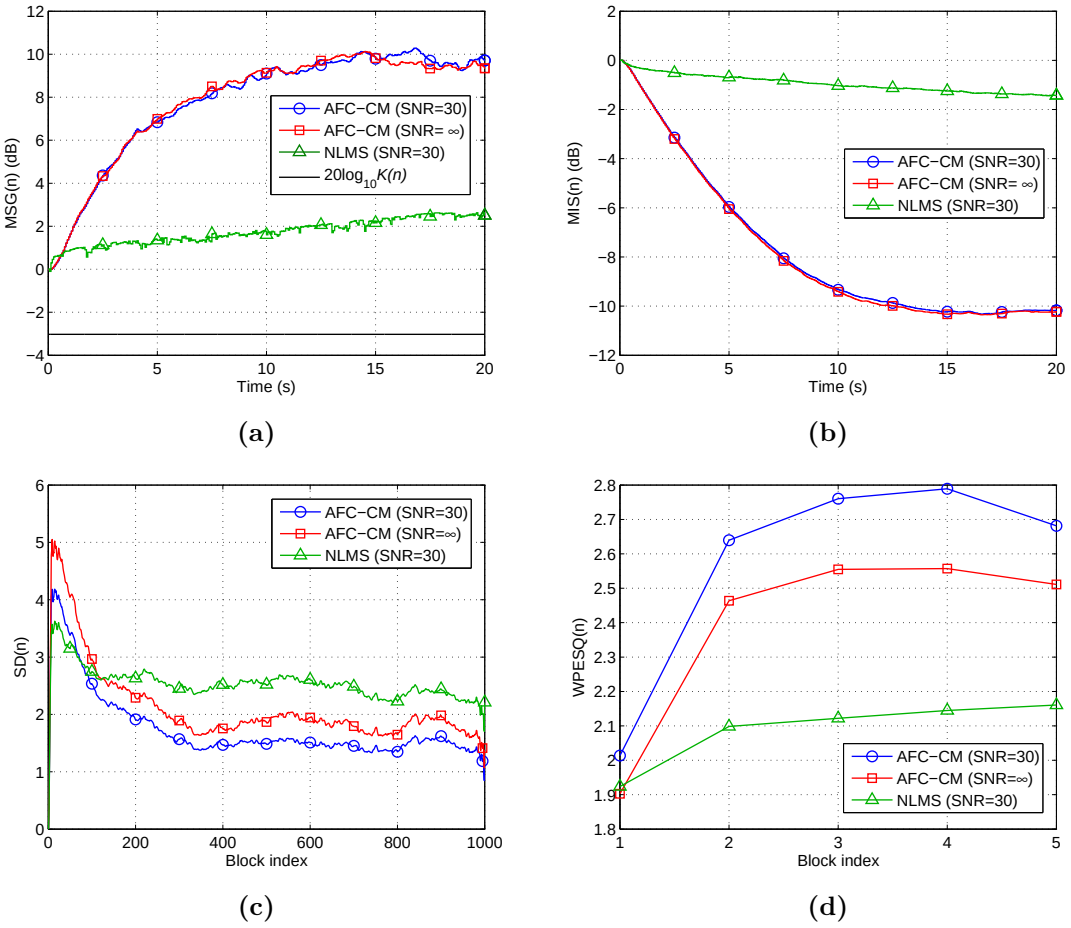


Figure 4.19: Average results of the AFC-CM method for speech signals and $\Delta K = 0$: (a) MSG(n); (b) MIS(n); (c) SD(n); (d) WPESQ(n).

and the error signal $e(n)$, respectively. When the source signal $v(n)$ is speech, there are always short-time segments of very low energy (almost silence) between words or phonemes. Then, when SNR $= \infty$, the frames of $u(n)$ may contain only these very low-intensity segments of $v(n)$, leading to $S_u(e^{j\omega}, n)$ with very low values. However, because of the uncancelled feedback signal $x(n) * [\mathbf{f}(n) - \mathbf{h}(n)]$, the corresponsing segments in $e(n)$ always contain a significant energy which results in an $S_e(e^{j\omega}, n)$ with considerable values. Consequently, for these signal segments, the value of the ratio in SD(n) may be very high and increases $\overline{\text{SD}}$. On the other hand, the decrease in SNR (increase in the level of $r(n)$) causes the energy of the corresponding segments in the system input signal $u(n)$ to increase as well as the values of their $S_u(e^{j\omega}, n)$. As a result, for these segments, the value of the ratio in SD(n) is now not so high and then has a lower influence on $\overline{\text{SD}}$. When $u(n)$ is essentially white noise, these short-time segments of very low energy no longer exist.

Furthermore, as also occurred with the PEM-AFROW, the results obtained with SNR $= 30$ dB are slightly better than those obtained with SNR $= \infty$. The ambient noise $r(n)$, being white noise, contributes to approach the cepstrum $\mathbf{c_u}(n)$ of the system

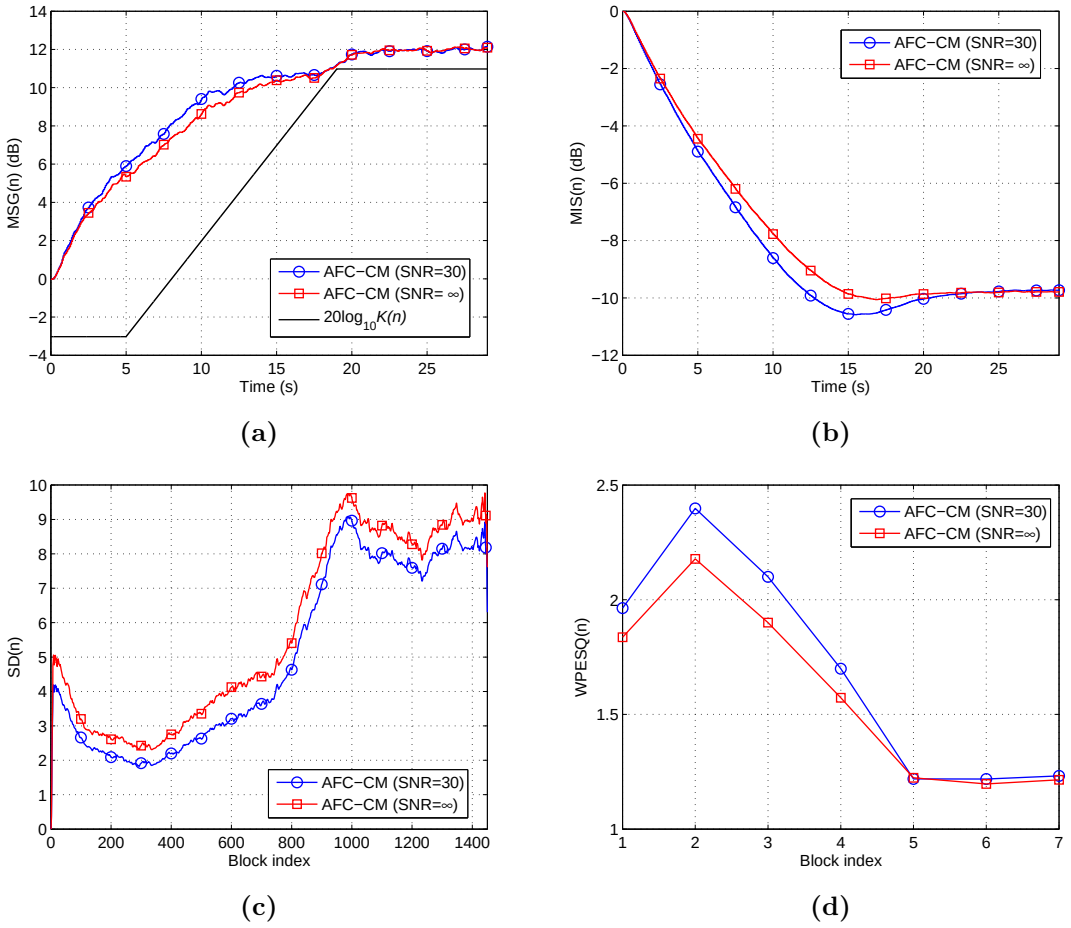


Figure 4.20: Average results of the AFC-CM method for speech signals and $\Delta K = 14$ dB: (a) MSG(n); (b) MIS(n); (c) SD(n); (d) WPESQ(n).

input signal to an impulse-like waveform, which may improve the estimation of $\mathbf{g}(n) * \mathbf{f}(n)$ from $\mathbf{c_y}(n)$ provided by the AFC-CM method.

In Section 4.3.1, a detailed explanation was given on how the performance of the AFC-CM method is theoretically limited by the need to fulfill the condition $\left|G(e^{j\omega}, n)B(e^{j\omega})H(e^{j\omega}, n)\right| < 1$, which ultimately becomes the NGC of the PA system. The results presented in this section demonstrated it in practice. In the first configuration of the forward path $G(q, n)$, where $\Delta K = 0$, the condition $\left|G(e^{j\omega}, n)B(e^{j\omega})H(e^{j\omega}, n)\right| < 1$ was always fulfilled. Then, $\mathbf{c_y}(n)$ was accurately defined by (4.16) and the AFC-CM worked optimally throughout the simulation time. In this case, the performance of the AFC-CM method was limited by the cepstrum $\mathbf{c_u}(n)$ of the system input signal that acts as noise in the estimation of $\mathbf{g}(n) * \mathbf{f}(n)$ from $\mathbf{c_y}(n)$, as explained in Section 4.4.3.1.

In the second configuration of $G(q, n)$, where $K(n)$ increases over time, the AFC-CM performed well until $t = 12$ s as can be observed in Figures 4.20a and 4.20b. In this time interval, the method worked properly because the condition $\left|G(e^{j\omega}, n)B(e^{j\omega})H(e^{j\omega}, n)\right| < 1$ was fulfilled at all frequency components and then (4.16) was accurately defined or, at least,

it was partially fulfilled such that the inaccuracy of (4.16) was small. But after this time interval, (4.16) becomes inaccurate to the point of disrupting the estimate of the feedback path provided by the AFC-CM method and thereby limits its performance. This behavior is easily noticed in the MIS(n) presented in Figure 4.20b. The need to fulfill the condition $\left|G(e^{j\omega},n)B(e^{j\omega})H(e^{j\omega},n)\right| < 1$ limited the increase in the broadband gain, ΔK, in 14 dB and, consequently, the performance of the AFC-CM method.

4.7.2.2 AFC-CE Method

Similarly, the performance of the AFC-CE method is shown in Figures 4.21 and 4.22. Figure 4.21 shows the results obtained for $\Delta K = 0$. Once again, the results obtained by the NLMS adaptive filtering algorithm when SNR = 30 dB are also included. The AFC-CE method achieved $\overrightarrow{\Delta\text{MSG}} \approx 11.0$ dB and $\overrightarrow{\text{MIS}} \approx -11.3$ dB when SNR $= \infty$, and $\overrightarrow{\Delta\text{MSG}} \approx 10.7$ dB and $\overrightarrow{\text{MIS}} \approx -11.0$ dB when SNR = 30 dB. The relative efficiency of the AFC-CE method is also evident when comparing its results with those of the NLMS. Regarding sound quality, the AFC-CE achieved $\overrightarrow{\text{SD}} \approx 1.4$ and $\overrightarrow{\text{WPESQ}} \approx 2.73$ when SNR $= \infty$, and $\overrightarrow{\text{SD}} \approx 1.2$ and $\overrightarrow{\text{WPESQ}} \approx 2.90$ when SNR = 30 dB.

Hereupon, $K(n)$ was increased in order to determine the MSBG achievable by the AFC-CE method. This situation occurred with an impressive $\Delta K = 30$ dB for both ambient noise conditions. Figures 4.22a and 4.22b shows the results obtained by the AFC-CE method for $\Delta K = 30$ dB. The AFC-CE method achieved $\overrightarrow{\Delta\text{MSG}} \approx 29.6$ dB and $\overrightarrow{\text{MIS}} \approx -22$ dB when SNR $= \infty$, and $\overrightarrow{\Delta\text{MSG}} \approx 30.0$ dB and $\overrightarrow{\text{MIS}} \approx -25.0$ dB when SNR = 30 dB. With respect to sound quality, the AFC-CE achieved $\overrightarrow{\text{SD}} \approx 4.6$ and $\overrightarrow{\text{WPESQ}} \approx 1.46$ when SNR $= \infty$, and $\overrightarrow{\text{SD}} \approx 4.0$ and $\overrightarrow{\text{WPESQ}} \approx 1.54$ when SNR = 30 dB.

It can be observed that, as occurred with the PEM-AFROW and AFC-CM, the results of MSG(n) and MIS(n) improve as ΔK increases. As explained in Section 4.4.3.1, when $\Delta K = 0$, the magnitude of the impulse response $\mathbf{g}(n) * [\mathbf{f}(n) - \mathbf{h}(n)]$ decreases as $H(q,n)$ approaches $F(q,n)$ while the cepstrum $\mathbf{c_u}(n)$ of the system input signal is not affected. But, when the broadband gain $K(n)$ of the forward path increases, this magnitude decrease that would be caused by $\mathbf{h}(n)$ is compensated. Then, the estimation of $\mathbf{g}(n) * [\mathbf{f}(n) - \mathbf{h}(n)]$ from the cepstrum $\mathbf{c_e}(n)$ of the error signal becomes more accurate which, consequently, improves the performance of the AFC-CE method.

On the other hand, as also occurred with the PEM-AFROW and AFC-CM, the results of SD(n) and WPESQ(n) worsen as ΔK increases. This is because, despite the improvement in the estimates of the feedback path provided by the adaptive filters, the increase in the gain of $G(q,n)$ ultimately results in an increase in the energy of the uncancelled feedback signal $[\mathbf{f}(n) - \mathbf{h}(n)] * x(n)$. From an MSG point of view, this can be concluded by observing that the stability margins of the systems decreased. When $\Delta K = 0$, the stability margin was always higher than 3 dB and reached 14 dB. But, when $\Delta K = 30$ dB, the stability margin never exceeded 6 dB, was less then 3 dB for approximately 40% of

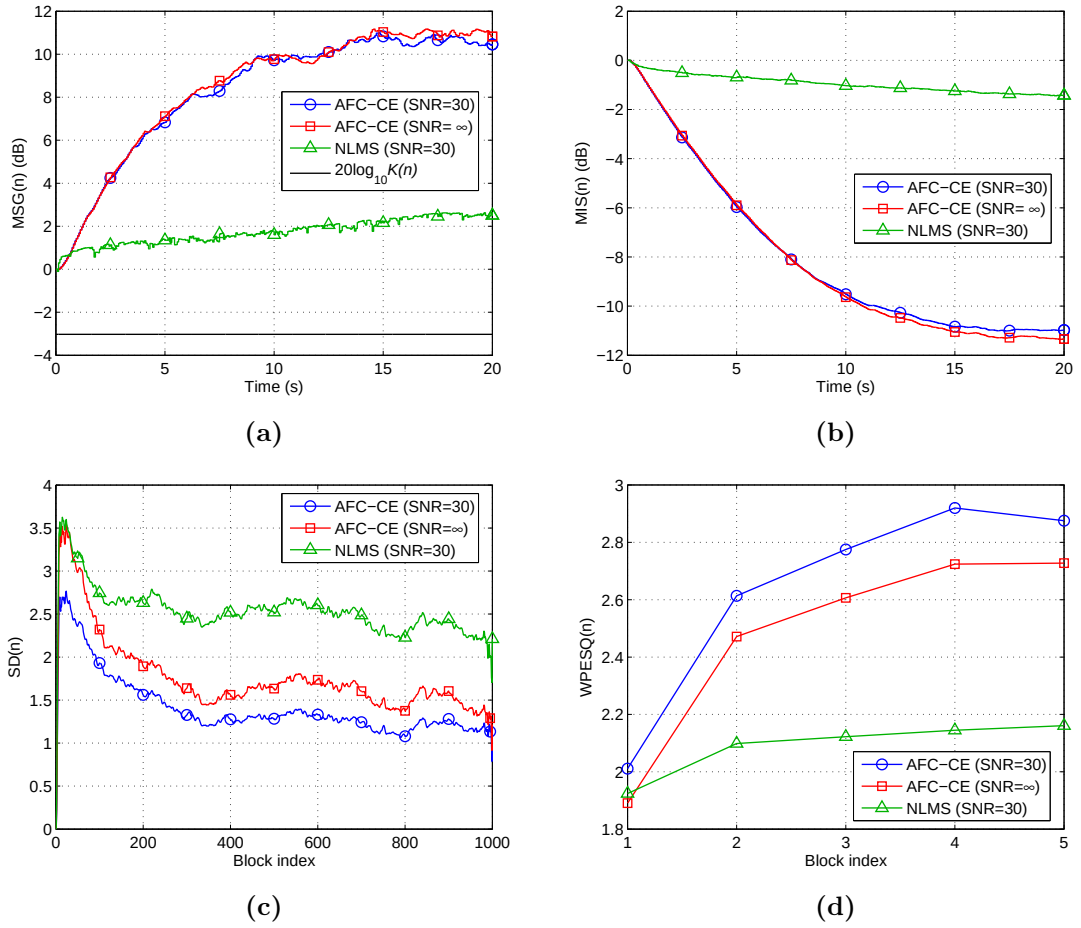


Figure 4.21: Average results of the AFC-CE method for speech signals and $\Delta K = 0$: (a) MSG(n); (b) MIS(n); (c) SD(n); (d) WPESQ(n).

the simulation time and, mainly, was very low for $30 \leq t \leq 40$ s. Although the MSG(n) is completely stable, some instability occurred for a few signals but no howling was audible.

With respect to the level of the ambient noise $r(n)$, the results showed that the performance of the AFC-CE in terms of MSG and MIS does not have a well-defined behavior. For $\Delta K = 0, 14$ and 16 dB, the method performed better with SNR $= \infty$. For $\Delta K = 30$ dB, the method performed better when SNR $= 30$ dB. But, with the exception of the MIS when $\Delta K =$ SNR $= 30$ dB, the difference in performance was very small as can be noticed from Table 4.3. This indicates that the AFC-CE method achieves similar performances for low-intensity noise environments when the source signal $v(n)$ is speech.

Finally, it can be concluded that the AFC-CE method outperforms the AFC-CM. This was expected because, as previously discussed in Sections 4.3.1 and 4.3.2, the only requirement in order for $\mathbf{c_e}(n)$ to be defined by (4.23) is the fulfillment of the NGC of the AFC system whereas the condition $\left|G(e^{j\omega}, n)B(e^{j\omega})H(e^{j\omega}, n)\right| < 1$ must also be fulfilled in order for $\mathbf{c_y}(n)$ to be defined by (4.16). Then, when this additional condition is fulfilled as occurred with $\Delta K = 0$, both methods present similar performances as can

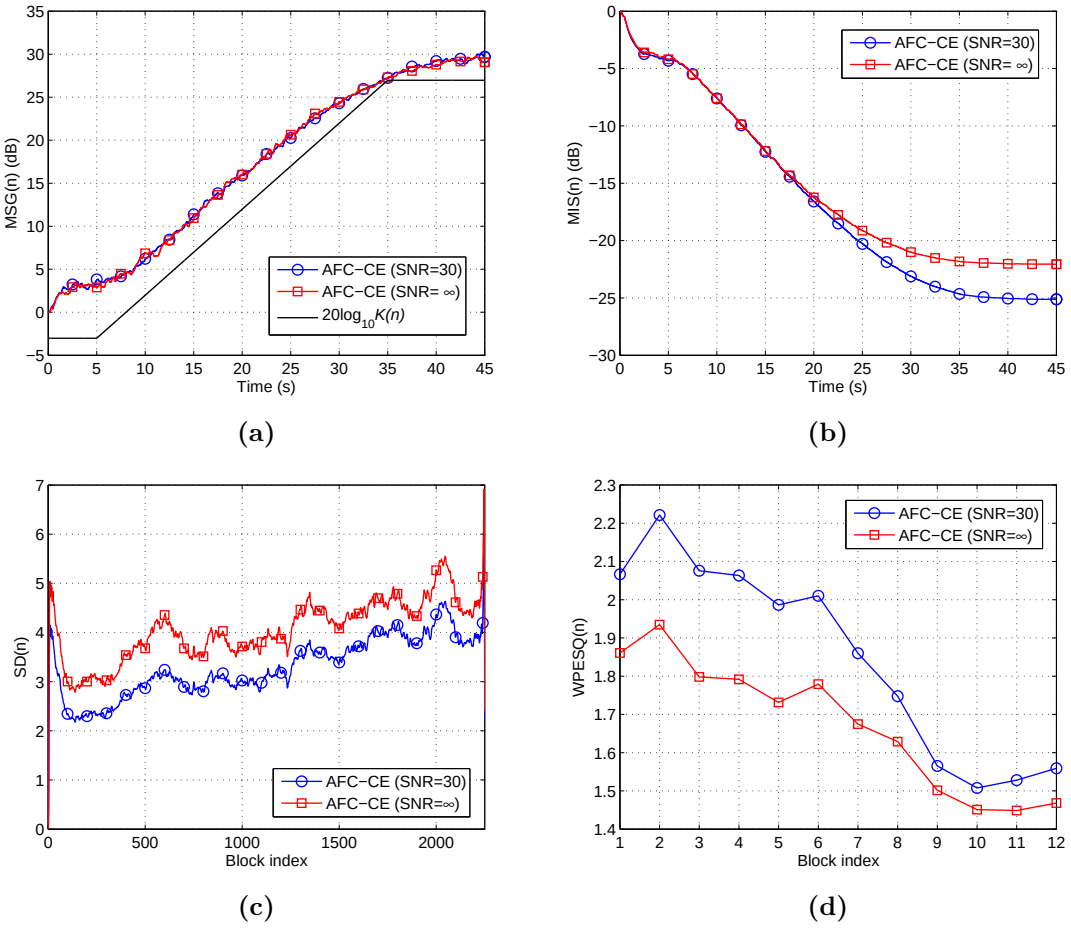


Figure 4.22: Average results of the AFC-CE method for speech signals and $\Delta K = 30$ dB: (a) MSG(n); (b) MIS(n); (c) SD(n); (d) WPESQ(n).

be observed from Table 4.3 and Figures 4.23a and 4.23b. But when the broadband gain $K(n)$ of the forward path increases, $\Delta K > 0$, as $H(q,n)$ converges to $F(q,n)$, the condition $\left|G(e^{j\omega},n)B(e^{j\omega})H(e^{j\omega},n)\right| < 1$ is no longer satisfied after a certain time and thereby limits the performance of the AFC-CM method. Meanwhile, the AFC-CE method works properly because the NGC of the AFC system is still fulfilled.

In fact, the performance of the AFC-CE method was only limited by the influence of the cepstrum $\mathbf{c_u}(n)$ of the system input signal that acts as noise in the estimation of $\mathbf{g}(n) * [\mathbf{f}(n) - \mathbf{h}(n)]$ from $\mathbf{c_e}(n)$. And, as explained in Section 4.4.3.1, the influence of $\mathbf{c_u}(n)$ on the performance of the AFC-CE method has a lower bound that is obtained with $\left|G(e^{j\omega},n)\right| = \left[\max_\omega \left|B(e^{j\omega})\left[F(e^{j\omega},n) - H(e^{j\omega},n)\right]\right|\right]^{-1}$. For $\Delta K = 0, 14$ and 16 dB, this lower bound was not reached. But, in general, the influence of $\mathbf{c_u}(n)$ proved to be, in practice, quite small which allows the proposed AFC-CE method to increase the MSG of the PA system by 30 dB. Furthermore, the performance of the AFC-CE could be even better if the growth rate of the broadband gain $K(n)$ of the forward path were smaller.

4.7.2.3 Comparison with PEM-AFROW

After the evaluation and discussion of their individual performances, the proposed AFC-CM and AFC-CE methods will be now compared with the state-of-art PEM-AFROW method. The performance of the PEM-AFROW method was presented and discussed in Section 3.7. The comparison will focus on the results obtained with SNR = 30 dB because this ambient noise condition is closer to real-world conditions.

Figure 4.23 compares the results obtained by the AFC methods under evaluation for $\Delta K = 0$. It can be observed that the AFC-CM and AFC-CE methods presented similar performances, with a slight advantage for the AFC-CE, and both methods outperformed the PEM-AFROW. The proposed AFC-CE method achieved $\overrightarrow{\Delta\text{MSG}} \approx 10.7$ dB and $\overrightarrow{\text{MIS}} \approx -11$ dB, outscoring respectively the AFC-CM by 0.7 dB and 0.8 dB and the PEM-AFROW by 2.7 dB and 1.7 dB.

With respect to sound quality, the AFC-CE achieved $\overrightarrow{\text{SD}} \approx 1.2$ and $\overrightarrow{\text{WPESQ}} \approx 2.90$, outscoring respectively the AFC-CM by 0.2 and 0.16 and the PEM-AFROW by 0.3 and 0.26. These differences are almost imperceptible because, with such constant value of $K(n)$ and the increase in MSG provided by all the AFC methods, the systems were too far from instability as can be observed in Figure 4.23a.

Consider now the second configuration of the broadband gain $K(n)$ of the forward path where it was linearly (in dB scale) increased over time, as explained in Section 3.6.1, in order to determine the MSBG of each method. The AFC-CE method achieved an MSBG of the forward path $G(q, n)$ equal to 27 dB, outperforming the AFC-CM and the state-of-art PEM-AFROW by impressive 16 dB and 14 dB, respectively. This would be enough to conclude that the proposed AFC-CE method has the best performance. However, aiming to enrich the discussion, the performance of the AFC methods under evaluation will be compared considering the results obtained with all the values of ΔK used in this work.

Figure 4.24 compares the results obtained by the AFC methods under evaluation for $\Delta K = 14$ dB. It can be noticed that the AFC-CM performed well, even better than the PEM-AFROW, until 10 s of simulation. After this time, as previously explained in detail, the performance of the AFC-CM method was limited by the inaccuracy of (4.16). This behavior is easily observed in MIS(n) shown in Figure 4.24b. However, it is evident that the AFC-CE stood out from both methods by achieving $\overrightarrow{\Delta\text{MSG}} \approx 20$ dB and $\overrightarrow{\text{MIS}} \approx -20.9$ dB, outscoring respectively the AFC-CM by 8 dB and 11.1 dB and the PEM-AFROW by 6.5 dB and 5.6 dB. Moreover, it should be noted that the AFC-CM method outperformed the PEM-AFROW by 0.5 dB with regard to $\overline{\Delta\text{MSG}}$, which was the cost function in the optimization of the adaptive filter parameters for all methods.

Regarding sound quality, the AFC-CM method presented the worst performance by obtaining $\overrightarrow{\text{SD}} \approx 8.1$ and $\overrightarrow{\text{WPESQ}} \approx 1.23$ because its very low stability margin after $t = 17$ s, as can be observed in Figure 4.24a. Although its MSG(n) is completely stable, some instability occurred for a few signals which resulted in excessive reverberation or

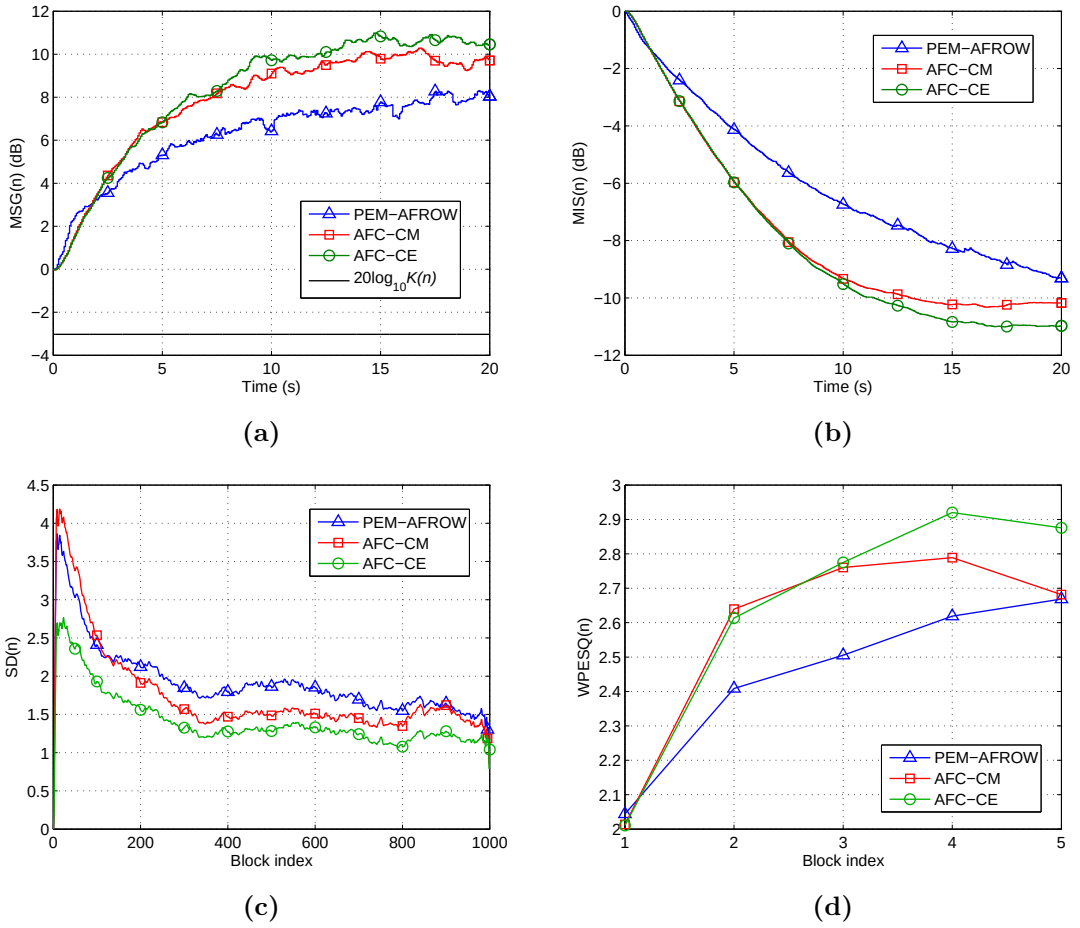


Figure 4.23: Performance comparison between the PEM-AFROW, AFC-CM and AFC-CE methods for speech signals and $\Delta K = 0$: (a) MSG(n); (b) MIS(n); (c) SD(n); (d) WPESQ(n).

even in some howlings in the error signal $e(n)$. On the other hand, the AFC-CE method presented the best sound quality by achieving $\overrightarrow{\text{SD}} \approx 2.0$ and $\overrightarrow{\text{WPESQ}} \approx 2.32$ because its largest stability margin and outscored the PEM-AFROW by, respectively, 1.9 and 0.69.

Figure 4.25 compares the results obtained by the PEM-AFROW and AFC-CE methods for $\Delta K = 16$ dB. Once again, it can be observed that the AFC-CE method outperformed the PEM-AFROW. The PEM-AFROW obtained $\overrightarrow{\Delta\text{MSG}} \approx 15$ dB and $\overrightarrow{\text{MIS}} \approx -16.2$ dB while the AFC-CE method achieved $\overrightarrow{\Delta\text{MSG}} \approx 21$ dB and $\overrightarrow{\text{MIS}} \approx -22.4$ dB. Regarding sound quality, the AFC-CE method achieved $\overrightarrow{\text{SD}} \approx 2.1$ and $\overrightarrow{\text{WPESQ}} \approx 2.29$ while the PEM-AFROW obtained $\overrightarrow{\text{SD}} \approx 3.9$ and $\overrightarrow{\text{WPESQ}} \approx 1.58$.

In conclusion, the proposed AFC-CE method increased by 30 dB the MSG of the PA system, outperforming the AFC-CM and PEM-AFROW by, respectively, 18 and 15 dB. Moreover, the AFC-CE method estimated the impulse response of the feedback path with an MIS of -25 dB, outperforming the AFC-CM and PEM-AFROW by, respectively, 15.2 and 8.8 dB. And even with the same variation in the broadband gain $K(n)$ of the forward

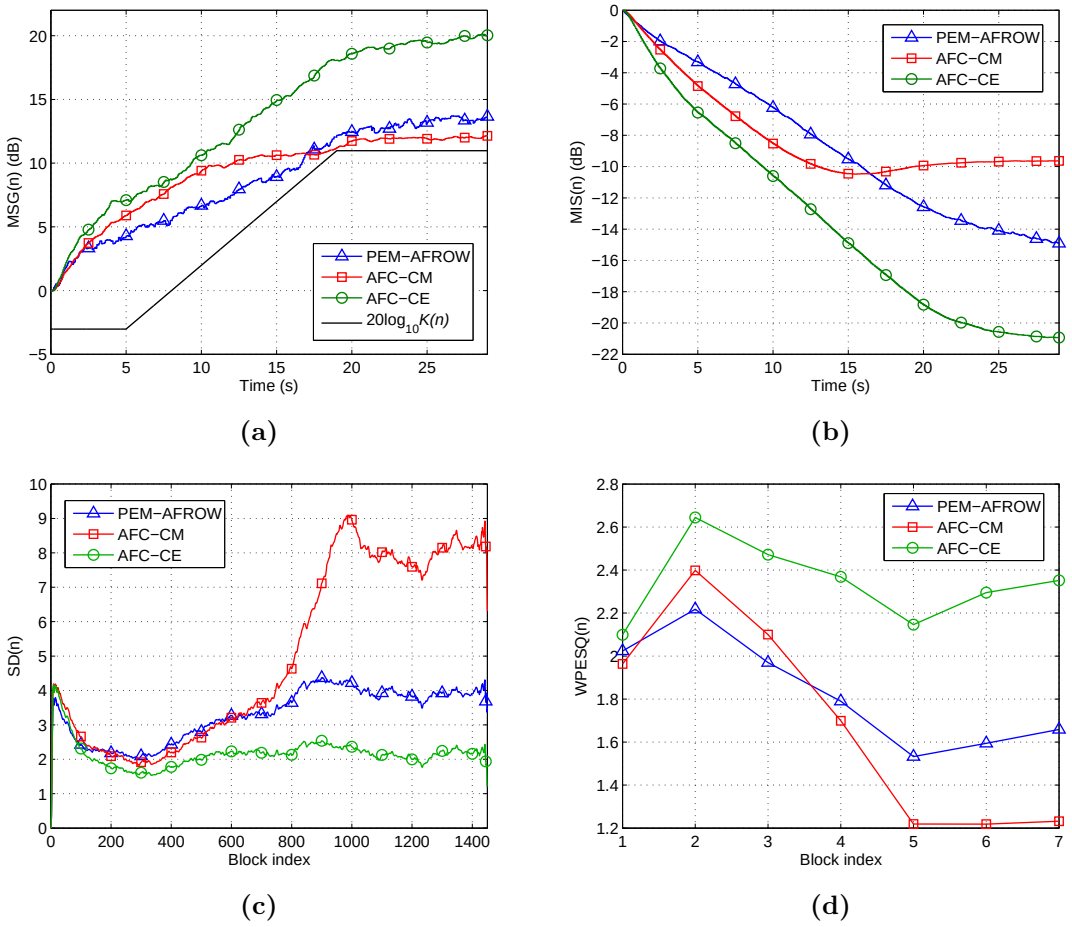


Figure 4.24: Performance comparison between the PEM-AFROW, AFC-CM and AFC-CE methods for speech signals and $\Delta K = 14$ dB: (a) MSG(n); (b) MIS(n); (c) SD(n); (d) WPESQ(n).

path, ΔK, the AFC-CE always outperformed the other methods not only in MSG(n) and MIS(n) but also in SD(n) and WPESQ(n).

Moreover, the structure of the PEM-AFROW method uses a source model that generally works well only for a restricted group of signals such as speech. If the nature of the source signal $v(n)$ changes over time, the PEM-AFROW method may not work properly unless its source model is modified appropriately to the nature of the new source signal. On the other hand, the definitions of the cepstra $\mathbf{c_y}(n)$ and $\mathbf{c_e}(n)$ of the microphone and error signals according to (4.16) and (4.23), respectively, as well as the basic equations of the proposed AFC-CM and AFC-CE methods, described respectively in Sections 4.4.1 and 4.4.2, are valid independently of the source signal $v(n)$.

In fact, the source signal $v(n)$ (through the system input signal $u(n) = v(n) + r(n)$) can interfere in the methods because the cepstrum $\mathbf{c_u}(n)$ acts as noise in the estimation of the 1-fold impulse responses from $\mathbf{c_y}(n)$ and $\mathbf{c_e}(n)$. When $v(n)$ is white noise or speech, it was proved that $\mathbf{c_u}(n)$ has, on average, a fast decay over sample and consequently has

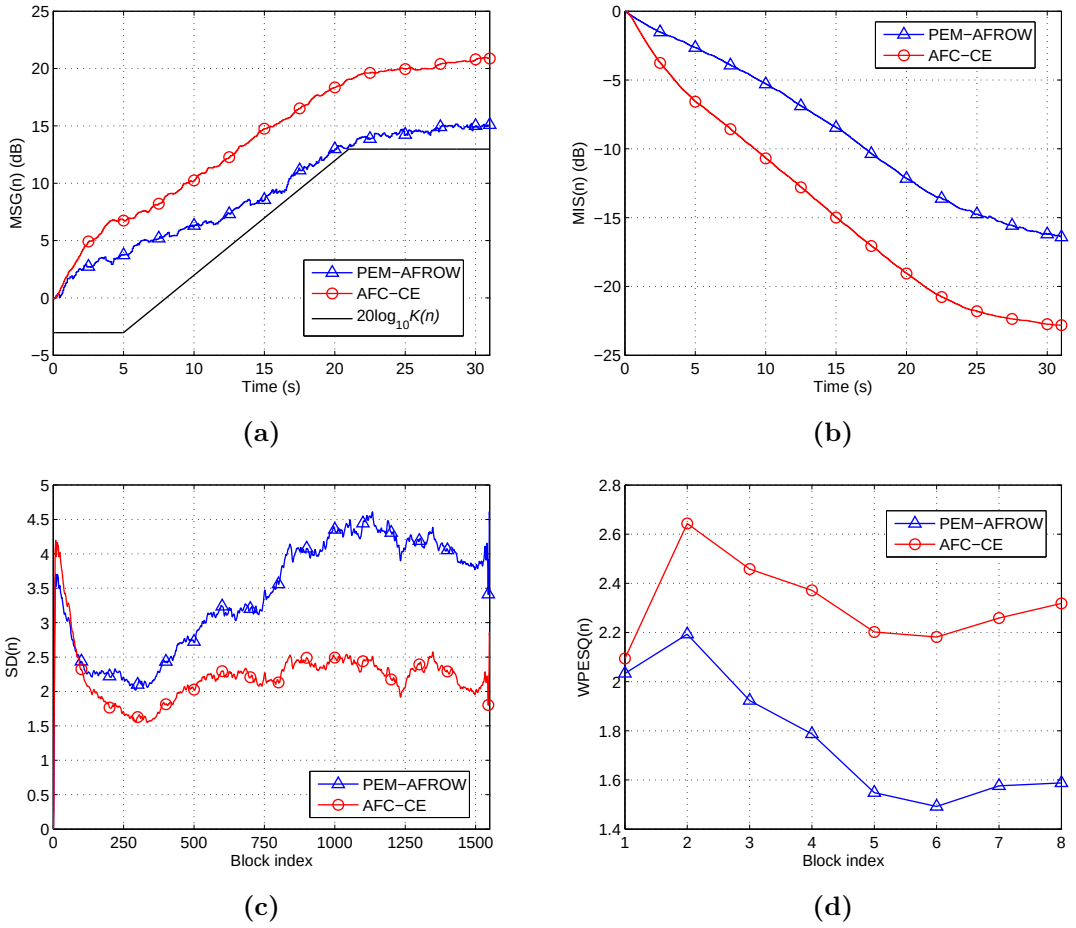


Figure 4.25: Performance comparison between the PEM-AFROW and AFC-CE methods for speech signals and $\Delta K = 16$ dB: (a) MSG(n); (b) MIS(n); (c) SD(n); (d) WPESQ(n).

low absolute values in the region where the 1-fold impulse responses are located in $\mathbf{c_y}(n)$ and $\mathbf{c_e}(n)$, which enables the methods to work properly. However, as a cepstrum, $\mathbf{c_u}(n)$ will always have a decay at least as fast as $1/m$, where m is the sample index, regardless of the signal nature. In the worst case, a higher $\frac{L_B-1}{2}$ (time delay caused by $B(q)$) will be required to accurately estimate the 1-fold impulse responses from $\mathbf{c_y}(n)$ and $\mathbf{c_e}(n)$.

Therefore, as with the PEM-AFROW, the nature of the source signal may affect the AFC-CM and AFC-CE methods. But, certainly, it is much easier to adapt the time delay caused by the cascade $G(q,n)B(q)$ through L_B in the proposed methods than to adapt the source model in the PEM-AFROW in order to suit the nature of the source signal $v(n)$ over time. Furthermore, a sufficient large value of L_B in the proposed AFC-CM and AFC-CE methods will probably suit the great majority of the signals.

4.8 Conclusion

This chapter detailed a cepstral analysis of a PA system. It was proved that the cepstrum of the microphone signal contains time domain information about the system, including its open-loop impulse response, if the NGC of the PA system is fulfilled. In addition, it was demonstrated that it is possible to remove the acoustic feedback by removing all the system information from the cepstrum of the microphone signal. Moreover, this work aimed to use this information to update an adaptive filter in a typical AFC system.

To this purpose, a cepstral analysis of an AFC system, where an error signal is generated from the microphone signal, was also detailed. It was proved that, in an AFC system, the cepstrum of the microphone signal may also contain time domain information about the system, including the open-loop impulse response of the PA system. But for this, the NGC of the AFC system and a gain condition as a function of the frequency responses of the forward path and adaptive filter must be fulfilled. A new AFC method based on the cepstral analysis of the microphone signal, called as AFC-CM, was proposed. The AFC-CM method estimates the feedback path impulse response from the cepstrum of the microphone signal to update the adaptive filter. A theoretical discussion on why the second aforementioned condition limits the use of the cepstrum of the microphone signal in an AFC system was presented and it was also demonstrated in practice by the proposed AFC-CM method.

Furthermore, in an AFC system, it was also proved that the cepstrum of the error signal may contain time domain information about the system, including the open-loop impulse response of the AFC system. But for this, as an advantage over the microphone signal, only the NGC of the AFC system must be fulfilled. Finally, a new AFC method based on the cepstral analysis of the error signal, called as AFC-CE, was proposed. The AFC-CE method estimates the feedback path impulse response from the cepstrum of the error signal to update the adaptive filter.

Simulation results demonstrated that, when the source signal is speech, the proposed AFC-CE method can estimate the feedback path impulse response with a MIS of -25 dB, outperforming the PEM-AFROW and the proposed AFC-CM by respectively 8.8 and 15.2 dB. Moreover, the AFC-CE method can increase by 30 dB the MSG of the PA system, outperforming the PEM-AFROW and AFC-CM by respectively 15 and 18 dB. It may be concluded that the proposed AFC-CE method achieves a less biased estimate of the acoustic feedback path and further increases the MSG of the PA system in comparison with the proposed AFC-CM method and state-of-art PEM-AFROW method.

Chapter 5

Acoustic Feedback Cancellation with Multiple Feedback Paths

5.1 Introduction

Chapters 3 and 4 addressed the AFC problem considering PA systems with only one microphone and one loudspeaker. In fact, this configuration is nearly the only one found in the literature and represents several practical applications of PA systems as, for instance, in hearing aids. However, this configuration may not precisely represent the use of PA systems in other practical applications as, for instance, in large environments.

This chapter deals with the AFC problem considering PA systems with one microphone and four loudspeakers. The acoustic coupling between the loudspeakers and the microphone result in four feedback paths. It is proved that, in this configuration of the PA system, the feedback signal is completely removed from the microphone signal if the adaptive filter impulse response is equal to the sum of the impulse responses of the single feedback paths. Moreover, the impulse response resulting from the sum of the impulse responses of the single feedback paths generally has a large number of prominent peaks and lower sparseness. It also has, in general, frequency components with higher energy.

The influence of a room impulse response with lower sparseness and higher energy in its frequency components on the performance of the PEM-AFROW, AFC-CM and AFC-CE methods is discussed. Finally, an evaluation of the AFC methods is carried out in a simulated environment. It is demonstrated that, for the same value of the increase in the broadband gain of the forward path, the AFC methods usually perform worse with multiple feedback paths as regards misalignment but the system sound quality is improved.

5.2 AFC with Multiple Feedback Paths

Typically, aiming to be heard by a large audience in the same acoustic environment, a speaker uses a PA system with one microphone, responsible for picking up his/her own voice, one amplification system, responsible for amplifying the voice signal, and several loudspeakers placed in different positions, responsible for playback and distributing the voice signal in the acoustic environment so that everyone in the audience can hear it.

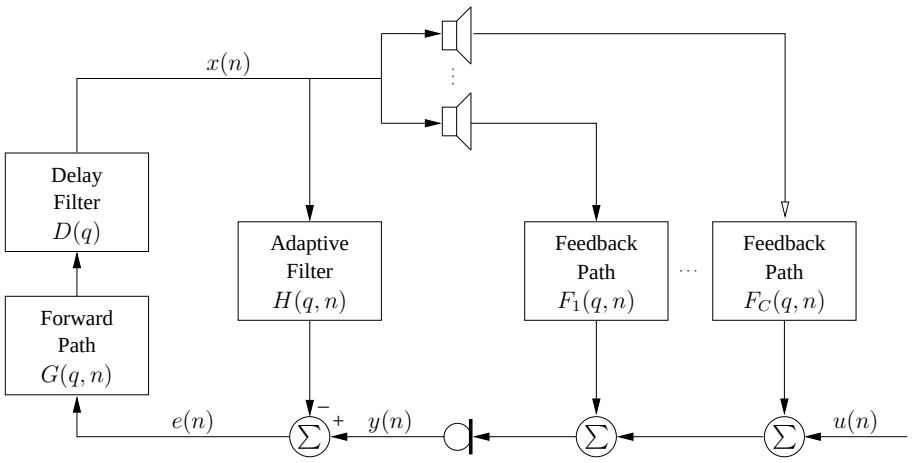


Figure 5.1: Typical AFC system with multiple feedback paths.

A typical PA system with 1 microphone and C loudspeakers is depicted in Figure 5.1. The loudspeaker signal $x(n)$, after played back by the kth-loudspeaker, may be fed back into the microphone through the feedback path $F_k(q,n)$. The C acoustic feedback signals $\mathbf{f}_k(n) * x(n)$ are added to the system input signal $u(n)$, generating the microphone signal

$$y(n) = u(n) + \sum_{k=1}^{C} \mathbf{f}_k(n) * x(n). \tag{5.1}$$

Then, an estimate of the overall feedback signal is calculated as $\mathbf{h}(n) * x(n)$ and subtracted from the microphone signal $y(n)$, generating the error signal

$$\begin{aligned} e(n) &= u(n) + \sum_{k=1}^{C} \mathbf{f}_k(n) * x(n) - \mathbf{h}(n) * x(n) \\ &= u(n) + \left[\sum_{k=1}^{C} \mathbf{f}_k(n) - \mathbf{h}(n) \right] * x(n), \end{aligned} \tag{5.2}$$

which is effectively the signal to be fed to the forward path $G(q,n)$. The error signal $e(n)$ will contain no acoustic feedback as desired if

$$H(q,n) = \sum_{k=1}^{C} F_k(q,n). \tag{5.3}$$

In this scenario with multiple feedback paths, the adaptive filter has optimum solution equal to the sum of the single acoustic feedback paths. Indeed, the AFC system with multiple feedback paths in Figure 5.1 can be simplified to the AFC system with single feedback path in Figure 3.1 by considering $F(q,n)$ as the overall acoustic feedback path such that

$$F(q,n) = \sum_{k=1}^{C} F_k(q,n). \tag{5.4}$$

However, in this case, the impulse response $\mathbf{f}(n)$ generally has a larger number of prominent peaks and, consequently, lower sparseness as will be demonstrated in Section 5.3.1.1. An impulse response is sparse if a small percentage of its coefficients have a significant magnitude while the rest are small or zero [75]. Another definition follows: an impulse response is sparse if a large fraction of its energy is concentrated in a small fraction of its coefficients. In general, a room impulse response is sparse because its magnitude typically decays exponentially over time. And the sparseness measure of a room impulse response is inversely proportional to its reverberation time (decay speed).

The traditional adaptive filtering algorithms, as the NLMS, have slow convergence when identifying sparse impulse responses [75, 76]. This fact has led to the development of several adaptive algorithms for the identification of sparse impulse responses as, for example, in [76, 77, 78, 79, 80, 81, 82, 83]. These new adaptive algorithms improve the performance of the traditional algorithms by changing their update equation so that the sparseness of the impulse response under identification is taken into account.

Therefore, the importance of evaluating AFC methods considering multiple feedback paths is twofold. First, it corresponds to a more realistic configuration of a typical PA system. Second, the resulting feedback path has lower sparseness which may affect the performance of the traditional adaptive algorithms and, thus, of the PEM-AFROW method. And, as it will de demonstrated, the decrease in sparseness may also affect the performance of the proposed AFC-CM and AFC-CE methods.

5.3 Simulation Configurations

With the aim to assess the performance of the proposed AFC-CM and AFC-CE methods in a PA system with multiple feedback paths, an experiment was carried out in a simulated environment to measure their ability to estimate the feedback path impulse response and increase the MSG of a PA system. The resulting distortion in the error signal $e(n)$ was also measured. To this purpose, the following configuration was used.

5.3.1 Simulated Environment

5.3.1.1 Feedback Path

The impulse responses $\mathbf{f}_k(n)$ of the acoustic feedback paths were 4 measured room impulse response of the same room available in [60], where each one was measured with the sound emitter placed in a different position, and thus $\mathbf{f}_k(n) = \mathbf{f}_k$. The impulse responses were downsampled to $f_s = 16$ kHz and then truncated to length $L_F = 4000$ samples, and are illustrated in Figure 5.2.

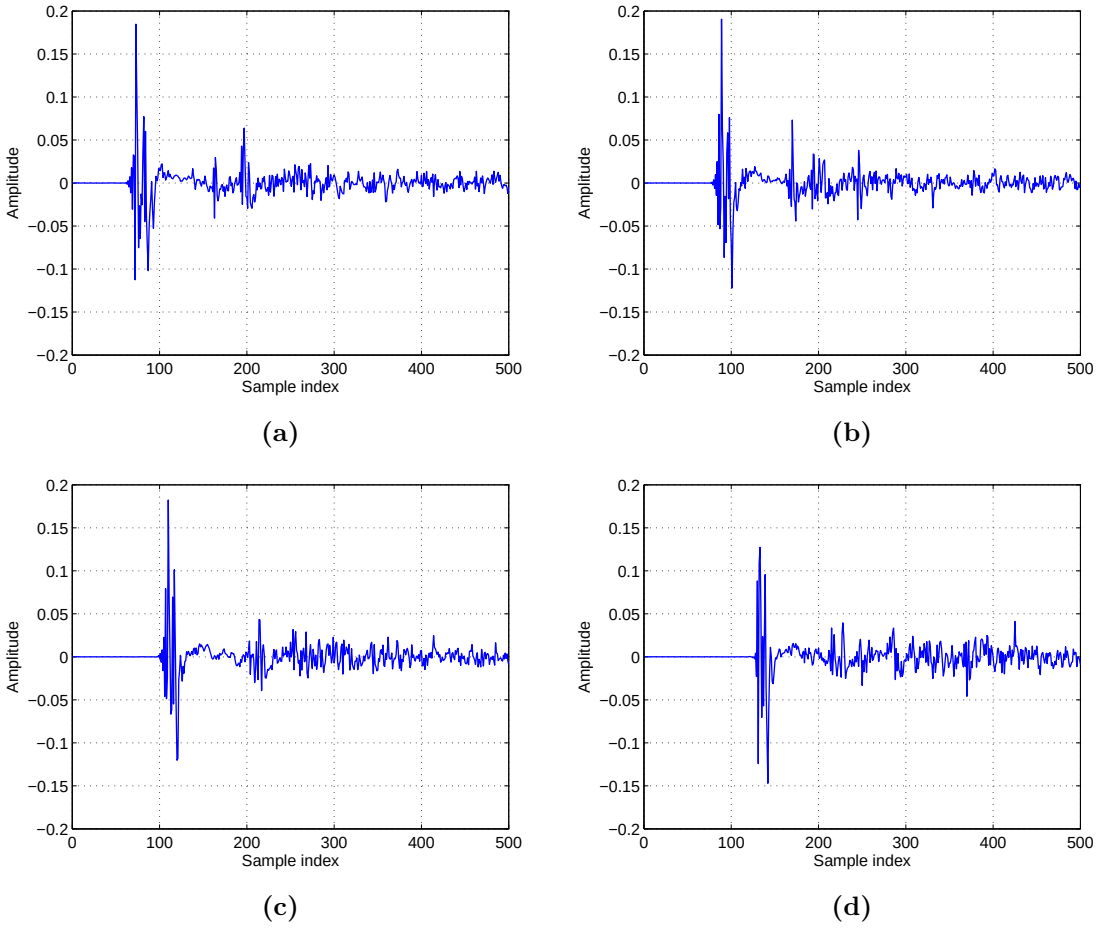
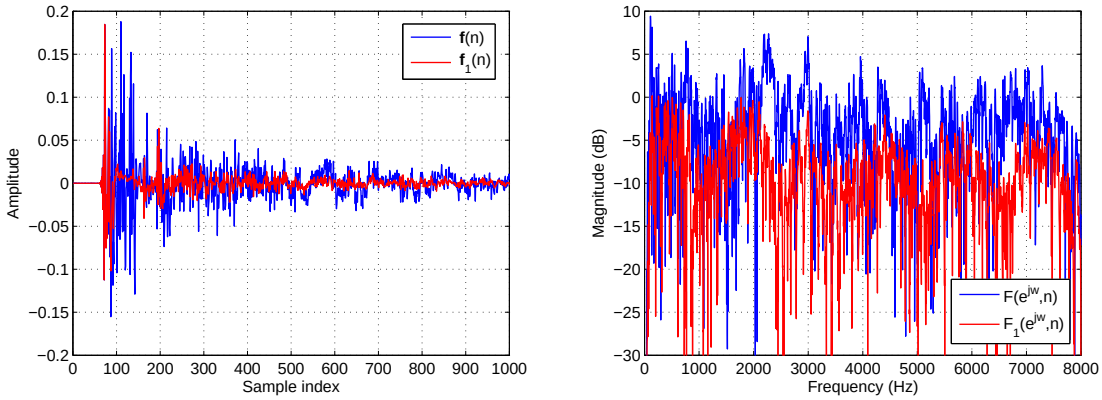


Figure 5.2: Impulse responses of the acoustic feedback paths (zoom in the first 500 samples): (a) $\mathbf{f}_1(n)$; (b) $\mathbf{f}_2(n)$; (c) $\mathbf{f}_3(n)$; (d) $\mathbf{f}_4(n)$.

Figure 5.3 compares the single feedback path $F_1(q,n)$, which was used in Chapters 3 and 4, and the overall feedback path $F(q,n)$. It can be observed from Figure 5.3a that, compared with the impulse response of $F_1(q,n)$, the impulse response of $F(q,n)$ has coefficients with absolute values generally higher but the highest absolute value is almost the same. This indicates a reduction in sparseness.



(a) **(b)**

Figure 5.3: Comparison between single $F_1(q,n)$ and multiple $F(q,n)$ acoustic feedback paths: (a) impulse response; (b) frequency response.

The sparseness of an impulse response $\mathbf{f}(n)$ can be quantified by [76]

$$\xi(n) = \frac{L_F}{L_F - \sqrt{L_F}} \left[1 - \frac{\|\mathbf{f}(n)\|_1}{\sqrt{L_F}\|\mathbf{f}(n)\|_2} \right], \tag{5.5}$$

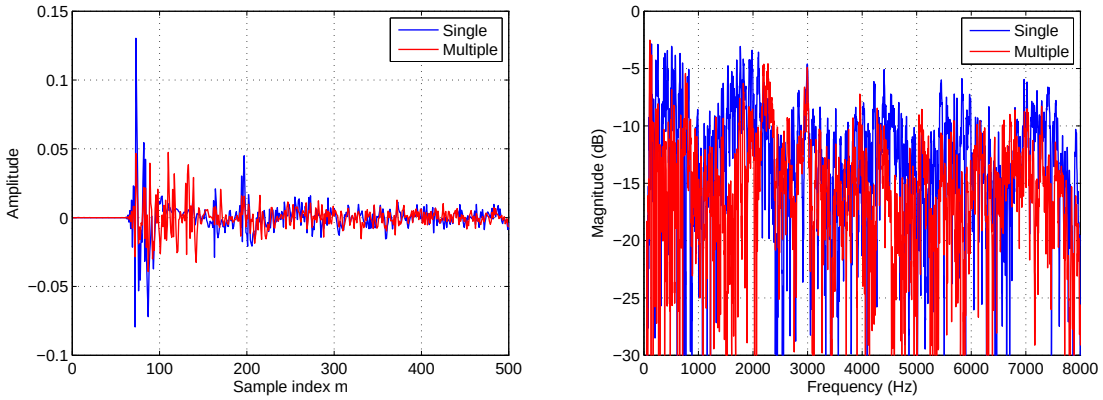
where $\|\cdot\|_1$ and $\|\cdot\|_2$ denote the l_1 and l_2-norm, respectively. According to (5.5), $\mathbf{f}_1(n)$ has $\xi = 0.75$ and $\mathbf{f}(n)$ has $\xi = 0.67$. It can be concluded that, when the system has multiple feedback paths, the sparseness of the impulse response of the overall feedback path decreases, in this case by 11%, which may affect the performance of adaptive filtering algorithms [75, 76].

Moreover, it can be observed from Figure 5.3 that $F(q,n)$ has higher energy than $F_1(q,n)$. In fact, $\mathbf{f}(n)$ has an energy 6.13 dB higher than $\mathbf{f}_1(n)$. This will influence the performance of the proposed AFC-CM and AFC-CE methods, as will be shown and discussed in Section 5.4.

5.3.1.2 Forward Path

As in Chapters 3 and 4, the forward path $G(q,n)$ was simply defined as an unit delay and a gain according to (3.42). The two configurations of the broadband gain $K(n)$ of the forward path, explained in detail in Section 3.6.1, were applied. For the PEM-AFROW method, as explained in Section 3.6.1, $G(q,n)$ was followed by the delay filter $D(q)$ with $L_D = 401$. For the proposed AFC-CM and AFC-CE methods, as explained in Section 4.4.3.3, $G(q,n)$ was followed by the highpass filter $B(q)$ with $L_B = 801$. Note that the highpass filter $B(q)$ and delay filter $D(q)$ generate the same time delay.

With multiple feedback paths, the initial broadband gain $K(0)$ is lower due to the increase in magnitude of the frequency response $F(e^{j\omega},n)$ of the feedback path, which can be observed in Figure 5.3b. With $F_1(q,n)$, the MSG of the PA system is around 0 dB and thus $20\log_{10} K(0) \approx -3$ dB. With $F(q,n)$, the MSG of the PA system is around -9 dB



(a) **(b)**

Figure 5.4: Comparison between open-loop responses with single and multiple acoustic feedback paths: (a) impulse response; (b) frequency response.

and thus $20\log_{10} K(0) \approx -12$ dB. Therefore, for the same ΔK, the broadband gain $K(n)$ of the forward path is 9 dB lower when the system has multiple feedback paths.

Although $\mathbf{f}(n)$ has higher absolute values than $\mathbf{f}_1(n)$, the values of $\mathbf{g}(n)*\mathbf{f}(n)$ and $\mathbf{g}(n)*\mathbf{f}_1(n)$ depend on the value of $K(n)$. Figure 5.4 shows $G(q,0)F(q,0)$ and $G(q,0)F_1(q,0)$. It can be observed that, due to the lower value of $K(0)$, the highest absolute values of $\mathbf{g}(0)*\mathbf{f}(0)$ are smaller than those of $\mathbf{g}(0)*\mathbf{f}_1(0)$. In fact, for the same value of ΔK, the proportion between the values of $\mathbf{g}(n)*\mathbf{f}(n)$ and $\mathbf{g}_1(n)*\mathbf{f}_1(n)$ is the same shown in Figure 5.4a and, therefore, the highest absolute values of $\mathbf{g}(n)*\mathbf{f}(n)$ are smaller than those of $\mathbf{g}(n)*\mathbf{f}_1(n)$. Hence, for the same value of ΔK, this may make it more difficult to estimate the highest absolute values of $\mathbf{g}(n)*\mathbf{f}(n)$ from $\mathbf{c_y}(n)$ or $\mathbf{c_e}(n)$. Since these values are the ones that contribute most to the feedback problem, this fact may impair the performance of the proposed AFC-CM and AFC-CE methods.

5.3.2 Maximum Stable Gain

The main goal of any AFC method is to increase the MSG of the PA system that has an upper limit due the acoustic feedback. Therefore, the MSG is the most important metric in evaluating AFC methods.

For an AFC system that uses the PEM-AFROW, AFC-CM or AFC-CE methods, as discussed in 3.6.2 and 4.6.2, the MSG of the AFC system and the increase in MSG achieved by the AFC methods, ΔMSG, were measured according to (3.6) and (3.8), respectively.

The frequency responses in (3.6) and (3.8) were computed using an N_{FFT_e}-point FFT with $N_{FFT_e} = 2^{17}$. The sets of critical frequencies $P(n)$ and $P_H(n)$ were obtained by searching, in the corresponding unwrapped phase, each crossing by integer multiples of 2π. A detailed explanation can be found in Section 3.6.2.

5.3.3 Misalignment

In addition to the MSG, the performance of the AFC methods were also evaluated through the normalized misalignment (MIS) metric. The MIS(n) measures the mismatch between the adaptive filter and the feedback path according to (3.43). A detailed description can be found in Section 3.6.3.

5.3.4 Frequency-weighted Log-spectral Signal Distortion

The sound quality of the AFC systems was evaluated through the frequency-weighted log-spectral signal distortion (SD). The SD(n) measures the spectral distance (in dB) between the error signal $e(n)$ and the system input signal $u(n)$ according to (3.44). A detailed description can be found in Section 3.6.4.

5.3.5 Wideband Perceptual Evaluation of Speech Quality

Moreover, the sound quality of the AFC systems was perceptually evaluated through the standardized W-PESQ algorithm. The W-PESQ quantifies the perceptible distortion in the error signal $e(n)$ due to the acoustic feedback by comparing it with the system input signal $u(n)$ according to the degradation category rating. A detailed description can be found in Section 3.6.5.

5.3.6 Signal Database

The signal database was formed by the same 10 speech signals used in Chapters 3 and 4. A detailed description can be found in Section 3.6.6.

5.4 Simulation Results

This section presents and discusses the performance of the AFC-CM and AFC-CE methods proposed in Chapter 4 using the configuration of the PA system, the evaluation metrics and the signals described in Section 5.3. The state-of-art PEM-AFROW method, presented in Chapter 3, was also evaluated and used for performance comparison.

As in Chapters 3 and 4, the evaluation of the AFC methods was done in two ambient noise conditions. The first was an ideal condition where the ambient noise signal $r(n) = 0$ and thus the source-signal-to-noise ratio SNR $= \infty$. The second was close to real-world conditions where $r(n) \neq 0$ such that SNR $= 30$ dB. Table 5.1 summarizes the results obtained by the AFC methods for speech signals.

Table 5.1: Summary of the results obtained by the PEM-AFROW, AFC-CM and AFC-CE methods for speech signals.

			$\overline{\Delta\text{MSG}}$	$\overrightarrow{\Delta\text{MSG}}$	$\overline{\text{MIS}}$	$\overrightarrow{\text{MIS}}$	$\overline{\text{SD}}$	$\overrightarrow{\text{SD}}$	$\overline{\text{WPESQ}}$	$\overrightarrow{\text{WPESQ}}$
NLMS	$\Delta K = 0$	SNR $= 30$	2.6	3.4	-0.5	-0.9	2.2	2.0	2.22	2.30
		SNR $= \infty$	2.6	3.4	-0.5	-0.8	2.6	2.4	2.08	2.14
PEM-AFROW	$\Delta K = 0$	SNR $= 30$	5.7	8.0	-4.3	-6.6	1.5	1.4	2.66	2.80
		SNR $= \infty$	5.5	7.7	-4.2	-6.3	2.0	1.7	2.46	2.58
	$\Delta K = 13$	SNR $= 30$	8.4	13.2	-7.5	-13.1	2.6	3.1	1.96	1.83
		SNR $= \infty$	7.8	12.6	-6.2	-12.1	3.3	4.0	1.79	1.69
	$\Delta K = 16$	SNR $= 30$	9.2	14.7	-7.8	-14.5	2.8	3.0	1.88	1.67
		SNR $= \infty$	8.8	14.4	-7.1	-13.4	3.6	3.8	1.73	1.55
AFC-CM	$\Delta K = 0$	SNR $= 30$	7.7	9.7	-5.9	-7.6	1.5	1.3	2.72	2.91
		SNR $= \infty$	7.8	9.6	-6.0	-7.7	1.8	1.5	2.53	2.67
	$\Delta K = 13$	SNR $= 30$	8.5	11.3	-7.6	-10.4	3.1	4.7	1.86	1.44
		SNR $= \infty$	8.4	11.2	-7.5	-10.4	3.7	5.5	1.75	1.38
AFC-CE	$\Delta K = 0$	SNR $= 30$	8.1	10.4	-6.1	-7.9	1.4	1.2	2.76	2.99
		SNR $= \infty$	8.2	10.6	-6.2	-8.2	1.8	1.4	2.58	2.79
	$\Delta K = 13$	SNR $= 30$	13.2	20.9	-10.7	-18.1	2.0	2.0	2.43	2.53
		SNR $= \infty$	13.5	21.1	-11.1	-18.2	2.4	2.5	2.26	2.40
	$\Delta K = 16$	SNR $= 30$	14.4	23.1	-11.8	-20.1	2.1	1.7	2.37	2.42
		SNR $= \infty$	14.7	22.9	-12.2	-20.1	2.5	2.1	2.21	2.29
	$\Delta K = 32$	SNR $= 30$	18.7	30.6	-15.3	-25.0	3.0	3.6	1.96	1.55
		SNR $= \infty$	19.1	30.7	-15.5	-23.8	3.6	4.0	1.80	1.48

5.4.1 PEM-AFROW Method

In this section, the performance of the state-of-art PEM-AFROW method is presented. Figure 5.5 shows the results obtained by the PEM-AFROW method for $\Delta K = 0$. In order to illustrate the bias problem in AFC, the results obtained by the NLMS algorithm when SNR = 30 dB are also considered. The PEM-AFROW method achieved $\overrightarrow{\Delta\text{MSG}} \approx 7.7$ dB and $\overrightarrow{\text{MIS}} \approx -6.3$ dB when SNR $= \infty$, and $\overrightarrow{\Delta\text{MSG}} \approx 8.0$ dB and $\overrightarrow{\text{MIS}} \approx -6.6$ dB when SNR = 30 dB. With respect to sound quality, the PEM-AFROW achieved $\overrightarrow{\text{SD}} \approx 1.7$ and $\overrightarrow{\text{WPESQ}} \approx 2.58$ when SNR $= \infty$, and $\overrightarrow{\text{SD}} \approx 1.4$ and $\overrightarrow{\text{WPESQ}} \approx 2.80$ when SNR = 30 dB.

Hereupon, $K(n)$ was increased in order to determine the MSBG achievable by the PEM-AFROW method. Such situation occurred with $\Delta K = 16$ dB for both ambient noise conditions. When SNR $= \infty$, this can be interpreted as an improvement in the method performance because the MSBG was achieved with $\Delta K = 14$ dB in the case of single feedback path. Figure 5.6 shows the results obtained by the PEM-AFROW method for $\Delta K = 16$ dB. The PEM-AFROW method achieved $\overrightarrow{\Delta\text{MSG}} \approx 14.4$ dB and $\overrightarrow{\text{MIS}} \approx -13.4$ dB when SNR $= \infty$, and $\overrightarrow{\Delta\text{MSG}} \approx 14.7$ dB and $\overrightarrow{\text{MIS}} \approx -14.5$ dB when SNR =

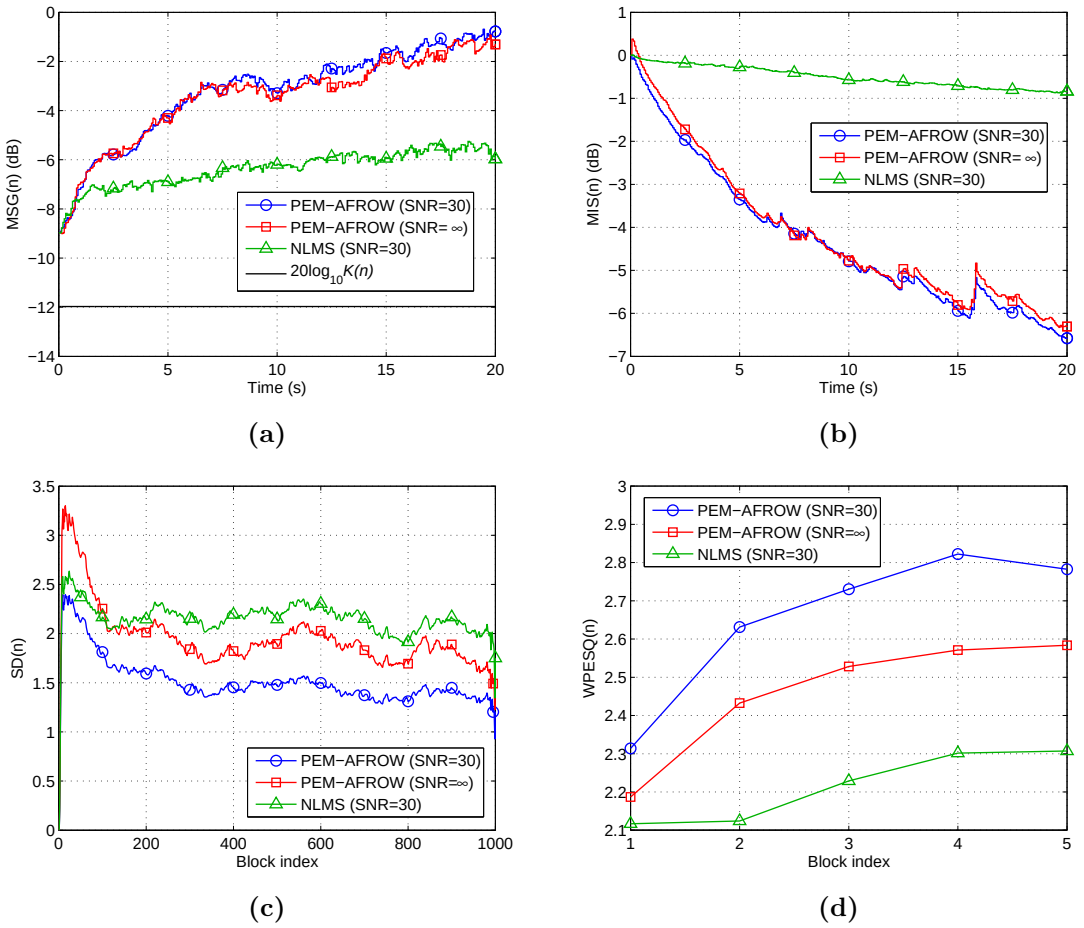


Figure 5.5: Average results of the PEM-AFROW method for speech signals and $\Delta K = 0$: (a) MSG(n); (b) MIS(n); (c) SD(n); (b) WPESQ(n).

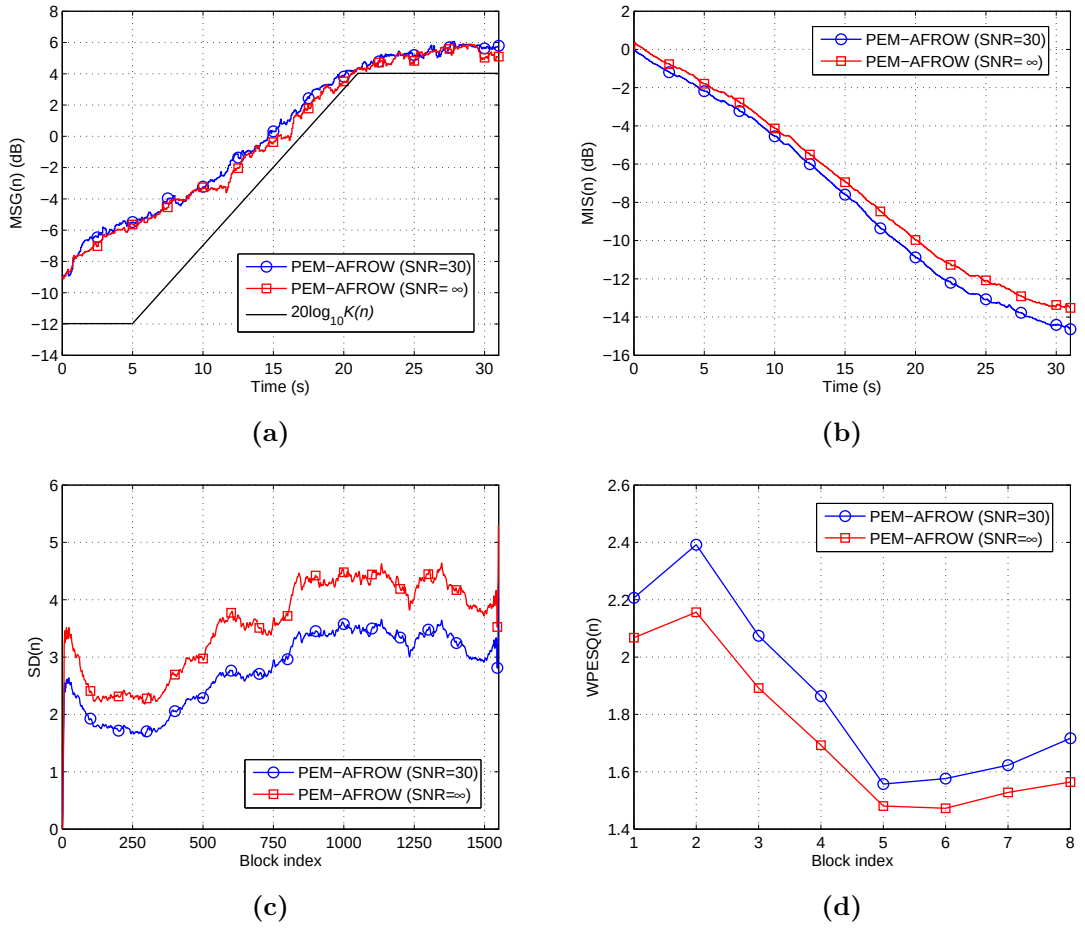


Figure 5.6: Average results of the PEM-AFROW method for speech signals and $\Delta K = 16$ dB: (a) MSG(n); (b) MIS(n); (c) SD(n); (b) WPESQ(n).

30 dB. Regarding sound quality, the PEM-AFROW achieved $\overrightarrow{\text{SD}} \approx 3.8$ and $\overrightarrow{\text{WPESQ}} \approx 1.55$ when SNR $= \infty$, and $\overrightarrow{\text{SD}} \approx 3.0$ and $\overrightarrow{\text{WPESQ}} \approx 1.67$ when SNR $= 30$ dB.

For $\Delta K = 0$ and 16 dB, the PEM-AFROW method performed worse in MIS(n) but performed better in SD(n) and WPESQ(n) when the system had multiple feedback paths. Regarding MSG(n), the PEM-AFROW method did not present a well-defined behavior which can be explained by the fact that the MSG(n) depends on the accuracy of $H(e^{j\omega}, n)$ in only one frequency component. This general behavior occurred because, for the same value of ΔK, there is an increase of 6.13 dB in the energy of the impulse response $\mathbf{f}(n)$ of the feedback path and a decrease of 9 dB in the broadband gain $K(n)$, as explained in Section 5.3.1. The combination of these two factors leads to a decrease in the energy of the feedback signal $\mathbf{f}(n) * x(n)$ while the energy of the system input signal $u(n)$ is unchanged. For instance, considering a PA system with white noise as $u(n)$ and $\Delta K = 0$, the feedback signal has 53% less energy when the system has multiple feedback paths.

Consequently, with multiple feedback paths, the ratio between the energies of the feedback signal (desired signal to the adaptive filter) and system input signal (interference

signal to the adaptive filter) is decreased for the same value of ΔK. This worsens the performance of the NLMS algorithm and, consequently, of the PEM-AFROW method. On the other hand, the feedback signal inserts less distortion in the error signal $e(n)$ even without any AFC method which improves the sound quality.

However, it could be expected that the reduction of 53% in the energy of the feedback signal $\mathbf{f}(n) * x(n)$ would imply a more pronounced worsening in method performance. On the other hand, it could be expected that the reduction of 11% in the sparseness of the impulse response $\mathbf{f}(n)$ of the feedback path would imply an improvement in method performance. The slight worsening in performance of the PEM-AFROW method is the outcome of the combination of these two factors.

In conclusion, ensuring on average the stability of the AFC system throughout the simulation time, the state-of-art PEM-AFROW method increased by 13.3 and 15.0 dB the MSG of the PA system with single feedback path when SNR $= \infty$ and 30 dB, respectively. When the system has multiple feedback paths, the PEM-AFROW method increased by 14.4 and 14.7 dB the MSG of the PA system when SNR $= \infty$ and 30 dB, respectively.

5.4.2 AFC-CM Method

This section presents and discusses the performance of the AFC-CM method. Figure 5.7 shows the results obtained by the AFC-CM method for $\Delta K = 0$. Again, the results obtained by the NLMS algorithm when SNR = 30 dB are considered in order to illustrate the bias problem in AFC. The AFC-CM method achieved $\overrightarrow{\Delta\text{MSG}} \approx 9.6$ dB and $\overrightarrow{\text{MIS}} \approx -7.7$ dB when SNR $= \infty$, and $\overrightarrow{\Delta\text{MSG}} \approx 9.7$ dB and $\overrightarrow{\text{MIS}} \approx -7.6$ dB when SNR = 30 dB. With respect to sound quality, the AFC-CM achieved $\overrightarrow{\text{SD}} \approx 1.5$ and $\overrightarrow{\text{WPESQ}} \approx 2.67$ when SNR $= \infty$, and $\overrightarrow{\text{SD}} \approx 1.3$ and $\overrightarrow{\text{WPESQ}} \approx 2.91$ when SNR = 30 dB.

For $\Delta K = 0$, as occurred with the PEM-AFROW, the AFC-CM method performed worse in MIS(n) but performed better in SD(n) and WPESQ(n) when the system had multiple feedback paths. Regarding MSG(n), the AFC-CM method did not present a well-defined behavior which can be explained by the fact that the MSG(n) depends on the accuracy of $H(e^{j\omega}, n)$ in only one frequency component. The worsening in MSG(n) and MIS(n) is due to the decrease in the highest absolute values of $\mathbf{g}(n) * \mathbf{f}(n)$ for the same value of ΔK, as can be observed in Figure 5.4a. For the same system input signal $u(n)$, this makes the estimation of these values of $\mathbf{g}(n) * \mathbf{f}(n)$ from $\mathbf{c_y}(n)$ more difficult. And, as these are the values that contribute most to the feedback problem, this fact worsens the performance of the proposed AFC-CM method. On the other hand, the improvement in SD(n) and WPESQ(n) is due to the lower energy of the feedback signal $\mathbf{f}(n) * x(n)$ for the same value of ΔK, as explained in Section 5.4.1. This leads to less distortion in the error signal $e(n)$ resulting from the feedback signal even without any AFC method and improves the system sound quality.

Hereupon, $K(n)$ was increased in order to determine the MSBG achievable by the AFC-CM method. Such situation occurred with $\Delta K = 13$ dB for both ambient noise

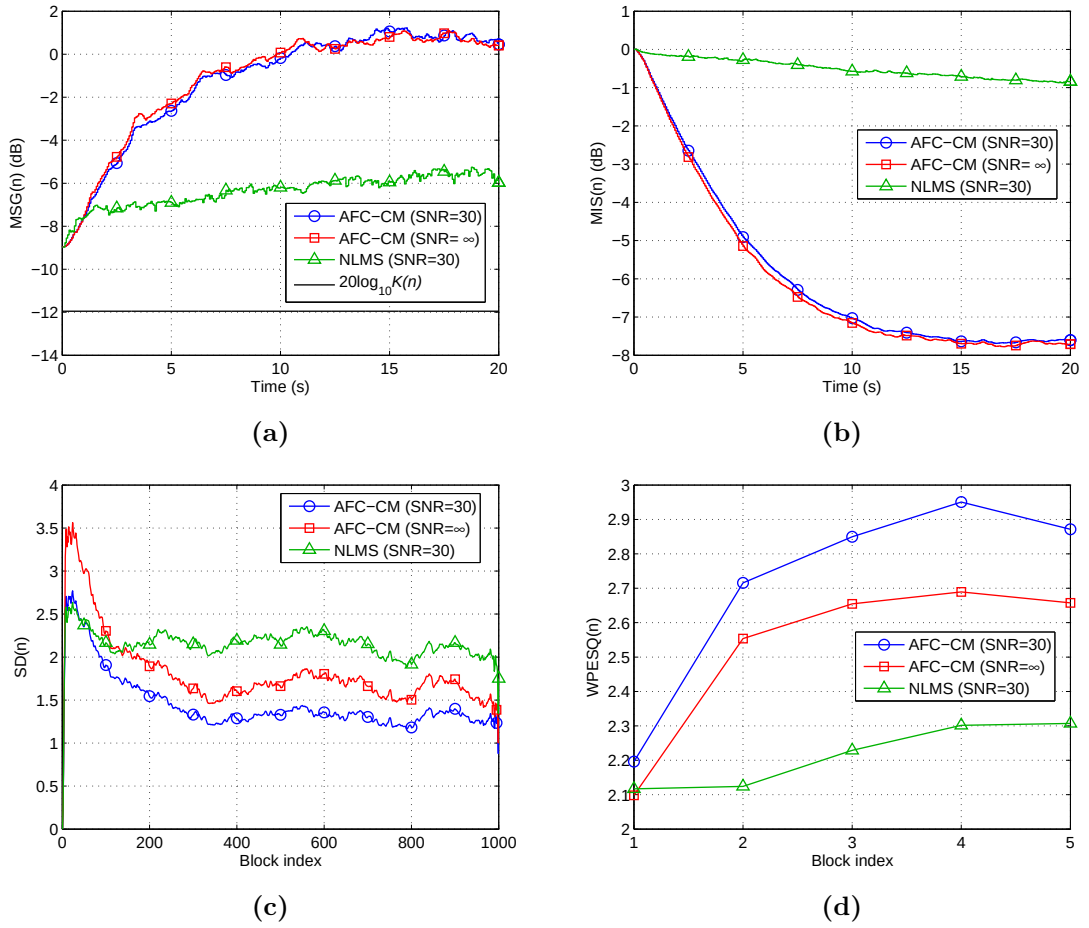


Figure 5.7: Average results of the AFC-CM method for speech signals and $\Delta K = 0$: (a) MSG(n); (b) MIS(n); (c) SD(n); (b) WPESQ(n).

conditions. This represents a worsening in the method performance because its MSBG was achieved with $\Delta K = 14$ dB when the system had a single feedback path. Figure 5.8 shows the results obtained by the AFC-CM method for $\Delta K = 13$ dB. The AFC-CM method achieved $\overrightarrow{\Delta \text{MSG}} \approx 11.2$ dB and $\overrightarrow{\text{MIS}} \approx -10.4$ dB when SNR $= \infty$, and $\overrightarrow{\Delta \text{MSG}} \approx 11.3$ dB and $\overrightarrow{\text{MIS}} \approx -10.4$ dB when SNR $= 30$ dB. Regarding sound quality, the AFC-CM achieved $\overrightarrow{\text{SD}} \approx 5.5$ and $\overrightarrow{\text{WPESQ}} \approx 1.38$ when SNR $= \infty$, and$\overrightarrow{\text{SD}} \approx 4.7$ and $\overrightarrow{\text{WPESQ}} \approx 1.44$ when SNR $= 30$ dB.

Section 4.3.1 explained in detail that the broadband gain $K(n)$ of the forward path must be lower than the MSG of the PA system in order to simultaneously fulfill the conditions $\left|G(e^{j\omega}, n)D(e^{j\omega})H(e^{j\omega}, n)\right| < 1$ and $\left|G(e^{j\omega}, n)D(e^{j\omega})\left[F(e^{j\omega}, n) - H(e^{j\omega}, n)\right]\right| < 1$, which are required to defined $\mathbf{c_y}(n)$ according to (4.16). The results presented in the Section 4.7.2.1 demonstrated that, when the system had a single feedback path, these conditions limited the increase in the broadband gain, ΔK, in 14 dB. The results presented in this section demonstrated that these conditions limited ΔK in 13 dB when the system had multiple feedback paths, thereby limiting even more the performance of the AFC-CM

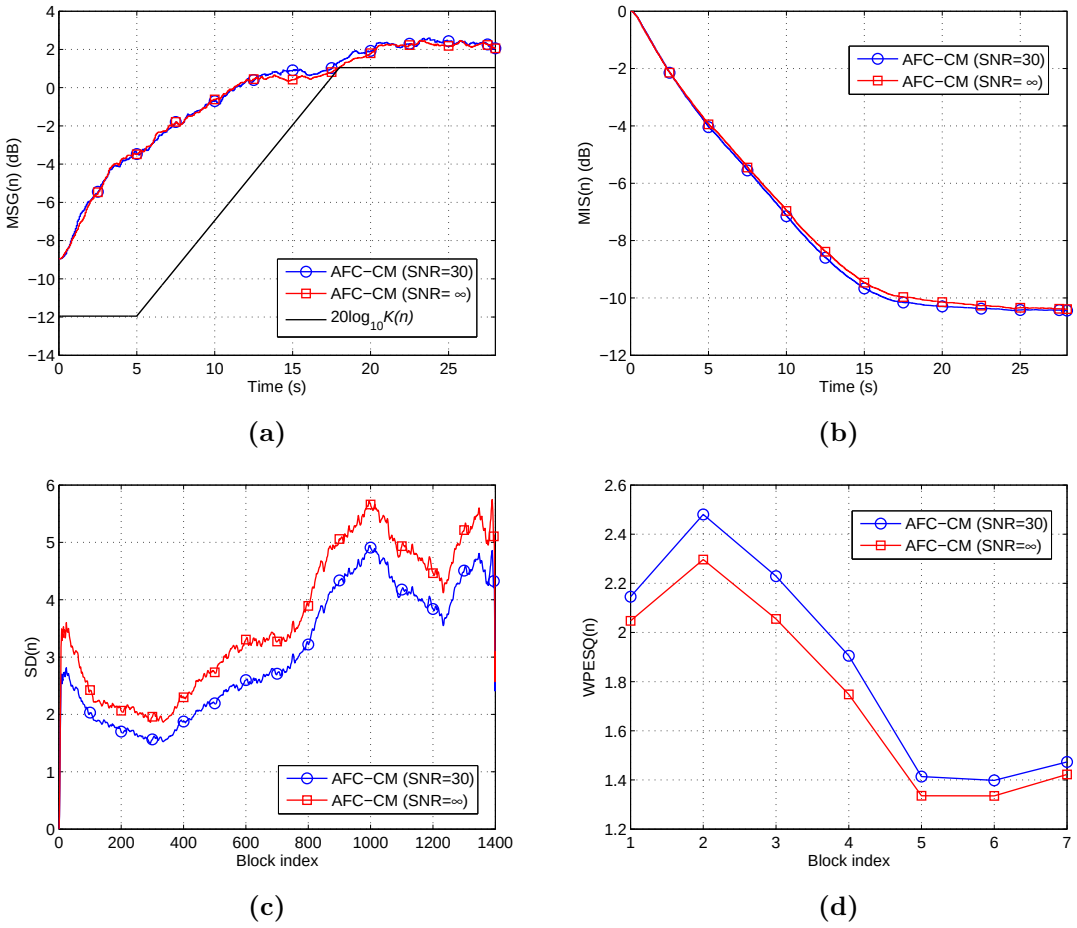


Figure 5.8: Average results of the AFC-CM method for speech signals and $\Delta K = 13$ dB: (a) MSG(n); (b) MIS(n); (c) SD(n); (b) WPESQ(n).

method in comparison with the case of single feedback path.

In conclusion, ensuring on average the stability of the AFC system throughout the simulation time, the proposed AFC-CM method increased by 12 dB the MSG of the PA system with single feedback path when SNR $= \infty$ or 30 dB. When the system has multiple feedback paths, the proposed AFC-CM method increased by 11.2 and 11.3 dB the MSG of the PA system when SNR $= \infty$ and 30 dB, respectively.

5.4.3 AFC-CE Method

Similarly, this section addresses the performance of the proposed AFC-CE method. Figure 5.9 shows the results obtained by the AFC-CE method for $\Delta K = 0$. Once again, the results obtained by the NLMS algorithm when SNR $= 30$ dB are also included. The AFC-CE method achieved $\overrightarrow{\Delta\text{MSG}} \approx 10.6$ dB and $\overrightarrow{\text{MIS}} \approx -8.2$ dB when SNR $= \infty$, and $\overrightarrow{\Delta\text{MSG}} \approx 10.4$ dB and $\overrightarrow{\text{MIS}} \approx -7.9$ dB when SNR $= 30$ dB. Regarding the sound quality, the AFC-CE achieved $\overrightarrow{\text{SD}} \approx 1.4$ and $\overrightarrow{\text{WPESQ}} \approx 2.79$ when SNR $= \infty$, and $\overrightarrow{\text{SD}} \approx 1.2$ and $\overrightarrow{\text{WPESQ}} \approx 2.99$ when SNR $= 30$ dB.

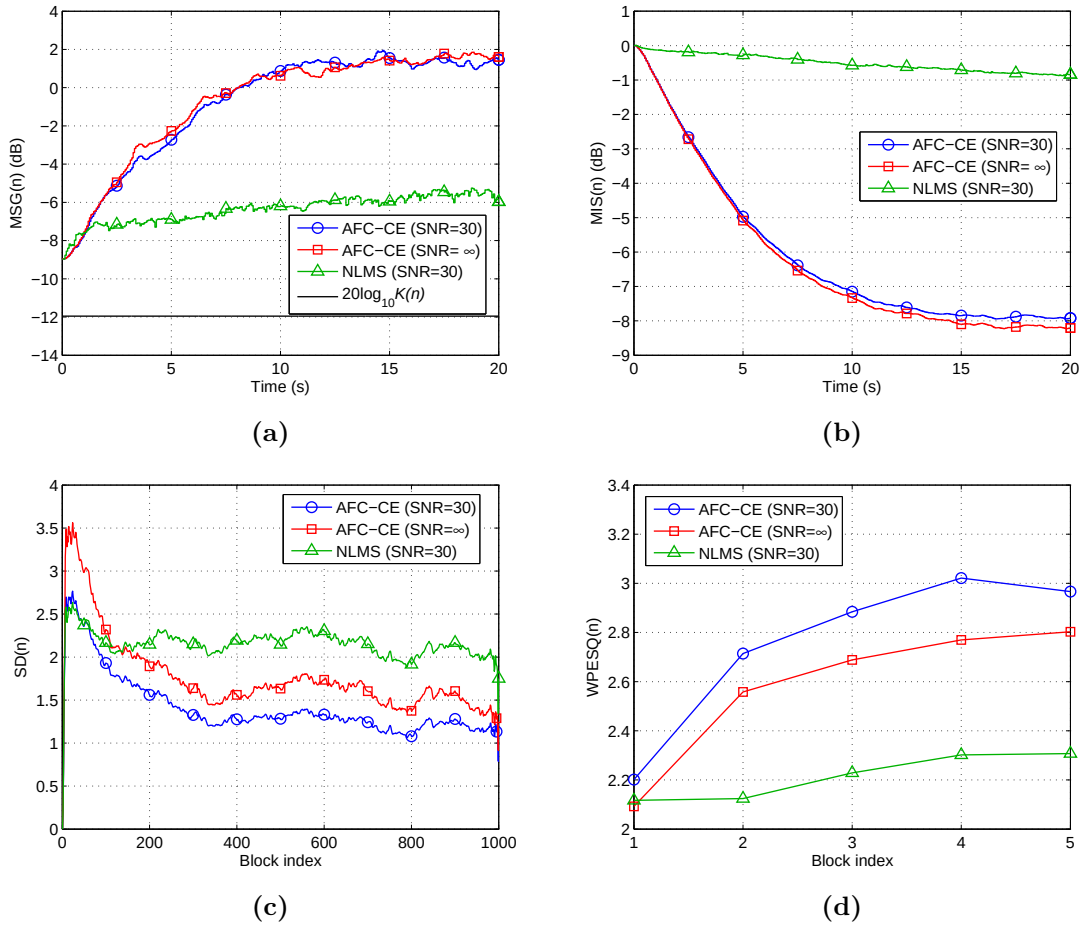


Figure 5.9: Average results of the AFC-CE method for speech signals and $\Delta K = 0$: (a) MSG(n); (b) MIS(n); (c) SD(n); (b) WPESQ(n).

For $\Delta K = 0$, the AFC-CE performed worse in MSG(n) and MIS(n) but performed better in SD(n) and WPESQ(n) when the system had multiple feedback paths. Similarly to the AFC-CM, the worsening in MSG(n) and MIS(n) is due to the decrease in the highest absolute values of $\mathbf{g}(n) * [\mathbf{f}(n) - \mathbf{h}(n)]$ for the same value of ΔK. For the same system input signal $u(n)$, this makes the estimation of these values of $\mathbf{g}(n) * [\mathbf{f}(n) - \mathbf{h}(n)]$ from $\mathbf{c_e}(n)$ more difficult. And, since these are the values that contribute most to the feedback problem, this worsens the performance of the proposed AFC-CE method. On the other hand, the improvement in SD(n) and WPESQ(n) is due to the lower energy of the feedback signal $\mathbf{f}(n) * x(n)$ for the same value of ΔK, as explained in Section 5.4.1. This leads to less distortion in the error signal $e(n)$ resulting from the feedback signal even without any AFC method and improves the system sound quality.

Hereupon, $K(n)$ was increased in order to determine the MSBG achievable by the AFC-CE method. Such situation occurred with an impressive $\Delta K = 32$ dB for both ambient noise conditions. This represents an improvement in the method performance because its MSBG was achieved with $\Delta K = 30$ dB when the system had a single feedback

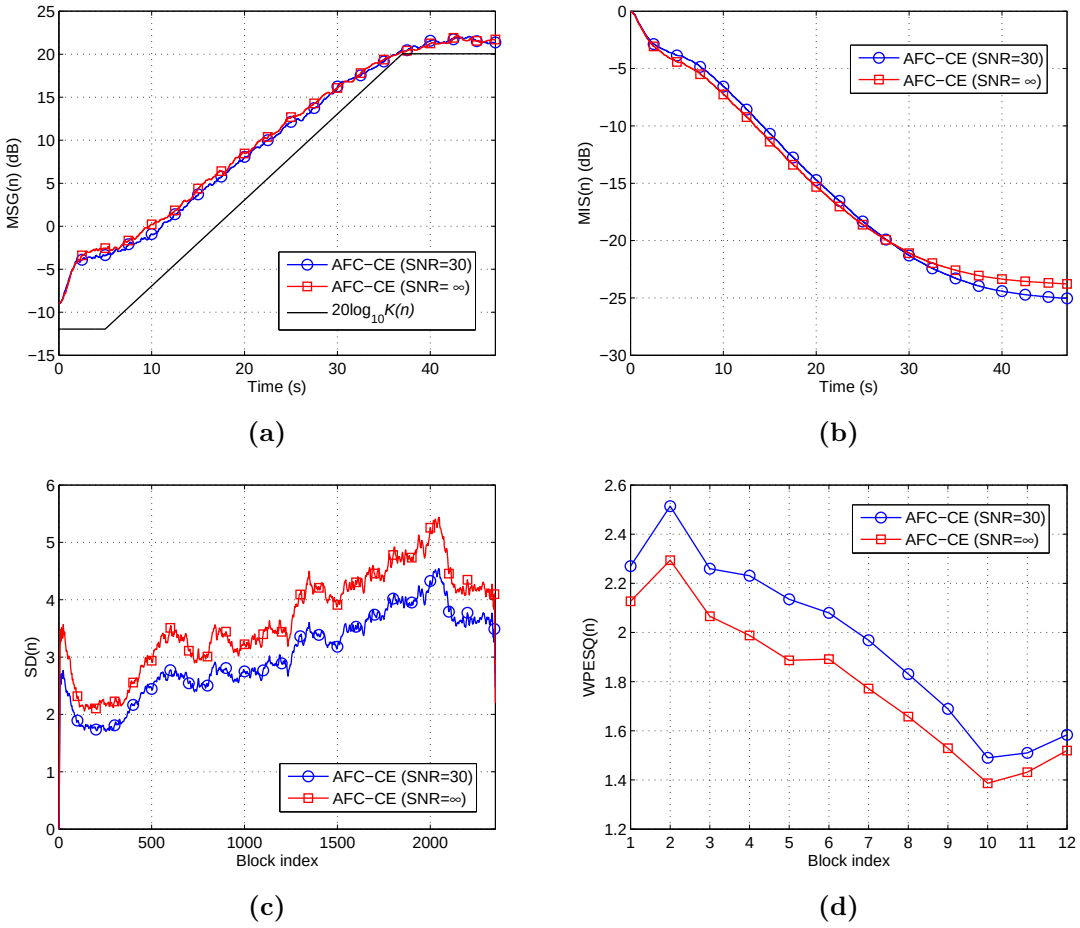


Figure 5.10: Average results of the AFC-CE method for speech signals and $\Delta K = 32$ dB: (a) MSG(n); (b) MIS(n); (c) SD(n); (b) WPESQ(n).

path. Figure 5.10 shows the results obtained by the AFC-CE method for $\Delta K = 32$ dB. The AFC-CE method achieved $\overrightarrow{\Delta\text{MSG}} \approx 30.7$ dB and $\overrightarrow{\text{MIS}} \approx -23.8$ dB when SNR $= \infty$, and $\overrightarrow{\Delta\text{MSG}} \approx 30.6$ dB and $\overrightarrow{\text{MIS}} \approx -25.0$ dB when SNR $= 30$ dB. With respect to sound quality, the AFC-CE achieved $\overrightarrow{\text{SD}} \approx 4.0$ and $\overrightarrow{\text{WPESQ}} \approx 1.48$ when SNR $= \infty$, and $\overrightarrow{\text{SD}} \approx 3.6$ and $\overrightarrow{\text{WPESQ}} \approx 1.55$ when SNR $= 30$ dB.

In the second configuration of the forward path $G(q, n)$ where its broadband $K(n)$ increases over time, i.e., $\Delta K > 0$, the proposed AFC-CE method performed better when the system had multiple feedback paths. Regarding MSG(n) and MIS(n), this is due to the lower sparseness of the impulse response $\mathbf{f}(n)$ of the feedback path that causes a larger number of samples of $\mathbf{g}(n) * [\mathbf{f}(n) - \mathbf{h}(n)]$ to be accurately estimated from $\mathbf{c_e}(n)$ as ΔK increases. As a consequence, a larger number of samples of $\mathbf{f}(n)$ is also accurately estimated which improves the performance of the AFC-CE method. With respect to sound quality, the improvement occurs because, with $\Delta K = 32$ dB and multiple feedback paths, the feedback signal $\mathbf{f}(n) * x(n)$ has lower energy than with $\Delta K = 30$ dB and single feedback path. Thus, the feedback signal inserts less distortion in the error signal $e(n)$ even without

any AFC method.

In conclusion, ensuring on average the stability of the AFC system throughout the simulation time, the proposed AFC-CE method increased by 29.6 and 30 dB the MSG of the PA system with single feedback path when SNR $= \infty$ and 30 dB, respectively. When the system had multiple feedback paths, the proposed AFC-CE method increased by 30.7 and 30.6 dB the MSG of the PA system when SNR $= \infty$ and 30 dB, respectively.

5.4.4 Performance Comparison

After the evaluation and discussion of their individual performances, the AFC-CM and AFC-CE will be now compared with the state-of-art PEM-AFROW method. The comparison will focus on the results obtained with SNR = 30 dB because this ambient noise condition is closer to real-world conditions.

Figure 5.11 compares the results obtained by the AFC methods under evaluation in the first configuration of forward path, where its broadband gain $K(n)$ remained constant, i.e., for $\Delta K = 0$. It can be observed that the AFC-CM and AFC-CE methods presented similar performances, with a slight advantage for the AFC-CE, and both methods outperformed the PEM-AFROW. The proposed AFC-CE method achieved $\overrightarrow{\Delta\text{MSG}} \approx 10.4$ dB and $\overrightarrow{\text{MIS}} \approx -7.9$ dB, outscoring respectively the AFC-CM by 0.7 dB and 0.3 dB and the PEM-AFROW by 2.4 dB and 1.3 dB. With respect to sound quality, the AFC-CE achieved $\overrightarrow{\text{SD}} \approx 1.2$ and $\overrightarrow{\text{WPESQ}} \approx 2.99$ outscoring respectively the AFC-CM by 0.1 and 0.08, and the PEM-AFROW by 0.2 and 0.19. These differences are hardly noticeable audibly and were caused by the fact that, with the constant value of $K(n)$ and the increase in MSG provided by all the AFC methods, the systems were too far from the instability as can be observed in Figure 5.11a.

Consider now the second configuration of the broadband gain $K(n)$ of the forward path where it was linearly (in dB scale) increased, as explained in Section 3.6.1, in order to determine the MSBG of each method. The MSBG was defined as the maximum value of K_2 with which an AFC method achieves a MSG(n) completely stable. The AFC-CE method achieved a MSBG of the forward path $G(q, n)$ equal to 20 dB, outperforming the AFC-CM and the state-of-art PEM-AFROW by impressive 19 dB and 16 dB, respectively. This would be enough to conclude that the proposed AFC-CE method has the best performance. However, aiming to enrich the discussion, the performance of the AFC methods under evaluation will be compared considering the results obtained with all the values of ΔK used in this work.

Figure 5.12 compares the results obtained by the AFC methods under evaluation for $\Delta K = 13$ dB. It can be observed that the AFC-CM performed well, even better than the PEM-AFROW, until 10 s of simulation. After this time, as explained in Section 4.7.2.1, the performance of the AFC-CM method was limited by the inaccuracy of (4.16). This behavior is easily observed in MIS(n) showed in Figure 5.12b. However, it is evident that the AFC-CE stood out from both methods by achieving $\overrightarrow{\Delta\text{MSG}} \approx 20.9$ dB and

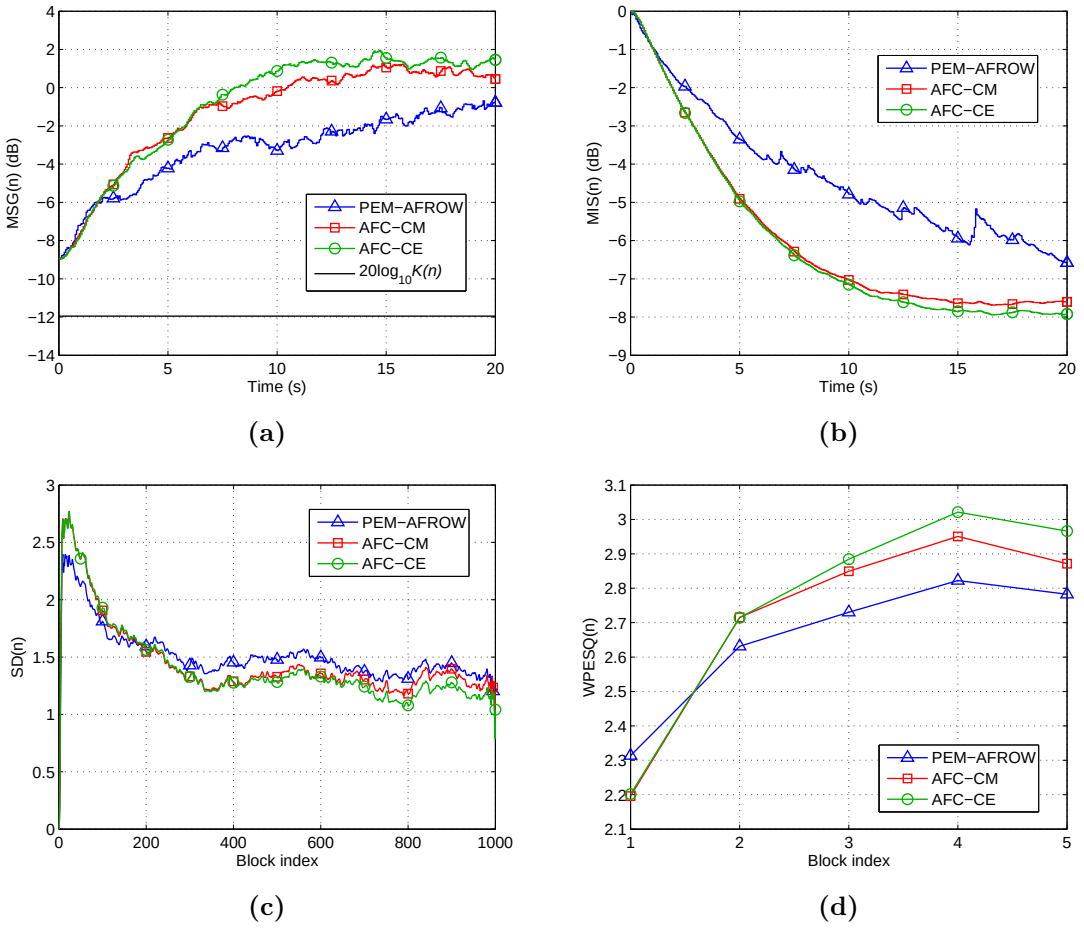


Figure 5.11: Performance comparison between the PEM-AFROW, AFC-CM and AFC-CE methods for speech signals and $\Delta K = 0$: (a) MSG(n); (b) MIS(n); (c) SD(n); (b) WPESQ(n).

$\overrightarrow{\text{MIS}} \approx -18.1$ dB, outscoring respectively the AFC-CM by 9.6 dB and 7.7 dB and the PEM-AFROW by 6.2 dB and 5.0 dB. Moreover, it should be noted that the AFC-CM method outperformed the PEM-AFROW by 0.1 dB with respect to $\overline{\Delta\text{MSG}}$, which was the cost function in the optimization of the adaptive filters parameters for all methods.

Regarding sound quality, the AFC-CM method presented the worst performance by obtaining $\overrightarrow{\text{SD}} \approx 4.7$ and $\overrightarrow{\text{WPESQ}} \approx 1.44$ because of its very low stability margin after $t = 17$ s, as can be observed in Figure 5.12a. Although its MSG(n) is completely stable, some instability occurred for a few signals which resulted in excessive reverberation or even in some howlings in the error signal $e(n)$. On the other hand, the AFC-CE method presented the best sound quality by achieving $\overrightarrow{\text{SD}} \approx 2.0$ and $\overrightarrow{\text{WPESQ}} \approx 2.53$ because of its largest stability margin and outscored the PEM-AFROW by 1.1 and 0.7, respectively.

Finally, Figure 5.13 compares the results obtained by the PEM-AFROW and AFC-CE methods for $\Delta K = 16$ dB. Once again, it can be observed that the AFC-CE method outperformed the PEM-AFROW. The PEM-AFROW obtained $\overrightarrow{\Delta\text{MSG}} \approx 14.7$ dB and $\overrightarrow{\text{MIS}} \approx -14.5$ dB while the AFC-CE method achieved $\overrightarrow{\Delta\text{MSG}} \approx 23.1$ dB and $\overrightarrow{\text{MIS}} \approx -20.1$ dB.

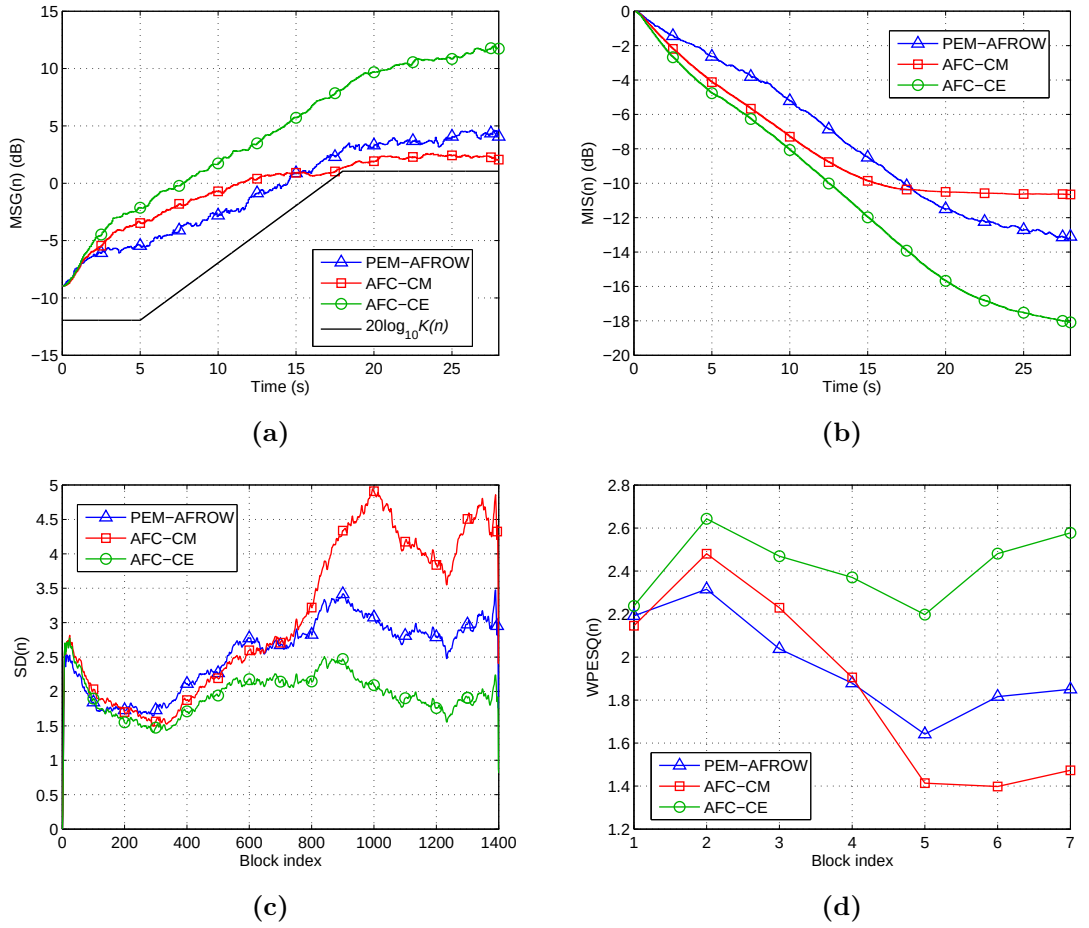


Figure 5.12: Performance comparison between the PEM-AFROW, AFC-CM and AFC-CE methods for speech signals and $\Delta K = 13$ dB: (a) MSG(n); (b) MIS(n); (c) SD(n); (b) WPESQ(n).

Regarding sound quality, the AFC-CE method also presented the best performance by achieving $\overrightarrow{\text{SD}} \approx 1.7$ and $\overrightarrow{\text{WPESQ}} \approx 2.42$ while the PEM-AFROW obtained $\overrightarrow{\text{SD}} \approx 3.0$ and $\overrightarrow{\text{WPESQ}} \approx 1.67$.

In conclusion, the proposed AFC-CE method increased by 30.6 dB the MSG of the PA system, outperforming the AFC-CM and PEM-AFROW by 19.3 and 15.9 dB, respectively. Moreover, the AFC-CE method estimated the impulse response of the feedback path with an MIS of -25 dB, outperforming the AFC-CM and PEM-AFROW by 14.6 and 10.5 dB, respectively. And even with the same variation in the broadband gain of the forward path, ΔK, the AFC-CE always outperformed the other methods not only in MSG(n) and MIS(n) but also in SD(n) and WPESQ(n).

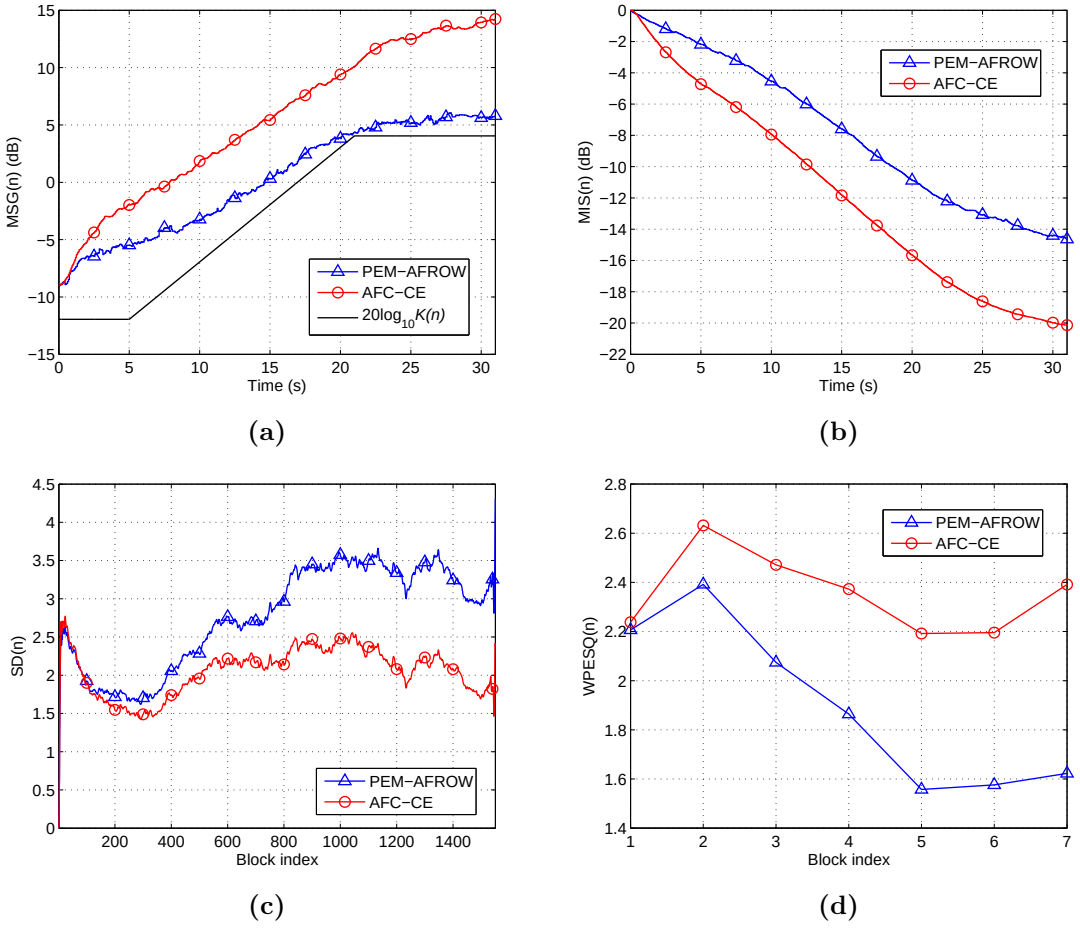


Figure 5.13: Performance comparison between the PEM-AFROW and AFC-CE methods for speech signals and $\Delta K = 16$ dB: (a) MSG(n); (b) MIS(n); (c) SD(n); (b) WPESQ(n).

5.5 Conclusion

Chapters 3 and 4 addressed the AFC problem considering PA systems with only one microphone and one loudspeaker. In fact, this configuration is practically the only one found in the literature and represents several practical applications of PA systems as, for instance, in hearing aid. However, this configuration may not precisely represent the use of PA systems in other practical applications as, for instance, in very large environments.

Typically, aiming to be heard by a large audience in the same acoustic environment, a speaker uses a PA system with one microphone, responsible for picking up his/her own voice, one amplification system, responsible for amplifying the voice signal, and several loudspeakers placed in different positions, responsible for playback and distributing the voice signal in the acoustic environment so that everyone in the audience can hear it. This results in an PA system with multiple acoustic feedback paths.

This chapter dealt with the AFC problem considering PA systems with one microphone and four loudspeakers. It was demonstrate that the impulse response of the resulting overall feedback path generally has a larger number of prominent peaks and lower sparseness. In addition, its frequency components have, in general, a higher energy. The influence of both characteristics on the performance of the state-of-art PEM-AFROW and the proposed AFC-CM and AFC-CE methods was discussed. Finally, an evaluation of the AFC methods was carried out in a simulated environment.

Simulation results demonstrated that, if the broadband gain of the forward path is not increased, all the AFC methods under evaluation performed worse in a PA system with multiple feedback paths than with single feedback path. On the other hand, the sound quality of the AFC systems was improved because the feedback signal had lower energy and thus inserted less distortion in the system input signal. Moreover, in comparison with the case of single feedback path, the MSBG achieved by the PEM-AFROW, AFC-CM and AFC-CE methods remained constant, decreased 1 dB and increased 2 dB, respectively.

In conclusion, when the source signal is speech, the proposed AFC-CE method can estimate the feedback path impulse response with a MIS of -25 dB, outperforming the state-of-art PEM-AFROW and the proposed AFC-CM by respectively 10.5 and 14.6 dB. Moreover, the proposed AFC-CE method can increase by 30.6 dB the MSG of the PA system, outperforming the PEM-AFROW and AFC-CM by respectively 15.9 and 19.3 dB. It may be concluded that, with multiple feedback paths, the proposed AFC-CE method achieved a less biased estimate of the acoustic feedback path and further increased the MSG of the PA system in comparison with the AFC-CM and PEM-AFROW methods.

Part II

Acoustic Echo Cancellation

Chapter 6

Acoustic Echo Cancellation

6.1 Introduction

This chapter addresses the topic of acoustic echo cancellation in teleconference systems. Similar to the AFC approach, the AEC approach uses an adaptive filter to identify the acoustic echo path and estimate the echo signal that is subtracted from the microphone signal. During the last decades, the use of the traditional gradient-based and least-squares-based adaptive filtering algorithms has been established in AEC applications.

The cepstral analysis, which was successfully applied to the AFC problem in the previous chapters, is now applied to the AEC problem. The independence between the loudspeaker and microphone signals of the same room in the AEC application is exploited to develop a new AEC method based on cepstral analysis. Moreover, two improved versions that perform the inverse of the overlap-and-add method using the adaptive filter as an estimate of the echo path are also proposed. An evaluation of the proposed AEC methods is carried out in a simulated environment. It is demonstrated that the AEC methods based on cesptral analysis are able to outperform the NLMS and BNDR-LMS, adaptive filtering algorithms widely used in practical applications, but they can present a worse performance in the first seconds of echo cancellation.

Hence, to combine the strengths of both methodologies, hybrid AEC methods are also proposed. The hybrid methods update the adaptive filter through the NLMS or BNDR-LMS algorithms most of the time and the AEC methods based on cepstral analysis are sporadically used to accelerate or straighten the learning process. An evaluation of the proposed AEC methods is carried out in the same simulated environment used for the individual methods. It is demonstrated that hybrid AEC methods are able to outperform the individual methods with regard to both misalignment and echo cancellation. This means that the AEC methods based on cepstral analysis can be used alone or to improve the performance of the traditional adaptive filtering algorithms in AEC applications.

6.2 The Acoustic Echo Problem

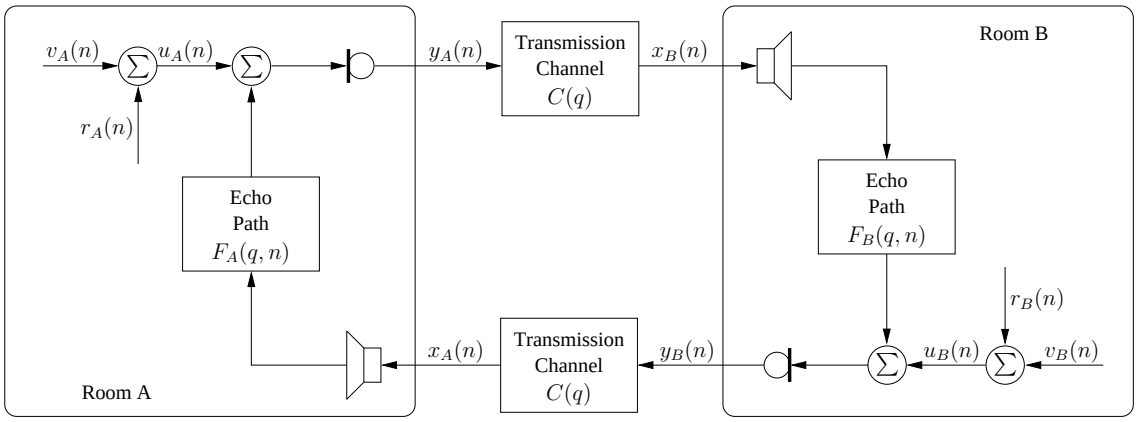


Figure 6.1: Acoustic echo in a teleconference system.

A typical teleconference system is illustrated in Figure 6.1 considering two acoustic environments, rooms A and B, with one microphone and one loudspeaker. In room B, the loudspeaker signal $x_B(n)$ may return to the microphone through the echo path $F_B(q,n)$. The acoustic echo signal $\mathbf{f}_B(n) * x_B(n)$ is added to the speaker signal $v_B(n)$ and to the ambient noise $r_B(n)$, generating the microphone signal $y_B(n)$. The same occurs in room A such that

$$\begin{aligned} y_A(n) &= \mathbf{f}_A(n) * x_A(n) + v_A(n) + r_A(n) \\ y_B(n) &= \mathbf{f}_B(n) * x_B(n) + v_B(n) + r_B(n). \end{aligned} \tag{6.1}$$

The transmission channel is the medium by which the speaker signals are transmitted from a room to another. It is usually defined as a time delay and is denoted as

$$\begin{aligned} C(q) &= c_{L_C-1} q^{-(L_C-1)} \\ &= \mathbf{c}^T \mathbf{q} \end{aligned} \tag{6.2}$$

Let the room input signals $u_A(n)$ and $u_B(n)$ be the sum of the respective speaker and ambient noise signals, i.e., $u_A(n) = v_A(n) + r_A(n)$ and $u_B(n) = v_B(n) + r_B(n)$, and also include the characteristics of the microphones and A/D converter. The room input signal $u_B(n)$ and the loudspeaker signal $x_B(n)$ are related by the transfer function

$$x_B(n) = \frac{C(q)\left[u_A(n) + F_A(q,n)C(q)u_B(n)\right]}{1 - F_A(q,n)C(q)C(q)F_B(q,n)}. \tag{6.3}$$

If $u_A(n) = 0$, (6.3) becomes

$$x_B(n) = \frac{F_A(q,n)C(q)C(q)}{1 - F_A(q,n)C(q)C(q)F_B(q,n)} u_B(n). \tag{6.4}$$

Comparing (6.4) and (2.4), it can be concluded that the teleconference system depicted in Figure 6.1 is equivalent to the PA system depicted in Figure 2.1 if $F_A(q,n)C(q)C(q) = G(q,n)D(q)$. Indeed, the two systems are equivalent if the echo path $F_A(q,n)$ is equal to the forward path $G(q,n)$ and the transmission channel $C(q)$ applies the half delay of the delay filter $D(q)$. The same analogy can be made for the room A. Therefore, a teleconference system has also a closed-loop signal and can become unstable, resulting in a howling artifact, the Larsen effect, that will be audible in both rooms. However, the instability issue is more critical in a PA system for two reasons. First, the frequency response $G(e^{j\omega},n)$ of the forward path generally has much higher magnitude than the frequency response $F_A(e^{j\omega},n)$ of the echo path. Second, the techniques to suppress or cancel the echo signals in a teleconference system are applied in both rooms, leading to a residual closed-loop signal with very little energy such that it is ignored.

The difference between the concepts of acoustic echo and feedback is straightforward: in acoustic echo, it is assumed that there is no closed-loop signal and thereby the communication system is always stable. Hence, the acoustic echo limits the performance of a teleconference or hands-free communication system only with regard to sound quality. If no signal processing is applied, the microphone signals $y_A(n)$ and $y_B(n)$, defined in (6.1), are sent over the transmission channel to the rooms B and A, respectively, containing the echo signals. As a consequence, after talking, a speaker receives back his own voice that, owing to the delay of hundreds of milliseconds caused by the transmission channel, is easily distinguished from the speaker's signal and sounds like an echo. The occurrence of this acoustic echo is annoying for the audience in both rooms and disturbs the communication. Therefore, the acoustic echo signals should be eliminated or, at least, attenuated.

In order to attenuate the acoustic echo, two approaches have been developed over the last 20 years: acoustic echo suppression (AES) and acoustic echo cancellation (AEC). The former, also called loss control, attenuates the loudspeaker and/or microphone signals depending on the comparison between their energies with pre-defined thresholds and between themselves [12, 20]. Similarly to AFC, the latter estimates the echo signal by means of adaptive filter and subtracts it from the microphone signal [12, 21].

The operation of AES is simple. If only the loudspeaker signal is active, it attenuates the microphone signal in order to avoid the transmission of acoustic echo. If only the speaker signal is active, it attenuates the loudspeaker signal in order to avoid the reception of noise. The problem occurs when both loudspeaker and speaker signals are simultaneously active, which is defined as a double-talk situation [12]. In this case, the method decides which signal, of the loudspeaker or microphone, is attenuated. Therefore, AES methods preclude full-duplex communication [12]. In fact, the AES approach assumes the existence of the acoustic echo and only concerns to control it.

Nowadays, the AES approach is practically in disuse and the AEC approach is widely used in teleconference and hands-free communication systems. Its drawback compared with the AES approach is a higher computational complexity.

6.3 Mono-channel Acoustic Echo Cancelation

The AEC has stabilized in the past years as the state-of-art approach to remove or, at least, attenuate the effects of acoustic echo in teleconference and hands-free communication systems [72, 74, 84]. The AEC methods identify and track the echo path $F(q,n)$ using an adaptive filter that is generally defined as a FIR filter

$$\begin{aligned} H(q,n) &= h_0(n) + h_1(n)q^{-1} + \ldots + h_{L_H-1}(n)q^{-(L_H-1)} \\ &= \mathbf{h}^T(n)\mathbf{q} \end{aligned} \tag{6.5}$$

with length L_H.

Then, an estimate of the echo signal $\mathbf{f}(n) * x(n)$ is computed as $\mathbf{h}(n) * x(n)$ and subtracted from the microphone signal $y(n)$, generating the error signal

$$\begin{aligned} e(n) &= y(n) - \mathbf{h}(n) * x(n) \\ &= u(n) + \mathbf{f}(n) * x(n) - \mathbf{h}(n) * x(n) \\ &= u(n) + [\mathbf{f}(n) - \mathbf{h}(n)] * x(n), \end{aligned} \tag{6.6}$$

which is effectively the signal to be sent over the transmission channel. Such a scheme is shown in Figure 6.2 [72, 74, 84]. It is noteworthy that, from (6.6), the amount of acoustic echo present in the error signal $e(n)$ depends on $\mathbf{f}(n) - \mathbf{h}(n)$, the waveform of the mismatch between the impulse responses of the echo path and adaptive filter. If the adaptive filter exactly matches the echo path, i.e., $H(q,n) = F(q,n)$, the error signal $e(n)$ will contain no acoustic echo.

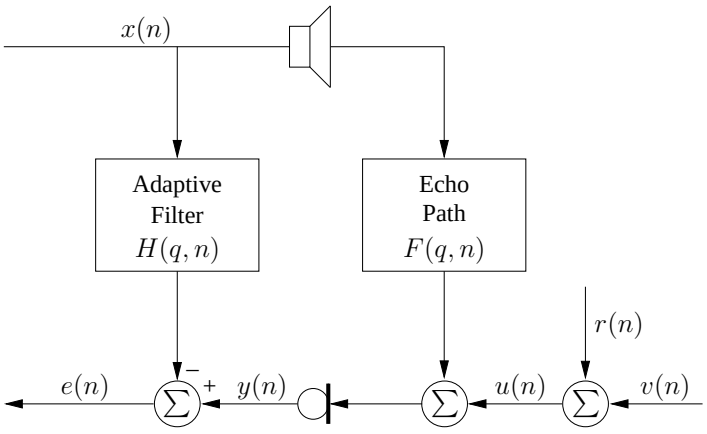


Figure 6.2: Mono-channel AEC.

Obviously, the adaptive filter $H(q,n)$ should only be updated when the microphone signal $y(n)$ is active and contains acoustic echo, i.e., when $y(n) \neq 0$ and $x(n) \neq 0$. Voice activity detectors (VAD) are generally used to detect this situation. However, when the source signal $v(n)$ is also active, i.e., when $v(n) \neq 0$, $y(n) \neq 0$ and $x(n) \neq 0$, a situation

called double-talk is declared [85, 86, 87, 88, 89]. In this case, if the traditional gradient-based or least-square-based adaptive algorithms are used to update the adaptive filter $H(q,n)$, the speaker signal $v(n)$ acts as noise to $H(q,n)$ because it prevents the error signal $e(n)$ to approach zero even if the ideal solution $H(q,n) = F(q,n)$ is achieved, as can be observed from (6.6). In fact, both ambient noise $r(n)$ and speaker signal $v(n)$ act as noise to $H(q,n)$ but $v(n)$ is much more harmful due to its higher intensity. As a consequence, the speaker signal $v(n)$ can disrupt the adaptation of $H(q,n)$ and cause its divergence. Therefore, the adaptive filter should not be updated when $v(n) \neq 0$. A double-talk detector (DTD) is used to detect if the speaker signal $v(n)$ is active or not.

Any adaptive filtering algorithm can be used in AEC. However, mostly gradient-based or least-squares-based adaptive algorithms are generally found in the literature. For these cases, there are time-domain, time-domain block, fullband frequency-domain and subband frequency-domain algorithms. Some can perform better than others depending on their characteristics as, for example, convergence speed, robustness to noisy environments and to short-time disturbances, computational complexity and stability.

In this chapter, the cepstral analysis, which was successfully applied to AFC in the previous chapters, will be used to develop mono-channel AEC methods. As discussed in Chapter 4, the cepstral analysis is quite suitable for deconvolution due to the property of transforming a convolution into a linear combination. For a better explanation, consider a convolution between two signals, the desired and contaminant signals, and its output. The traditional deconvolution by means of cepstral analysis assumes that only the convolution output is available and can be done in two different ways. In the first way, it is considered that the cesptra of the convolution inputs do not overlap and thus a filtering operation, which is actually called liftering, is applied to the cepstrum of the convolution output in order to obtain the cepstrum of the desired signal. In the second way, it is considered that the cepstrum of the contaminant signal generates noticeable changes in the cepstrum of the desired signal. Thus, the corresponding samples of the cepstrum of the convolution output are forced to zero in order to remove the effects of the contaminant signal. Thereafter, in both cases, the inverse cepstrum transformation is applied to obtain an estimate of the desired signal in the time-domain. In both cases, therefore, the deconvolution is performed in the cepstral-domain, i.e., by processing directly the cepstrum of the convolution output.

However, in an AEC application, in addition to the convolution output (the echo signal $\mathbf{f}(n) * x(n)$), the contaminant signal (the loudspeaker signal $x(n)$) is also available. Hence, this work will exploit this fact to develop an AEC method, where an adaptive filter $H(q,n)$ estimates the echo path $F(q,n)$ and removes its influence from the system. But, instead of the traditional gradient-based or least-squares-based adaptive algorithms, the adaptive filter $H(q,n)$ will be updated based on cepstral analysis of the system signals.

6.4 Mono-channel AEC Based on Cepstral Analysis

In the mono-channel AEC depicted in Figure 6.2, the microphone signal is defined as

$$y(n) = \mathbf{f}(n) * x(n) + v(n) + r(n). \tag{6.7}$$

Assuming a low-intensity noisy environment such that $r(n) \approx 0$, the microphone signal $y(n)$ defined in (6.7) can be approximated by

$$y(n) \approx \mathbf{f}(n) * x(n) + v(n). \tag{6.8}$$

In order to successfully apply the cepstral analysis to the microphone signal $y(n)$ defined in (6.8), it is necessary to consider that the update of the adaptive filter $H(q, n)$ will not be performed during double-talk, similarly to the traditional gradient-based and least-squares-based adaptive algorithms. Thus, when $v(n) = 0$, (6.8) is simplified to

$$y(n) = \mathbf{f}(n) * x(n). \tag{6.9}$$

From (6.9), the frequency-domain relationship between the loudspeaker signal $x(n)$ and the microphone signal $y(n)$ is given by

$$Y(e^{j\omega}, n) = F(e^{j\omega}, n)X(e^{j\omega}, n), \tag{6.10}$$

which by applying the natural logarithm becomes

$$\ln\left[Y(e^{j\omega}, n)\right] = \ln\left[F(e^{j\omega}, n)\right] + \ln\left[X(e^{j\omega}, n)\right]. \tag{6.11}$$

Applying the inverse Fourier transform in (6.11) as follows

$$\mathcal{F}^{-1}\left\{\ln\left[Y(e^{j\omega}, n)\right]\right\} = \mathcal{F}^{-1}\left\{\ln\left[F(e^{j\omega}, n)\right]\right\} + \mathcal{F}^{-1}\left\{\ln\left[X(e^{j\omega}, n)\right]\right\}, \tag{6.12}$$

the cepstral-domain relationship between the microphone signal $y(n)$ and the loudspeaker signal $x(n)$ is obtained as

$$\mathbf{c_y}(n) = \mathbf{c_x}(n) + \mathbf{c_f}(n). \tag{6.13}$$

As in any filtering operation, the cepstrum $\mathbf{c_y}(n)$ of the microphone signal is the cepstrum $\mathbf{c_x}(n)$ of the loudspeaker signal added to the cepstrum $\mathbf{c_f}(n)$ of the echo path. The cepstra $\mathbf{c_x}(n)$ and $\mathbf{c_y}(n)$ can be simply computed over time from the loudspeaker signal $x(n)$ and microphone signal $y(n)$, respectively, since they are available in the system. Thus, an estimate of the cepstrum of the echo path can be calculated as

$$\hat{\mathbf{c}}_\mathbf{f}(n) = \mathbf{c_y}(n) - \mathbf{c_x}(n). \tag{6.14}$$

From $\hat{\mathbf{c}}_{\mathbf{f}}(n)$, an estimate $\hat{\mathbf{f}}(n)$ of the echo path impulse response can be computed by applying the inverse cepstral transformation. To this end, the cepstrum $\hat{\mathbf{c}}_{\mathbf{f}}(n)$ must contain not only the amplitude information but also the phase information of the spectrum $F(e^{j\omega}, n)$ of the echo path. Therefore, it is necessary to use the complex cepstrum.

In order to compute the complex cepstrum $\mathbf{c}_{\mathbf{y}}(n)$ of the microphone signal, as discussed in [70], it is necessary to unwrap the phase of the spectrum $Y(e^{j\omega}, n)$ of the microphone signal and then remove its linear component. Assuming that $\angle Y_u(e^{j\omega}, n)$ is the unwrapped phase of $Y(e^{j\omega}, n)$, the linear component is removed according to [70]

$$\angle Y'_u(e^{j\omega}, n) = \angle Y_u(e^{j\omega}, n) - \omega\, r_y(n), \tag{6.15}$$

where

$$r_y(n) = \frac{\angle Y_u(e^{j\pi}, n)}{\pi} \tag{6.16}$$

is the lag of the microphone signal $y(n)$. The same procedure is performed to compute the complex cepstrum $\mathbf{c}_{\mathbf{x}}(n)$ of the loudspeaker signal.

Applying the inverse transformation of the complex cepstrum, an estimate of the echo path impulse response can be calculated according to

$$\hat{\mathbf{f}}(n) = \mathcal{F}^{-1}\left\{\exp\left[\mathcal{F}\left\{\hat{\mathbf{c}}_{\mathbf{f}}(n)\right\}\right]\right\}, \tag{6.17}$$

where the linear component must be inserted in the phase of $\mathcal{F}\left\{\hat{\mathbf{c}}_{\mathbf{f}}(n)\right\}$ using

$$r_f(n) = r_y(n) - r_x(n) \tag{6.18}$$

and $\mathcal{F}\{\cdot\}$ denotes the Fourier Transform. Hereupon, the estimate $\hat{\mathbf{f}}(n)$ of the echo path impulse response must be truncated to length L_H samples.

Although the adaptive filter can be updated directly as $\mathbf{h}(n) = \hat{\mathbf{f}}(n)$, in order to increase robustness to short-burst disturbances, the adaptive filter will be updated according to

$$\mathbf{h}(n) = \lambda \mathbf{h}(n-1) + (1-\lambda)\hat{\mathbf{f}}(n), \tag{6.19}$$

where $0 \leq \lambda < 1$ is a factor that controls the trade-off between robustness and tracking rate of the adaptive filter.

In conclusion, the AEC based on cepstral analysis calculates an estimate of $\mathbf{f}(n)$ from $\mathbf{c}_{\mathbf{y}}(n)$ and $\mathbf{c}_{\mathbf{y}}(n)$ to update $H(q, n)$. Depending on the variations of $F(q, n)$ over time, it can be deduced that this computational effort may not be worth it, regarding performance, if the method is applied to each new sample of the microphone signal $y(n)$ and loudspeaker signal $x(n)$. Therefore, the cepstral analysis will be applied every N_{fr} samples, where N_{fr} is a parameter that controls the trade-off between performance (latency and tracking capability) and computational complexity.

The scheme of the proposed AEC based on cepstral analysis is illustrated in Figure 6.3

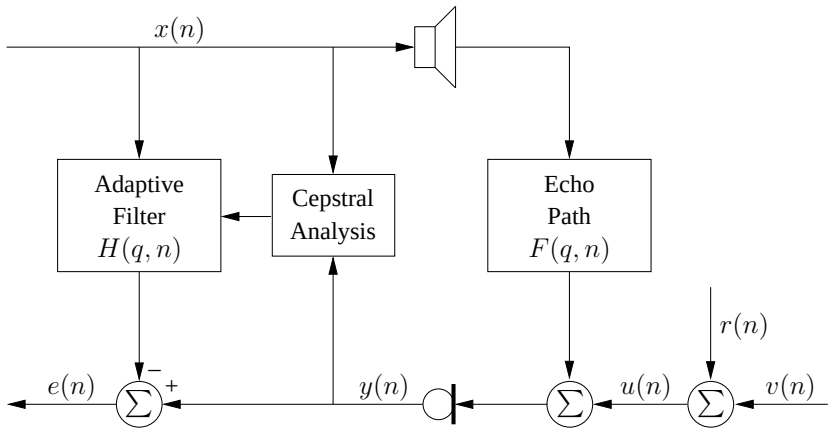


Figure 6.3: AEC based on cepstral analysis.

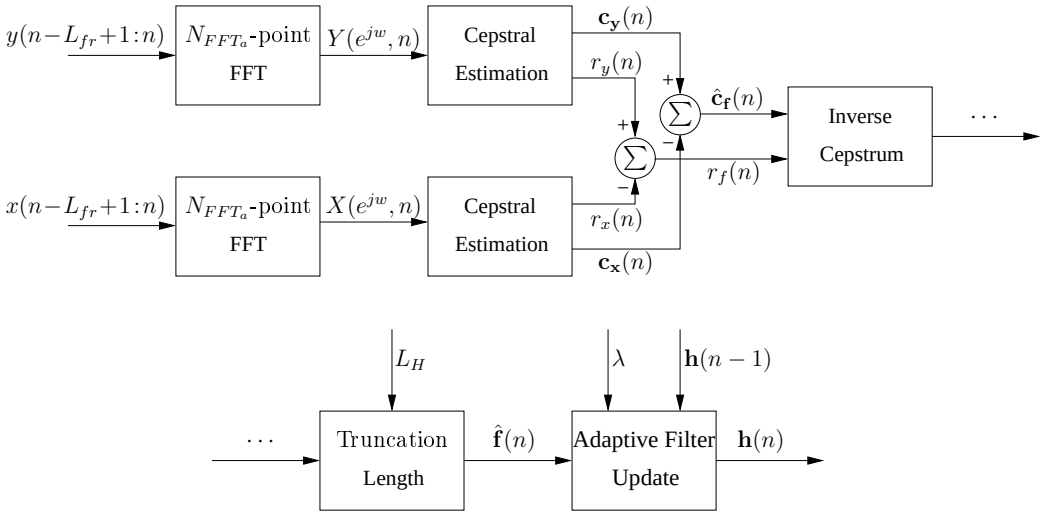


Figure 6.4: Detailed block diagram of the cepstral analysis.

and a detailed block diagram of the cepstral analysis is depicted in Figure 6.4. Every N_{fr} samples, a frame of the microphone signal $y(n)$ and loudspeaker signal $x(n)$ containing their newest L_{fr} samples is selected; the frames have their spectra, $Y(e^{j\omega}, n)$ and $X(e^{j\omega}, n)$, and complex cepstra, $\mathbf{c_y}(n)$ and $\mathbf{c_x}(n)$, calculated through an N_{FFT_a}-point Fast Fourier Transform (FFT); $\hat{\mathbf{c}}_\mathbf{f}(n)$ is calculated from $\mathbf{c_y}(n)$ and $\mathbf{c_x}(n)$; from $\hat{\mathbf{c}}_\mathbf{f}(n)$, $\hat{\mathbf{f}}(n)$ is computed and truncated to length L_H samples; finally, $\mathbf{h}(n)$ is updated.

6.4.1 AEC Based on Cepstral Analysis With No Lag

The first step of the cepstral analysis is to select a frame of loudspeaker signal $x(n)$ and microphone signal $y(n)$. The new AEC based on cepstral analysis with no lag (AEC-CA) method selects, as usually in frequency analysis, frames that contain the newest L_{fr}

samples of the signals as follows

$$\begin{aligned}\mathbf{x}(n) &= [x(n-L_{fr}+1) \ \ \ldots \ \ x(n-1) \ \ x(n)]^T \\ \mathbf{y}(n) &= [y(n-L_{fr}+1) \ \ \ldots \ \ y(n-1) \ \ y(n)]^T .\end{aligned} \tag{6.20}$$

However, the time-domain truncation of a signal causes inevitable oscillations in its frequency response, the so-called spectral leakage, and the only known way to moderate them is to use a smoothing window instead of the rectangular window as in (6.20). Then, the selected frames are multiplied by a smoothing window function $\mathbf{w}$ with length L_{fr} leading to the windowed frames

$$\begin{aligned}\mathbf{x_w}(n) &= \mathbf{x}(n) \circ \mathbf{w} \\ \mathbf{y_w}(n) &= \mathbf{y}(n) \circ \mathbf{w},\end{aligned} \tag{6.21}$$

where $\circ$ denotes the Hadamard or element-wise multiplication.

The advantage of the AEC-CA method is that it does not have any lag estimation, obtaining $\hat{\mathbf{f}}(n)$ at time index n. Its disadvantage is that its frame $\mathbf{y_w}(n)$ of the microphone signal does not accurately contain the echo signal generated by the frame $\mathbf{x_w}(n)$ of the loudspeaker signal. As a consequence, the method is able to estimate the echo path impulse response $\mathbf{f}(n)$ without lag but its estimate $\hat{\mathbf{f}}(n)$ may not be very accurate, as will be discussed in Sections 6.4.2 and 6.4.3.

6.4.2 AEC Based on Cepstral Analysis With No Lag - Improved

Figure 6.5 depicts the discrete convolution of (6.9) using the overlap-and-add method. The AEC-CA method selects the frames $\mathbf{x}(n)$ and $\mathbf{y}(n)$ of the loudspeaker and microphone signals, respectively, according to (6.20). In relation to Figure 6.5, the frame $\mathbf{x}(n)$ of the loudspeaker signal corresponds to the frame $\mathbf{x_2}$(n) and the frame $\mathbf{y}(n)$ of the microphone signal corresponds to the samples of $y(n)$ in the time interval $[n - L_{fr} + 1, n]$.

It can be observed that the frame $\mathbf{y}(n)$ of the microphone signal does not exactly match the convolution result $\mathbf{y}_2(n)$ generated by the frame $\mathbf{x}(n) = \mathbf{x_2}(n)$ of the loudspeaker signal for two reasons. First, $\mathbf{y}(n)$ contains the last L_F samples of $\mathbf{y_1}(n)$, the convolution result generated by the frame $\mathbf{x}_1(n)$ of the loudspeaker signal. Second, $\mathbf{y}(n)$ does not contain the last L_F samples of $\mathbf{y_2}(n)$. These facts probably degrade the estimate $\hat{\mathbf{f}}(n)$ of the echo path impulse response provided by the AEC-CA method.

However, it is possible to solve the first problem in order to obtain a more accurate frame of the microphone signal, approximating it to $\mathbf{y}_2(n)$, and thus improve the estimate $\hat{\mathbf{f}}(n)$ of the echo path impulse response provided by the cepstral analysis. To this end, the inverse of the overlap-and-add method can be performed using the adaptive filter impulse response $\mathbf{h}(n)$ as an estimate of the echo path impulse response $\mathbf{f}(n)$.

In order to remove the last L_F samples of $\mathbf{y_1}(n)$ from the frame $\mathbf{y}(n)$ of the microphone signal defined in (6.20), a new method, called improved AEC-CA (AEC-CAI), is proposed.

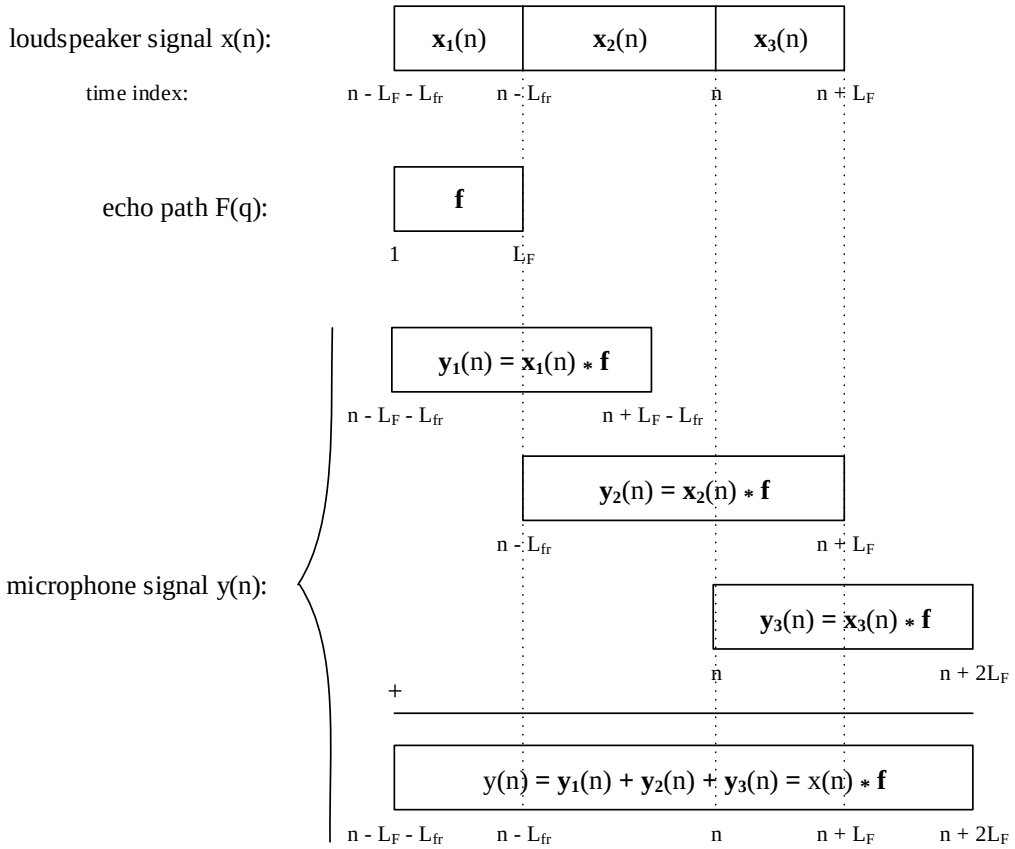


Figure 6.5: Illustration of the discrete convolution using the overlap-and-add method.

The AEC-CAI method estimates the convolution result $\mathbf{y_1}(n)$ as

$$\hat{\mathbf{y}}_1(n) = \mathbf{x}_1(n) * \mathbf{h}(n), \tag{6.22}$$

where

$$\mathbf{x_1}(n) = [x(n - L_{fr} - L_F) \ \ \ldots \ \ x(n - L_{fr} - 2) \ \ x(n - L_{fr} - 1)]^T . \tag{6.23}$$

Then, the method creates the auxiliary signal

$$\mathbf{y}'_1(n) = \begin{bmatrix} \underline{\hat{\mathbf{y}}_1(n)}_{L_F} \\ \mathbf{0}_{(L_{fr}-L_F)\times 1} \end{bmatrix}, \tag{6.24}$$

where $\mathbf{0}_{N\times 1}$ is a null matrix with dimension $N \times 1$ and $\underline{\mathbf{a}}_N$ is a vector formed by the N last samples of the vector $\mathbf{a}$.

From (6.22), it is noteworthy that the estimate $\hat{\mathbf{y}}_1(n)$ approaches $\mathbf{y}_1(n)$ as the match between the impulse responses of the adaptive filter and echo path improves. If $\mathbf{h}(n) = \mathbf{f}(n)$, then $\hat{\mathbf{y}}_1(n) = \mathbf{y}_1(n)$. However, in practice, the length L_F of the echo path is unknown

and hence the length L_H of the adaptive filter is actually used in (6.22) and (6.24).

Thus, the AEC-CAI method computes the frame of the microphone signal according to

$$\mathbf{y}'(n) = \mathbf{y}(n) - \mathbf{y}'_1(n), \tag{6.25}$$

where $\mathbf{y}(n)$ is defined in (6.20).

Finally, the AEC-CAI method defines the windowed frames of the loudspeaker and microphone signals to which the cepstral analyis will be applied, respectively, as

$$\begin{aligned} \mathbf{x_w}(n) &= \mathbf{x}(n) \circ \mathbf{w} \\ \mathbf{y_w}(n) &= \mathbf{y}'(n) \circ \mathbf{w}. \end{aligned} \tag{6.26}$$

The AEC-CAI and AEC-CA methods have the same frame $\mathbf{x}(n)$ of the loudspeaker signal defined in (6.20). The advantage of the AEC-CAI method is that its frame of the microphone signal, $\mathbf{y}'(n)$, is closer to the convolution result between $\mathbf{x}(n)$ and the impulse response $\mathbf{f}(n)$ of the echo path. On the other hand, in order to achieve such improvement in the frame of the microphone signal, the AEC-CAI has a higher computational complexity.

6.4.3 AEC Based on Cepstral Analysis With Lag

As discussed in Section 6.4.2, the problem of the frame $\mathbf{y}(n)$ of the microphone signal selected by the AEC-CA method, defined in (6.20), is twofold. First, it contains the last L_F samples of $\mathbf{y_1}(n)$, the convolution result generated by the frame $\mathbf{x}_1(n)$ of the loudspeaker signal. Second, it does not contain the last L_F samples of $\mathbf{y_2}(n)$, the convolution result generated by the selected frame $\mathbf{x}(n)$ of the loudspeaker signal. These facts degrade the estimate $\hat{\mathbf{f}}(n)$ of the echo path impulse response provided by the cepstral analysis.

The first problem is overcome in the AEC-CAI method by performing the inverse of the overlap-and-add method using $\mathbf{h}(n)$ as an estimate of $\mathbf{f}(n)$. However, the second problem still occurs. A first idea to overcome the second problem would be to increase the length of the frame of the microphone signal so that it corresponds to the samples of $y(n)$ in the time interval $[n - L_{fr} + 1, n + L_F]$. But it can be observed from Figure 6.5 that the resulting frame would also contain the first L_F samples of $\mathbf{y_3}(n)$, the convolution result generated by the frame $\mathbf{x}_3(n)$ of the loudspeaker signal.

In order to include the last L_F samples of $\mathbf{y_2}(n)$ in the frame of the microphone signal, a new method, called the AEC based on cepstral analysis with lag (AEC-CAL), is proposed. The AEC-CAL method extends the idea of performing the inverse of the overlap-and-add method, applied in the AEC-CAI method, to the frame $\mathbf{x}_3(n)$ of the loudspeaker signal. The method computes an estimate of the convolution output $\mathbf{y_3}(n)$ according to

$$\hat{\mathbf{y}}_3(n) = \mathbf{x}_3(n) * \mathbf{h}(n), \tag{6.27}$$

where

$$\mathbf{x_3}(n) = [x(n+1) \quad \dots \quad x(n+L_F-1) \quad x(n+L_F)]^T, \tag{6.28}$$

Then, the method creates the auxiliary signals

$$\mathbf{y}_1''(n) = \begin{bmatrix} \overline{\hat{\mathbf{y}}_1(n)}_{L_F} \\ \mathbf{0}_{(L_{fr}-1)\times 1} \end{bmatrix} \tag{6.29}$$

and

$$\mathbf{y}_3''(n) = \begin{bmatrix} \mathbf{0}_{(L_{fr}-1)\times 1} \\ \overline{\hat{\mathbf{y}}_3(n)}_{L_F} \end{bmatrix}, \tag{6.30}$$

where $\hat{\mathbf{y}}_1(n)$ is defined in (6.22) and $\overline{\mathbf{a}}_N$ denotes the N first samples of the vector $\mathbf{a}$.

From (6.27), it can be concluded that $\hat{\mathbf{y}}_3(n)$ approaches $\mathbf{y}_3(n)$ as the match between the impulse responses of the adaptive filter and echo path improves. If $\mathbf{h}(n) = \mathbf{f}(n)$, then $\hat{\mathbf{y}}_3(n) = \mathbf{y}_3(n)$. However, in practice, the length L_F of the echo path is unknown and hence the length L_H of the adaptive filter is actually used in (6.27), (6.29) and (6.30).

Thus, the AEC-CAL method calculates the frame of the microphone signal as follows

$$\mathbf{y}''(n) = \mathbf{y}(n) - \mathbf{y}_1''(n) - \mathbf{y}_3''(n), \tag{6.31}$$

where

$$\mathbf{y}(n) = [y(n-L_{fr}+2) \quad \dots \quad y(n+L_F-1) \quad y(n+L_F)]^T. \tag{6.32}$$

Finally, the AEC-CAL method defines the windowed frames of the loudspeaker and microphone signals to which the cepstral analyis will be applied, respectively, as

$$\begin{aligned} \mathbf{x_w}(n) &= \mathbf{x}(n) \circ \mathbf{w} \\ \mathbf{y_w}(n) &= \mathbf{y}''(n) \circ \mathbf{w}. \end{aligned} \tag{6.33}$$

It is noteworthy that, for real-time implementation, the proposed AEC-CAL method involves a lag of L_F samples for the update of the adaptive filter $H(q,n)$ because the frame $\mathbf{x_3}(n)$ of the loudspeaker signal, defined in (6.28), is only available at the time $n+L_F$. The lag is efficiently implemented as a delay line for the windowed frames $\mathbf{x_w}(n)$ and $\mathbf{y_w}(n)$ of the loudspeaker and microphone signals before computing their complex cepstra. This lag may be a problem depending on the length and variations of $F(q,n)$ over time. However, the value of the lag is equal to the number of samples from the last L_F samples of $\mathbf{y_2}(n)$ that the method intends to include in the frame of the microphone signal. Hence, a lower lag can be achieved by including a smaller number of the last samples of $\mathbf{y_2}(n)$. The drawback will be a less accurate frame of the microphone signal and thereby a less accurate estimate of the echo path impulse response provided by the cepstral analysis. Therefore, the number of samples from the last L_F samples of $\mathbf{y_2}(n)$ that the AEC-CAL method will include in the frame of the microphone signal is a trade-off between accuracy

in estimating the echo path and latency.

6.4.4 Simulation Configurations

With the aim to assess the performance of the proposed AEC-CA, AEC-CAI and AFC-CAL methods, an experiment was carried out in a simulated environment to measure their ability to estimate the echo path impulse response and attenuate the acoustic echo signal. To this purpose, the following configuration was used.

6.4.4.1 Simulated Environment

The impulse response $\mathbf{f}(n)$ of the acoustic echo path was a measured room impulse response, from [60], and thus $\mathbf{f}(n) = \mathbf{f}$. The impulse response was downsampled to $f_s = 16$ kHz and then truncated to length $L_F = 4000$ samples, and is illustrated in Figure 3.3.

6.4.4.2 Misalignment

The performance of the proposed AEC-CA, AEC-CAI and AFC-CAL methods methods were also evaluated through the normalized misalignment (MIS) metric. The MIS(n) measures the distance between the impulse responses of the adaptive filter and the echo path according to (3.43). A detailed description can be found in Section 3.6.3.

6.4.4.3 Echo Return Loss Enhancement

A standardized metric for echo cancellation is the Echo Return Loss Enhancement (ERLE) [90]. The ERLE measures the attenuation of the echo signal provided by the echo canceller and is the inverse of the Mean Square Error (MSE) often used in the literature for echo cancellation. In this work, the performance of the adaptive filter as an echo canceller was measured by the normalized ERLE defined as

$$\text{ERLE}(n) = \frac{\text{LPF}\{[y(n) - r(n)]^2\}}{\text{LPF}\{[e(n) - r(n)]^2\}}, \tag{6.34}$$

where LPF$\{\cdot\}$ denotes a low-pass filter with a single pole at 0.999. In a simulated environment, the normalized ERLE(n) provides a more accurate evaluation by removing the ambient noise signal $r(n)$ from the measurement. Moreover, the use of the low-pass filter is a common practice in AEC to obtain a smooth curve ERLE(n) by removing the high frequency components without significantly affecting the convergence behavior.

Its optimum value is ERLE(n) $= \infty$ and, as the MIS(n), is achieved when $\mathbf{h}(n) = \mathbf{f}(n)$. In general, ERLE($n$) $\to \infty$ as $\mathbf{h}(n) \to \mathbf{f}(n)$. The MIS($n$) and ERLE($n$) metrics are correlated which means that an improvement in one of them probably results in an improvement in the other. However, this may not occur because MIS(n) measures $\|\mathbf{f}(n) -$

$\mathbf{h}(n)\|$ while ERLE(n) uses the value of the error signal $e(n)$ that depends on the waveform of $\mathbf{f}(n) - \mathbf{h}(n)$, as defined in (6.6), and not only on its norm. Therefore, a solution $\mathbf{h}_1(n)$ can achieve a better MIS(n) and a worst ERLE(n), or otherwise, than a solution $\mathbf{h}_2(n)$.

6.4.4.4 Signal database

The signal database was formed by the same 10 speech signals used in Chapters 3, 4 and 5. A detailed description can be found in Section 3.6.6.

6.4.5 Simulation Results

This section presents and discusses the performance of the proposed AEC-CA, AEC-CAI and AEC-CAL methods using the configuration of the teleconference system, the evaluation metrics and the signals described in Section 6.4.4. The proposed methods started only after 125 ms of simulation to avoid initial inaccurate estimates of the cepstra of the microphone and loudspeaker signals, $N_{fr} = 1000$, $N_{FFT_a} = 2^{15}$ and $N_{FFT_e} = 2^{17}$.

The parameters λ and L_H of the adaptive filter were optimized for each signal. From pre-defined ranges, the values of λ and L_H were chosen empirically in order to optimize the curves MIS(n) and ERLE(n) with regard to mean value within the simulation time $T = 20$ s. The optimal curves for the kth signal were denoted as MIS$_k(n)$ and ERLE$_k(n)$. Then, the mean curves MIS(n) and ERLE(n) were obtained by averaging the curves of each signal according to

$$\begin{aligned} \text{MIS}(n) &= \frac{1}{10}\sum_{k=1}^{10}\text{MIS}_k(n) \\ \text{ERLE}(n) &= \frac{1}{10}\sum_{k=1}^{10}\text{ERLE}_k(n). \end{aligned} \tag{6.35}$$

And their respective mean values were defined as

$$\begin{aligned} \overline{\text{MIS}} &= \frac{1}{N_T}\sum_{n=1}^{N_T}\text{MIS}(n) \\ \overline{\text{ERLE}} &= \frac{1}{N_T}\sum_{n=1}^{N_T}\text{ERLE}(n), \end{aligned} \tag{6.36}$$

where N_T is the number of samples relating to the simulation time. In addition, the asymptotic values of MIS(n) and ERLE(n) were defined as $\overrightarrow{\text{MIS}}$ and $\overrightarrow{\text{ERLE}}$, respectively, and were estimated only by graphically inspecting of the curves.

The evaluation was done in several ambient noise conditions because, unlike the AFC methods based on cesptral analysis proposed in Chapter 4, the AEC-CA, AEC-CAI and AEC-CAL methods proved to be very sensitive to the level of the ambient noise. Table 6.1

summarizes the results obtained by the AEC methods based on cepstral analysis for different values of the frame length L_{fr} and echo-to-ambient-noise ratio (ENR) . A detailed discussion about the results will be held in the following sections.

Table 6.1: Summary of the results obtained by the proposed AEC methods based on cepstral analysis.

	L_{fr}	ENR = 30 dB		ENR = 40 dB		ENR = 50 dB		ENR = ∞	
		$\overline{\text{MIS}}$	$\overline{\text{ERLE}}$	$\overline{\text{MIS}}$	$\overline{\text{ERLE}}$	$\overline{\text{MIS}}$	$\overline{\text{ERLE}}$	$\overline{\text{MIS}}$	$\overline{\text{ERLE}}$
AEC-CA	8000	-12.94	22.04	-17.38	23.50	-22.19	23.83	-24.00	23.89
	12000	-14.41	24.96	-19.58	27.02	-24.47	27.50	-26.91	27.59
	16000	-15.35	26.69	-20.62	29.28	-25.69	29.94	-28.26	30.06
	32000	-17.14	29.67	-22.84	33.12	-28.44	34.09	-32.08	34.32
	80000	-19.10	32.37	-25.08	37.25	-32.54	39.04	-39.08	39.43
AEC-CAI	8000	-13.59	22.03	-17.48	23.99	-21.93	24.38	-23.87	24.52
	12000	-14.91	25.05	-19.61	27.44	-24.73	28.02	-27.21	28.13
	16000	-15.57	26.88	-20.58	29.70	-25.89	30.40	-28.67	30.57
	32000	-17.64	29.66	-22.72	33.18	-28.58	34.30	-32.49	34.53
	80000	-18.68	32.40	-25.08	37.30	-32.65	39.18	-39.62	39.58
AEC-CAL	8000	-14.15	26.90	-19.84	33.27	-25.09	40.45	-30.60	71.13
	12000	-15.37	29.62	-21.33	37.39	-27.20	44.69	-35.11	97.98
	16000	-16.04	31.07	-21.84	39.29	-27.83	47.68	-35.93	114.67
	32000	-17.65	33.93	-24.40	42.85	-31.03	51.58	-36.04	123.40
	80000	-18.59	36.07	-25.95	45.14	-31.76	53.81	-36.59	127.61

6.4.5.1 Influence of Parameters

The results showed that the frame length L_{fr} and the level of the ambient noise have a great influence on the performance of the proposed AEC method based on cepstral analysis. Figures 6.6, 6.7 and 6.8 show the values of $\overline{\text{MIS}}$ and $\overline{\text{ERLE}}$ obtained by the AEC-CA, AEC-CAI and AEC-CAL methods, respectively, as a function of L_{fr} and ENR.

It can be observed that the performance of all methods improves as ENR increases. The basis of the cepstral analysis of the AEC system presented in Section 6.4, which led to development of all proposed methods, was the definition of the microphone signal $y(n)$ according to (6.9). This definition is actually an approximation of (6.7) considering a low-intensity noisy environment such that $r(n) \approx 0$ and the absence of the near-end speaker signal $v(n)$. Therefore, the more the ENR increases, the more (6.9) approaches (6.7). Consequently, the cepstral analysis becomes more accurate and thus the performance of all the proposed AEC methods is improved.

It can also noticed that the performance of all methods improves as L_{fr} increases. This is explained by the fact that increasing L_{fr} increases the number of samples that are provided for the cepstral analysis. This results in a more accurate estimate of the cepstrum $c_{\mathbf{f}}(n)$ of the echo path and, consequently, of its impulse response $\mathbf{f}(n)$.

In the AEC-CA method, as discussed in Section 6.4.2, there are 2 sample blocks with length L_F, one included and the other excluded from the frame of the microphone signal, that degrade the estimation of the impulse response $\mathbf{f}(n)$ of the echo path. For fixed L_F, increasing L_{fr} also reduces the influence of these two sample blocks until they become irrelevant. As a consequence, the estimate of $\mathbf{f}(n)$ provided by the AEC-CA is improved. Similar behavior occurs with the proposed AFC methods based on cepstral analysis, AFC-CE and AFC-CM, as discussed in Section 4.4.3.2. The AEC-CAI method has the same sample block with length L_F excluded from the frame of the microphone signal of the AEC-CA method. Thus, for fixed L_F, increasing L_{fr} similarly improves the estimate of $\mathbf{f}(n)$ provided by the AEC-CAI.

However, the conclusion that increasing L_{fr} may improve the estimation of the echo path impulse response $\mathbf{f}(n)$ can be ensured if $\mathbf{f}(n)$ is time-invariant throughout the frame length L_{fr}. If it is time-varying, the cepstral analysis will estimate an average of $\mathbf{f}(n)$ over the frame length L_{fr}. Then, in this case, increasing L_{fr} may give a lower weight to the current values of the impulse response and thus worsen its estimate. Therefore, for time-varying echo path impulse response, the frame length L_{fr} controls the trade-off between the amount of useful samples provided for the cepstral analysis and the weight given by the cepstral analysis to the current impulse response.

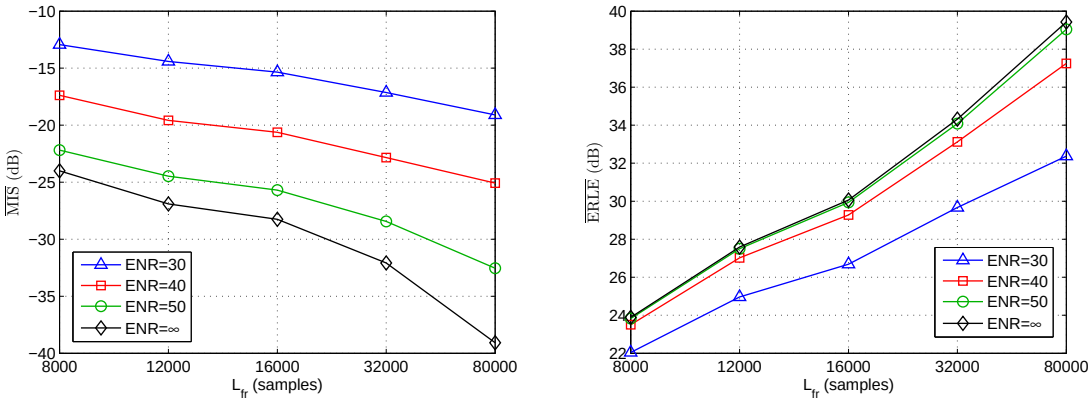


Figure 6.6: Influence of L_{fr} and ENR in the performance of AEC-CA: (a) $\overline{\text{MIS}}$; (b) $\overline{\text{ERLE}}$.

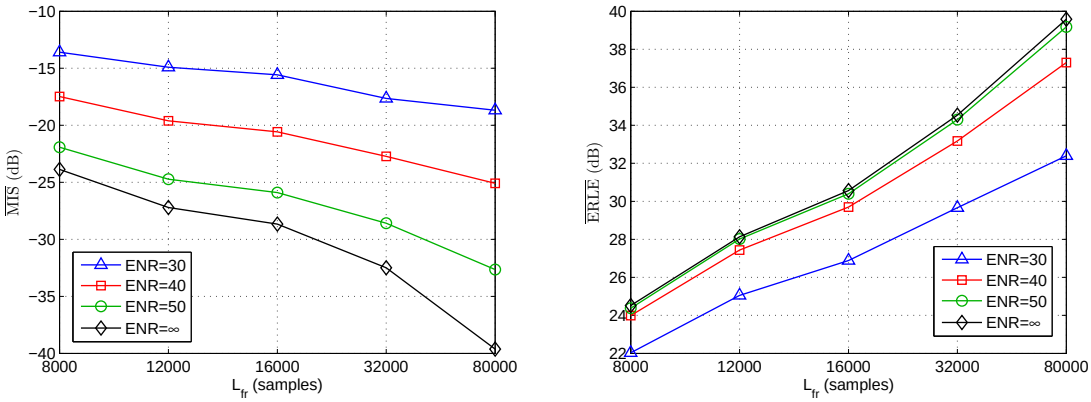


Figure 6.7: Influence of L_{fr} and ENR in the performance of AEC-CAI: (a) $\overline{\text{MIS}}$; (b) $\overline{\text{ERLE}}$.

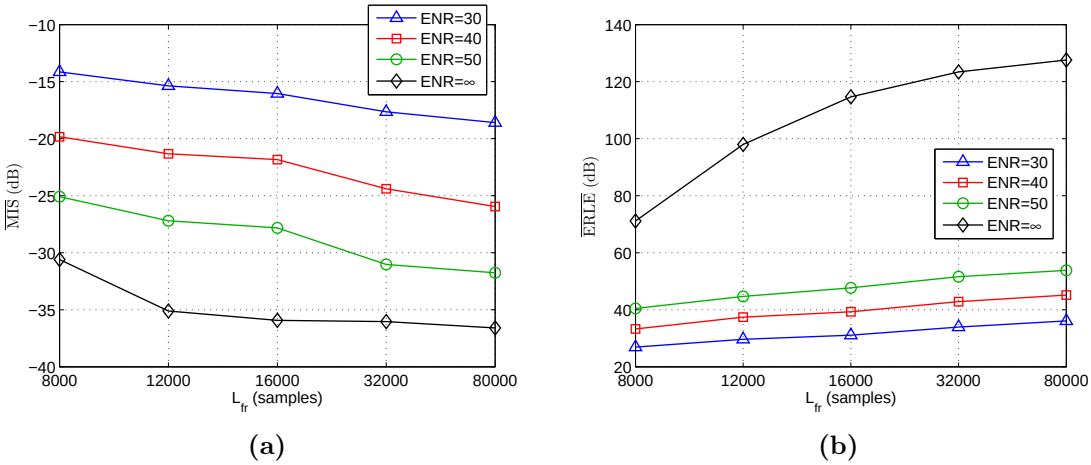


Figure 6.8: Influence of L_{fr} and ENR in the performance of AEC-CAL: (a) $\overline{\text{MIS}}$; (b) $\overline{\text{ERLE}}$.

6.4.5.2 Performance Comparison

This section will analyze and discuss the performance of the proposed AEC methods based on cepstral analysis. Since their adaptive filter parameters were chosen in order to optimize $\overline{\text{MIS}}$ and $\overline{\text{ERLE}}$, Table 6.1 will be the basis of this analysis. But to enrich the discussion, Figures 6.9 and 6.10 show the curves MIS(n) and ERLE(n) obtained by each method when ENR = 30 and 40 dB, respectively, and for $L_{fr} = 8000, 16000, 80000$ samples.

Among all assessed values, these values of ENR were chosen for illustration of the curves MIS(n) and ERLE(n) because they are commonly used in AEC as well as in line/network echo cancellation, an adaptive filter application similar to AEC where the echo signal is electrically generated. And these values of L_{fr} were chosen because they are the extreme and mean of the used values.

From Table 6.1, it can be concluded that the AEC-CAI method generally outperforms the AEC-CA with regard to both MIS(n) and ERLE(n). However, its advantage is very small such that it never exceeded 0.7 dB in both metrics. This small difference in performance can be also noticed by observing the difference between the MIS(n) and ERLE(n) curves obtained by the AEC-CAI and AEC-CA methods in Figures 6.9 and 6.10. With exception of Figure 6.9a, the curves are practically superimposed. As discussed in Section 6.4.2, the AEC-CAI method removes the last L_F samples of the convolution result between the previous frame of the loudspeaker signal and the echo path from the selected frame $\mathbf{y}(n)$ of the microphone signal. In fact, these L_F samples act as noise to the estimation of the convolution result between the current frame $\mathbf{x}(n)$ of the loudspeaker signal and the echo path from the microphone signal $y(n)$. But, these samples do not have high absolute values because they correspond to the convolution tail and thus the disturbance caused by them is actually small. Therefore, the AEC-CAI method is only capable of improving the AEC-CA method by a small amount.

In addition, from Table 6.1, it can be concluded that the AEC-CAL generally outperforms the AEC-CA and AEC-CAI with regard to both MIS(n) and ERLE(n). However, the advantage of the AEC-CAL is more significant in ERLE(n). This conclusion can be also observed from Figures 6.9 and 6.10 where the superior performance of AEC-CAL is more easily noticeable in ERLE(n) than in MIS(n). Moreover, the advantage of AEC-CAL tends to increase as ENR increases. This fact can be also inferred by observing that the distance between the values of $\overrightarrow{\text{ERLE}}$, the convergent value of ERLE(n), obtained by the AEC-CAL and the other methods increases from Figure 6.9 (with ENR = 30) to Figure 6.10 (with ENR = 40). In the ideal condition when ENR = ∞, the AEC-CAL achieves values of $\overline{\text{ERLE}}$ larger than twice those obtained by the AEC-CA and AEC-CAI.

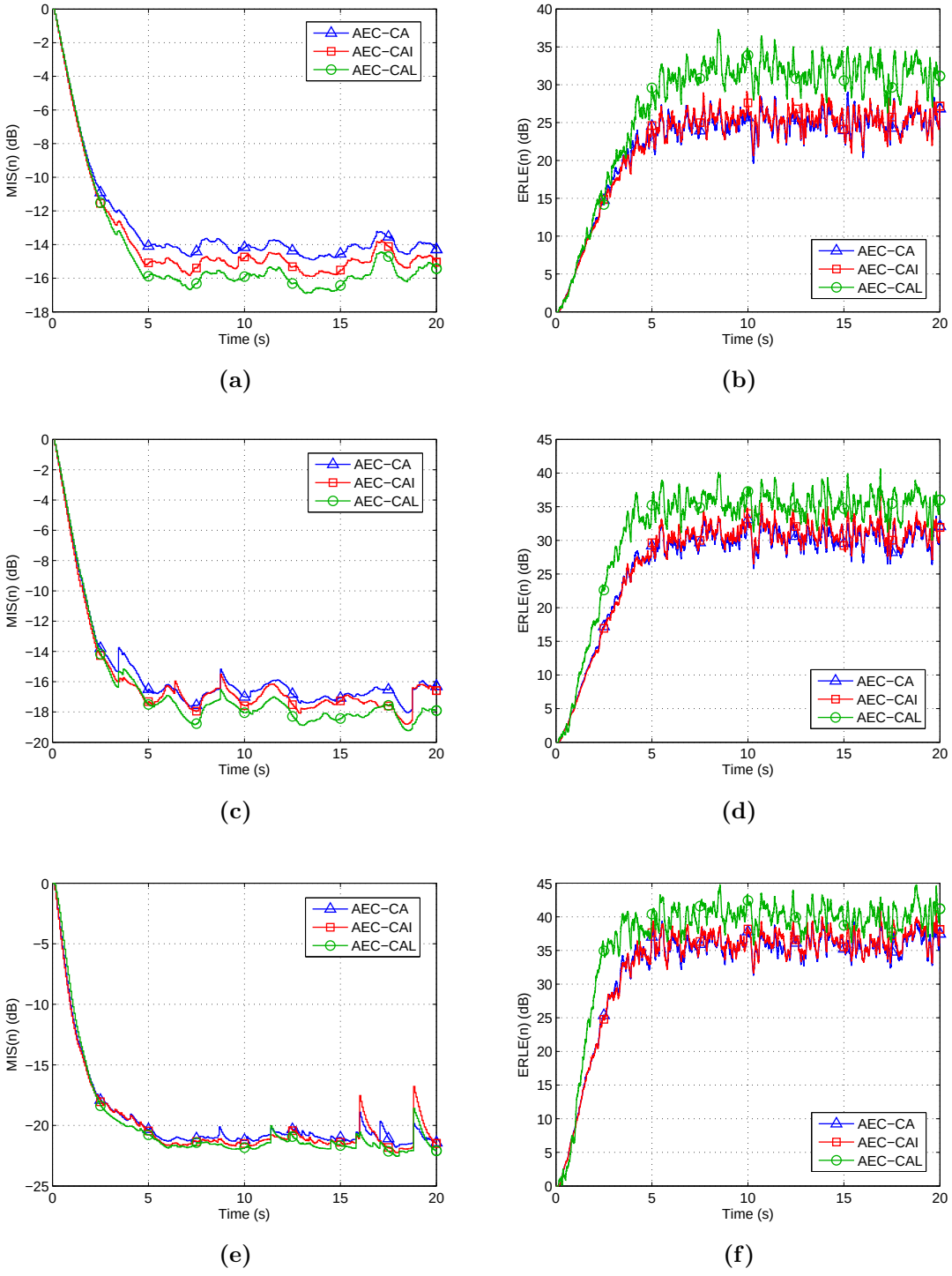


Figure 6.9: Performance comparison between the AEC-CA, AEC-CAI and AEC-CAL methods for ENR = 30 dB: (a),(c),(e) MIS(n); (b),(d),(f) ERLE(n); (a),(b) $L_{fr} = 8000$; (c),(d) $L_{fr} = 16000$; (e),(f) $L_{fr} = 80000$.

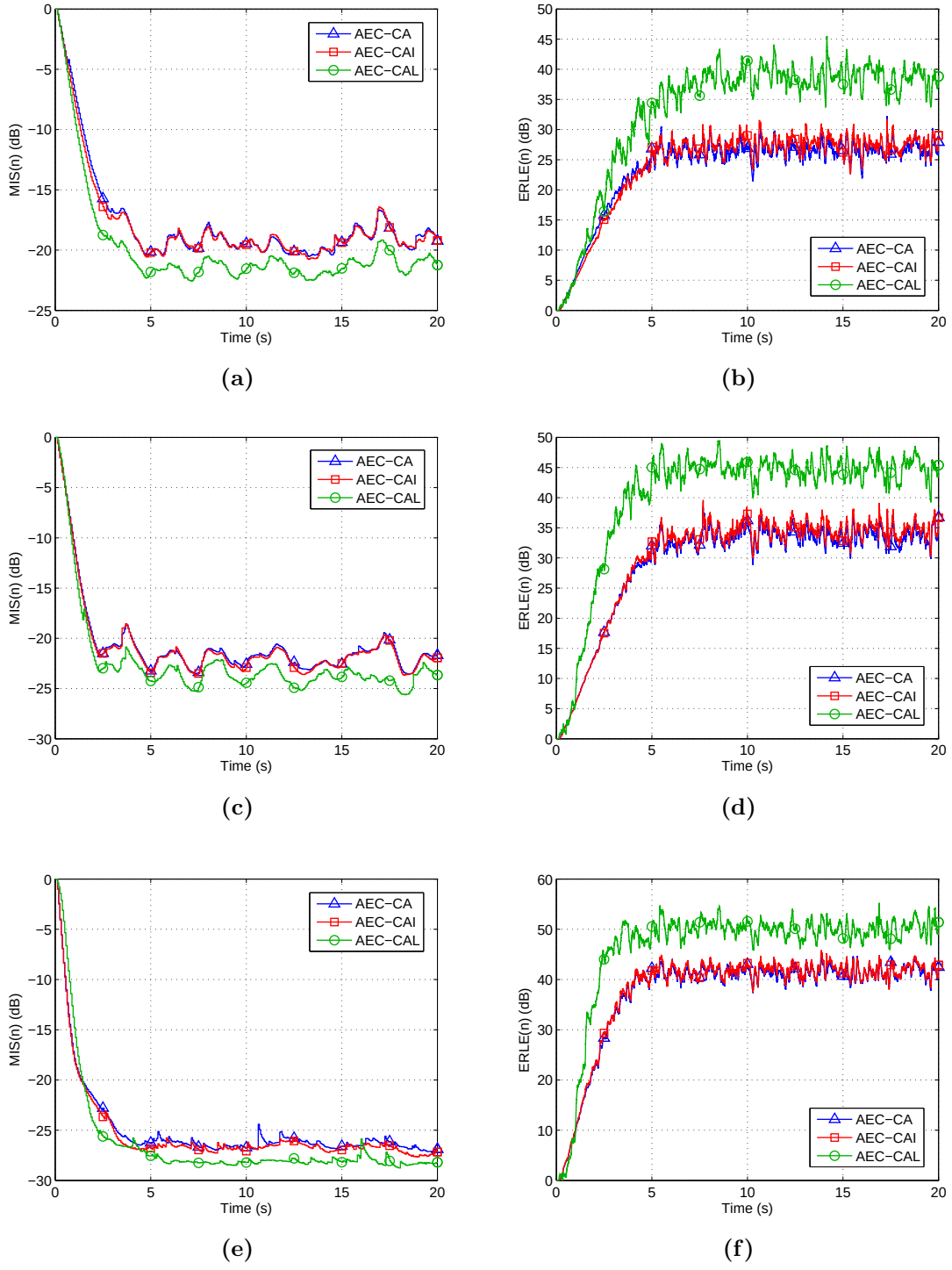


Figure 6.10: Performance comparison between the AEC-CA, AEC-CAI and AEC-CAL methods for ENR = 40 dB: (a),(c),(e) MIS(n); (b),(d),(f) ERLE(n); (a),(b) $L_{fr} = 8000$; (c),(d) $L_{fr} = 16000$; (e),(f) $L_{fr} = 80000$.

Table 6.2: Summary of the results obtained by the NLMS and BNDR-LMS.

	ENR = 30 dB		ENR = 40 dB		ENR = 50 dB		ENR = ∞	
	$\overline{\text{MIS}}$	$\overline{\text{ERLE}}$	$\overline{\text{MIS}}$	$\overline{\text{ERLE}}$	$\overline{\text{MIS}}$	$\overline{\text{ERLE}}$	$\overline{\text{MIS}}$	$\overline{\text{ERLE}}$
NLMS	-14.20	29.19	-15.63	33.26	-16.01	35.16	-16.03	35.65
BNDR-LMS	-17.58	32.74	-21.3	38.87	-23.22	44.61	-24.24	59.87

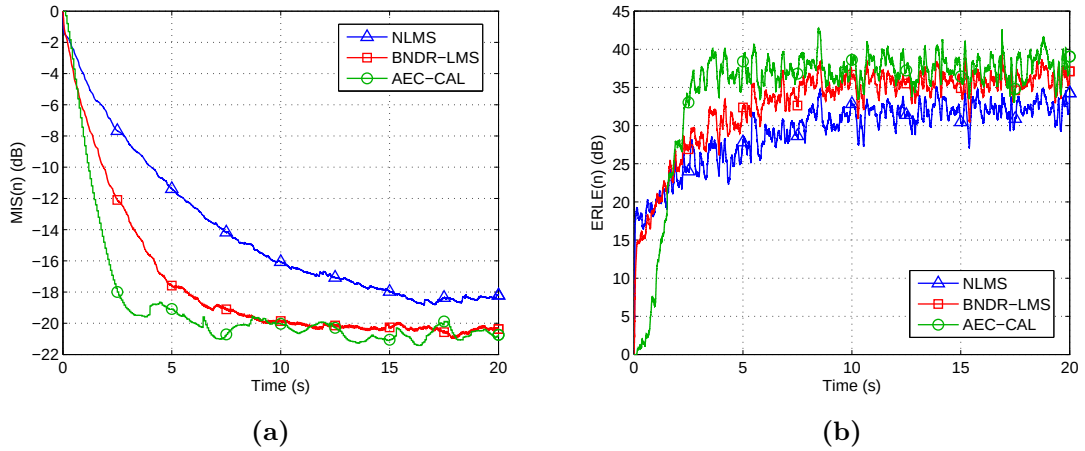
Afterwards, the Normalized Least Mean Square (NLMS) and the Binormalized Data-Reusing LMS (BNDR-LMS) algorithms were used to compare the performance of the proposed AEC methods. Both adaptive filtering algorithms are based on the Wiener theory. The NLMS is the most widely used algorithm in practical applications [12] and can be interpreted as the Affine Projection algorithm (APA) with no data reuse [72, 91, 92]. The BNDR-LMS can be interpreted as a special case of the APA with a single data reuse where the matrix inversion has closed form solution [72, 91, 92].

Their adaptive filter parameters are stepsize μ, normalization factor δ and L_H. And they were chosen empirically in order to optimize, for each signal, the curves MIS(n) and ERLE(n) with regard to mean values according to the same procedure used for the proposed AEC methods and described in Section 6.4.5. Table 6.2 summarizes the results obtained by the NLMS and BNDR-LMS algorithms for different values of ENR. As expected, the BNDR-LMS outperformed the NLMS.

For each one of the ENR values, there is at least one configuration of the proposed AEC methods based on cepstral analisys that outperformed the NLMS algorithm regarding both MIS(n) and ERLE(n). With respect to MIS(n), the maximum value of L_{fr} required for the AEC-CA, AEC-CAI and AEC-CAL to outperform the NLMS were 12000. With respect to ERLE(n), the maximum values of L_{fr} required for the AEC-CA, AEC-CAI and AEC-CAL to outperform the NLMS were 80000, 80000 and 12000, respectively.

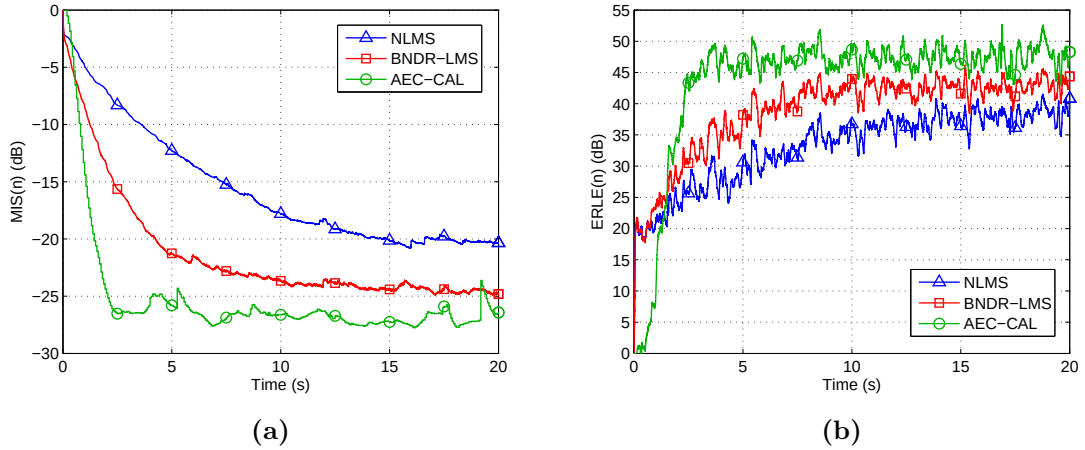
For each one of the ENR values, only the AEC-CAL method was able to outperform the BNDR-LMS with regard to MIS(n) and ERLE(n). The AEC-CA and AEC-CAI methods were only capable of outperforming the BNDR-LMS regarding MIS(n). With respect to MIS(n), the minimum values of L_{fr} required for the AEC-CA, AEC-CAI and AEC-CAL to outperform the BNDR-LMS were 80000, 32000 and 32000, respectively. With regard to ERLE(n), the maximum value of L_{fr} required for the AEC-CAL to outperform the BNDR-LMS was 32000 when ENR = 30 dB.

In order to enrich the discussion, Figures 6.11 and 6.12 compare the performance of the NLMS, BNDR-LMS and AEC-CAL with $L_{fr} = 32000$ when ERN = 30 and 40 dB, respectively. The AEC-CA and AEC-CAI methods are not included in this comparison because AEC-CAL outperformed them and many curves in the same figure would complicate the interpretation. The value $L_{fr} = 32000$ was used because it is the maximum value required for the AEC-CAL to outperform the BNDR-LMS with regard to mean values of MIS(n) and ERLE(n).



(a) **(b)**

Figure 6.11: Performance comparison between NLMS, BNDR-LMS and AEC-CAL for ENR = 30 dB and $L_{fr} = 32000$: (a) MIS(n); (b) ERLE(n).



(a) **(b)**

Figure 6.12: Performance comparison between NLMS, BNDR-LMS and AEC-CAL for ENR = 40 dB and $L_{fr} = 32000$: (a) MIS(n); (b) ERLE(n).

In general, compared with the NLMS and BNDR-LMS algorithms, the proposed AEC methods based on cepstral analysis proved to be more competitive regarding MIS(n) than ERLE(n). This can be explained by the fact that the proposed AEC methods identify directly the impulse response $\mathbf{f}(n)$ of the echo path through the cepstral analysis. On the other hand, the NLMS and BNDR-LMS, as adaptive filtering algorithms based on the Wiener theory, identify indirectly $\mathbf{f}(n)$ by minimizing the instantaneous squared error signal, $e^2(n)$. Therefore, the proposed AEC methods focus on identifying $\mathbf{f}(n)$, whose accuracy is measured by the MIS(n) metric, while the NLMS and BNDR-LMS focus on minimizing $e^2(n)$, whose value is measured by the ERLE(n) metric.

These consequences can be observed in Figures 6.11 and 6.12. The AEC-CAL outperformed the NLMS and BNDR-LMS algorithms with regard to MIS(n) during practically

all the simulation time. However, this advantage in MIS(n) is not reflected in the first seconds of the ERLE(n) although the AEC-CAL method achieved a higher $\overrightarrow{\text{ERLE}}$, convergent mean value. Even when ENR = 40 dB, situation in which the proposed AEC-CAL presented an evident superior performance regarding MIS(n), its advantage of around 10 dB in MIS(n) when $t = 1$ s results in a surprising disadvantage of around 12 dB in ERLE(n). This poor performance of the proposed AEC methods in the first seconds of ERLE(n) greatly reduces its mean value over time, $\overline{\text{ERLE}}$, on which is based the optimization of the adaptive filter parameters (λ and L_H) and the conclusions about performance. It is worth remembering that the proposed AEC methods started only after 125 ms of simulation which obviously implies a worse performance in the very first moments. On the other hand, even when there is little information on the system signals and the adaptive filter is at the beginning of the learning process, the NLMS and BNDR-LMS aim to minimize $e^2(n)$ and thereby achieve a significant attenuation of the echo signal.

In conclusion, the proposed AEC methods, which directly identify the impulse response $\mathbf{f}(n)$ of the echo path through cepstral analysis, proved to be able to outperform the NLMS and BNDR-LMS algorithms with regard to the $\overline{\text{MIS}}$ and $\overrightarrow{\text{ERLE}}$. On the other hand, the NLMS and BNDR-LMS algorithms, which indirectly identify $\mathbf{f}(n)$ by minimizing $e^2(n)$, presented a better performance in the first seconds of ERLE(n). Although the performance in the first seconds (≈ 2 s) may not be so relevant, it would be interesting to avoid this drawback of the proposed AEC methods. Therefore, in order to combine the strength of both methodologies, a hybrid approach emerged.

6.5 Hybrid AEC Based on Cepstral Analysis

Aiming to avoid the worst performance of the proposed AEC methods in the first seconds of ERLE(n), hybrid methods will be proposed to combine the strength of the methodologies of the traditional adaptive filtering algorithms and cepstral analysis. With this, it is expected that the proposed hybrid methods will not perform worse than each method individually with regard to both MIS(n) and ERLE(n). The hybrid methods will combine the AEC-CAI or AEC-CAL methods with the NLMS or BNDR algorithms.

By choice, the AEC-CAI and AEC-CAL methods are applied every $N_{fr} = 1000$ samples and start only after 125 ms to avoid inaccurate initial estimates. In the other time instants, the adaptive filter $H(q, n)$ is not updated. Instead of stopping the update of $H(q, n)$ in these moments, the hybrid methods will apply the NLMS or BNDR-LMS algorithms. Therefore, the adaptive filter $H(q, n)$ will be updated by the NLMS or BNDR-LMS algorithms most of the time. In fact, the AEC-CAI or AEC-CAL methods will be used only to provide an instantaneous estimate of the impulse response $\mathbf{f}(n)$ of the echo path and thereby to accelerate or straighten the learning process of the NLMS and BNDR-LMS algorithms.

Although there are four possible combinations, the evaluation of the methods will be carried out separately according to the traditional adaptive filtering algorithms, NLMS or BNDR-LMS, in order to facilitate the understanding of the benefits provided by the use of cepstal analysis. If the NLMS algorithm is used, the resulting two hybrid methods (combinations of NLMS with AEC-CA and AEC-CAL) will be called hybrid methods based on cepstral analysis and NLMS. If the BNDR-LMS is used, the resulting two hybrid methods will be called hybrid methods based on cepstral analysis and BNDR-LMS.

6.5.1 Simulation Configurations

With the aim to assess the performance of the proposed hybrid methods, an experiment was carried out in a simulated environment to measure their ability to estimate the echo path impulse response and attenuate the acoustic echo signal. To this purpose, the same simulation configuration described in Section 6.4.4 was used.

6.5.2 Simulation Results

The proposed hybrid methods have the following adaptive filter parameters: stepsize μ and normalization factor δ from NLMS or BNDR-LMS; λ from AEC-CAI or AEC-CAL; and the adaptive filter length L_H that is common to all methods.

Since the hybrid methods will update $H(q,n)$ through the NLMS or BNDR-LMS algorithms most of the time, the adaptive filter parameters μ, δ and L_H of the hybrid methods were the same of the NLMS or BNDR-LMS obtained in Section 6.4.5.2. On the other hand, the adaptive filter parameter λ was chosen empirically in order to optimize, for each signal, the curves MIS(n) and ERLE(n) with regard to mean values according to the same procedure described in Section 6.4.5. Note that the results obtained from this optimization are sub-optimal because the parameters were not all optimized jointly.

6.5.2.1 AEC Based on Cepstral Analysis and NLMS

Firstly, the proposed AEC-CAI and AEC-CAL methods were combined with the NLMS algorithm. Table 6.3 summarizes the results obtained by the hybrid methods based on cepstral analysis and NLMS for different values of L_{fr} and ENR. In order to enrich the discussion, Figures 6.13 and 6.14 show the curves MIS(n) and ERLE(n) obtained by the NLMS and the hybrid methods based on cepstral analysis and NLMS when ENR = 30 and 40 dB, respectively, and for $L_{fr} = 8000, 16000, 80000$ samples.

For the same value of ENR, the proposed hybrid AEC methods based on cepstral analysis and NLMS outperformed the individual NLMS algorithm with respect to both MIS(n) and ERLE(n) with any value of L_{fr}. And, in these comparisons, the improvements were in general more significant in MIS(n) than in ERLE(n). Moreover, since the performance of the AEC-CAI and AEC-CAL methods improves by increasing ENR and/or L_{fr} as discussed in Section 6.4.5.1, the improvement caused by the hybrid methods in comparison

with the NLMS increases as ENR and/or L_{fr} increases. For ENR = 30 and 40 dB, it can be observed from Figures 6.13 and 6.14 that, with respect to MIS(n), the increase in L_{fr} results in a significant improvement mainly in convergent value while, with respect to ERLE(n), it results in a significant improvement mainly in convergence speed. On the other hand, the increase in ENR results in a significant improvement mainly in convergent value of both MIS(n) and ERLE(n).

For the same value of ENR, the hybrid method based on cepstral analysis and NLMS outperformed the individual AEC-CAI and AEC-CAL methods with regard to both MIS(n) and ERLE(n), with exception of a few cases. And, in these comparisons, the improvements were in general more significant in ERLE(n) than in MIS(n). Moreover, the hybrid method based on AEC-CAL always performed better than the one based on AEC-CAI, which was an expected result because the AEC-CAL performs better than AEC-CAI as discussed in Section 6.4.5.2.

Therefore, it can be concluded that, except in a few cases, the proposed hybrid methods based on cepstral analysis and NLMS achieved their goal by outperforming the individual methods with regard to MIS(n) and ERLE(n). The hybrid methods based on cepstral analysis and NLMS were even able to outperform the BNDR-LMS algorithm depending on the values of L_{fr} and ENR.

Table 6.3: Summary of the results obtained by the hybrid AEC methods based on cepstral analysis and NLMS.

	L_{fr}	ENR = 30 dB		ENR = 40 dB		ENR = 50 dB		ENR = ∞	
		$\overline{\text{MIS}}$	$\overline{\text{ERLE}}$	$\overline{\text{MIS}}$	$\overline{\text{ERLE}}$	$\overline{\text{MIS}}$	$\overline{\text{ERLE}}$	$\overline{\text{MIS}}$	$\overline{\text{ERLE}}$
AEC-CAI	8000	-16.27	30.86	-19.63	36.51	-22.86	41.04	-24.33	43.03
	12000	-16.71	31.25	-21.29	37.48	-26.05	42.64	-28.34	45.44
	16000	-17.04	31.50	-22.20	37.86	-27.08	43.32	-29.62	46.40
	32000	-18.25	32.11	-23.84	38.69	-29.18	44.60	-31.96	48.15
	80000	-19.41	32.61	-26.18	39.49	-32.05	45.90	-35.02	50.49
AEC-CAL	8000	-16.05	31.15	-21.22	37.70	-26.40	43.86	-30.73	49.31
	12000	-16.84	31.54	-22.38	38.42	-27.86	45.06	-32.87	51.10
	16000	-17.24	31.79	-23.07	38.80	-28.87	45.56	-34.08	51.73
	32000	-18.41	32.37	-25.36	39.58	-31.03	46.30	-34.54	52.24
	80000	-19.38	32.81	-26.44	39.87	-31.28	46.53	-35.08	52.26

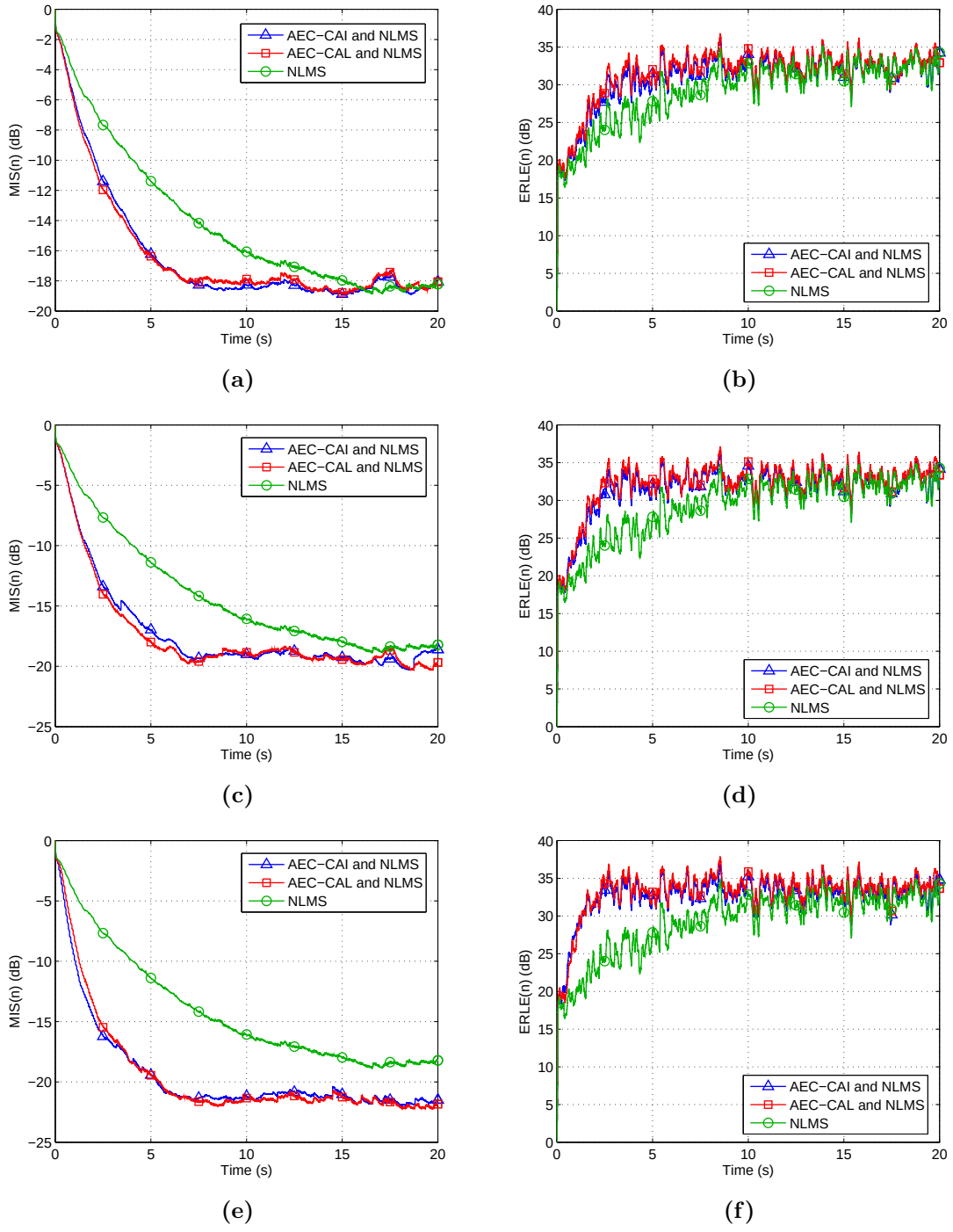


Figure 6.13: Performance comparison between the NLMS and AEC methods based on cepstral analysis and NLMS for ENR = 30: (a),(c),(e) MIS(n); (b),(d),(f) ERLE(n); (a),(b) $L_{fr} = 8000$; (c),(d) $L_{fr} = 16000$; (e),(f) $L_{fr} = 80000$.

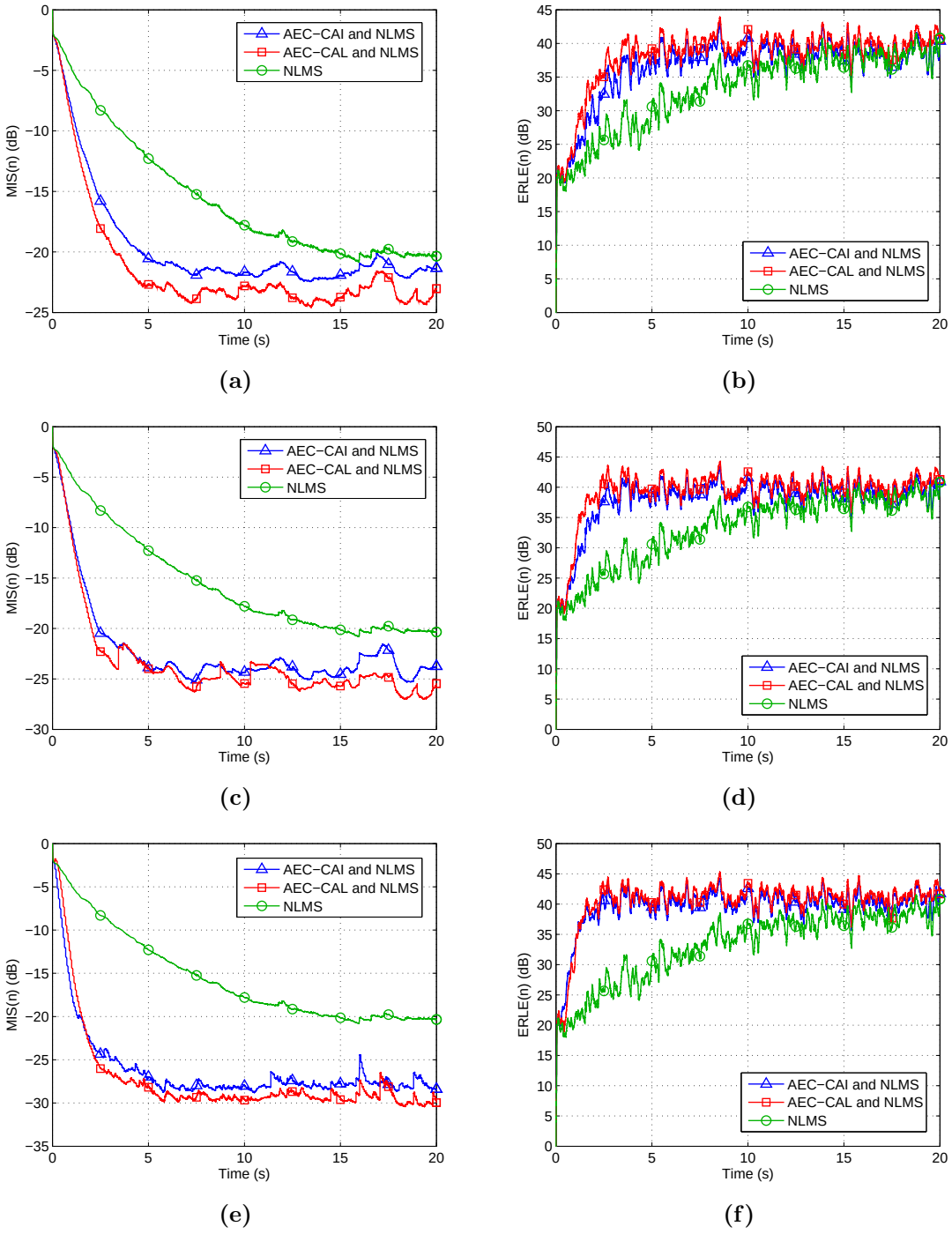


Figure 6.14: Performance comparison between the NLMS and AEC methods based on cepstral analysis and NLMS for ENR = 40: (a),(c),(e) MIS(n); (b),(d),(f) ERLE(n); (a),(b) $L_{fr} = 8000$; (c),(d) $L_{fr} = 16000$; (e),(f) $L_{fr} = 80000$.

6.5.2.2 AEC Based on Cepstral Analysis and BNDR-LMS

Analogously, the proposed AEC-CAI and AEC-CAL methods were combined with the BNDR-LMS algorithm. Table 6.4 summarizes the results obtained by the hybrid methods based on cepstral analysis and BNDR-LMS for different values of L_{fr} and ENR. In order to enrich the discussion, Figures 6.15 and 6.16 show the curves MIS(n) and ERLE(n) obtained by the BNDR-LMS and the hybrid methods based on cepstral analysis and BNDR-LMS when ENR = 30 and 40 dB, respectively, and for $L_{fr} = 8000, 16000, 80000$.

For the same value of ENR, the proposed hybrid AEC methods based on cepstral analysis and BNDR-LMS outperformed the individual BNDR-LMS algorithm regarding both MIS(n) and ERLE(n) with any value of L_{fr}. And, in these comparisons, the improvements were in general more significant in MIS(n) than in ERLE(n), except when ENR = ∞. In this ideal situation, the hybrid method based on AEC-CAL and BNDR-LMS achieved outstanding performances regarding ERLE(n) such that $\overline{\text{ERLE}} > 100$ dB.

Moreover, since the performance of the AEC-CAI and AEC-CAL methods improves by increasing ENR and/or L_{fr} as discussed in Section 6.4.5.1, the improvement caused by the hybrid methods in comparison with the BNDR-LMS increased as ENR and/or L_{fr} increases. For ENR = 30 and 40 dB, it can be observed from Figures 6.15 and 6.16 that, with respect to MIS(n), the increase in L_{fr} results in a significant improvement mainly in convergent value while, with respect to ERLE(n), it results in a significant improvement mainly in convergence speed. On the other hand, the increase in ENR results in a significant improvement mainly in convergent value of both MIS(n) and ERLE(n).

In addition, for the same value of ENR, the hybrid method based on cepstral analysis and BNDR-LMS outperformed the individual AEC-CAI and AEC-CAL methods with regard to both MIS(n) and ERLE(n), with exception of a few cases. And, in these comparisons, the improvements were more significant in ERLE(n) than in MIS(n). Moreover, the hybrid method based on AEC-CAL always performed better than the hybrid method based on AEC-CAI, which was an expected result because the AEC-CAL performs better than AEC-CAI as discussed in Section 6.4.5.2.

Therefore, it can be concluded that, except in some few cases, the proposed hybrid methods based on cepstral analysis and BNDR achieved their goal by outperforming the individual methods with regard to MIS(n) and ERLE(n). In general, these conclusions are very similar to those of Section 6.5.2.1. This means that the use of the proposed AEC based on cepstral analysis, AEC-CAI or AEC-CAL, even if sporadically, as every 1000 samples, can improve the results of the traditional adaptive filtering algorithms in AEC applications.

Table 6.4: Summary of the results obtained by the hybrid AEC methods based on cepstral analysis and BNDR-LMS.

	L_{fr}	ENR = 30 dB		ENR = 40 dB		ENR = 50 dB		ENR = ∞	
		$\overline{\text{MIS}}$	$\overline{\text{ERLE}}$	$\overline{\text{MIS}}$	$\overline{\text{ERLE}}$	$\overline{\text{MIS}}$	$\overline{\text{ERLE}}$	$\overline{\text{MIS}}$	$\overline{\text{ERLE}}$
AEC-CAI	8000	-17.84	33.29	-22.51	39.97	-26.92	46.04	-29.03	60.49
	12000	-18.06	33.68	-23.18	40.49	-28.12	46.68	-32.27	60.67
	16000	-18.17	33.88	-23.61	40.76	-29.16	47.03	-34.27	60.84
	32000	-19.07	34.40	-24.70	41.51	-30.93	47.81	-36.18	64.96
	80000	-19.98	35.14	-26.25	42.46	-33.53	48.86	-41.62	69.24
AEC-CAL	8000	-17.98	33.64	-22.51	40.53	-27.62	47.32	-37.15	107.45
	12000	-18.20	33.92	-23.07	41.03	-28.75	47.93	-38.41	130.07
	16000	-18.24	34.17	-23.54	41.36	-30.00	48.32	-43.06	142.23
	32000	-18.93	34.79	-24.67	42.21	-32.10	49.14	-50.21	152.73
	80000	-19.95	35.38	-26.10	42.77	-32.95	49.54	-52.94	157.30

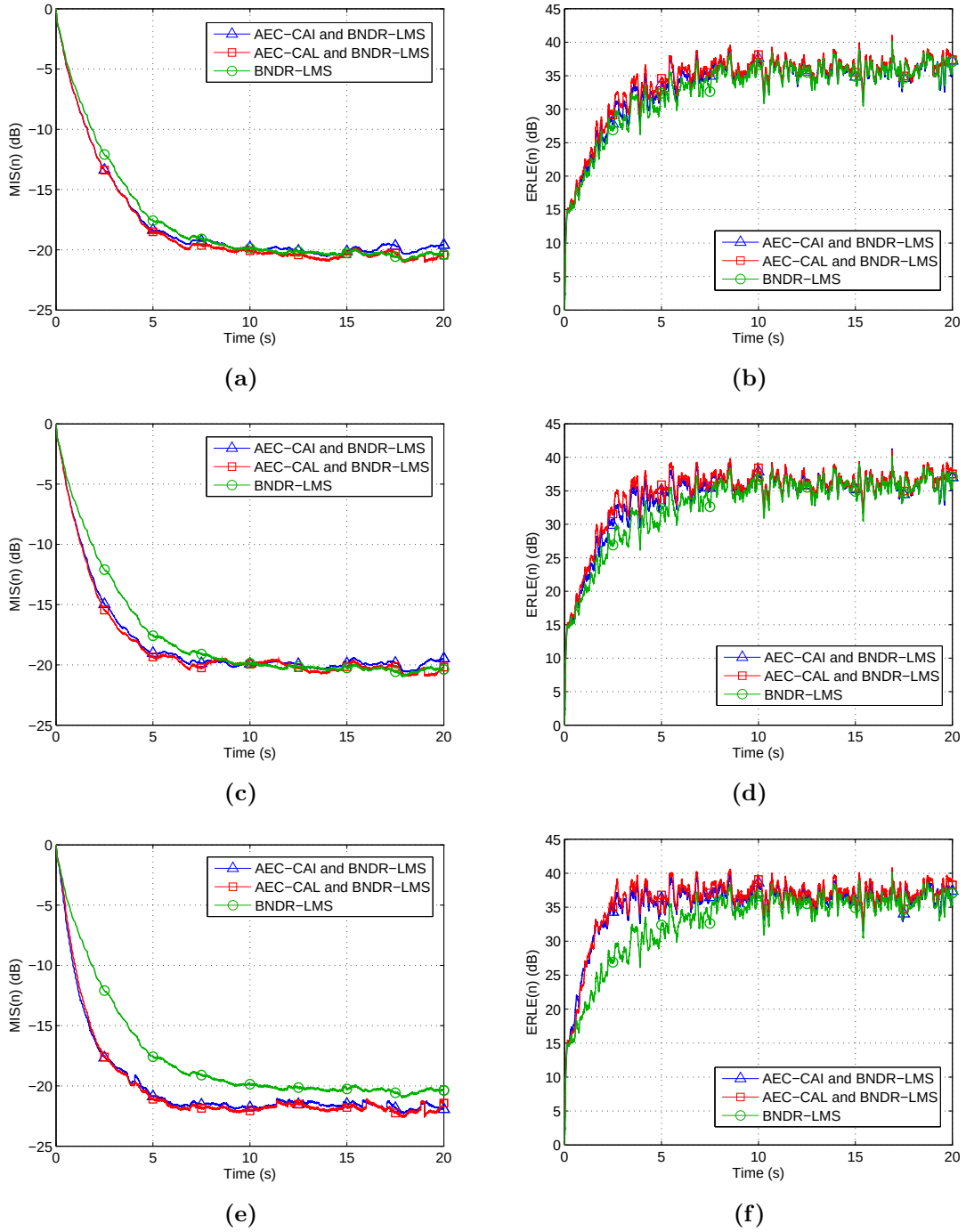


Figure 6.15: Performance comparison between the BNDR-LMS and AEC methods based on cepstral analysis and BNDR-LMS for ENR = 30: (a),(c),(e) MIS(n); (b),(d),(f) ERLE(n); (a),(b) $L_{fr} = 8000$; (c),(d) $L_{fr} = 16000$; (e),(f) $L_{fr} = 80000$.

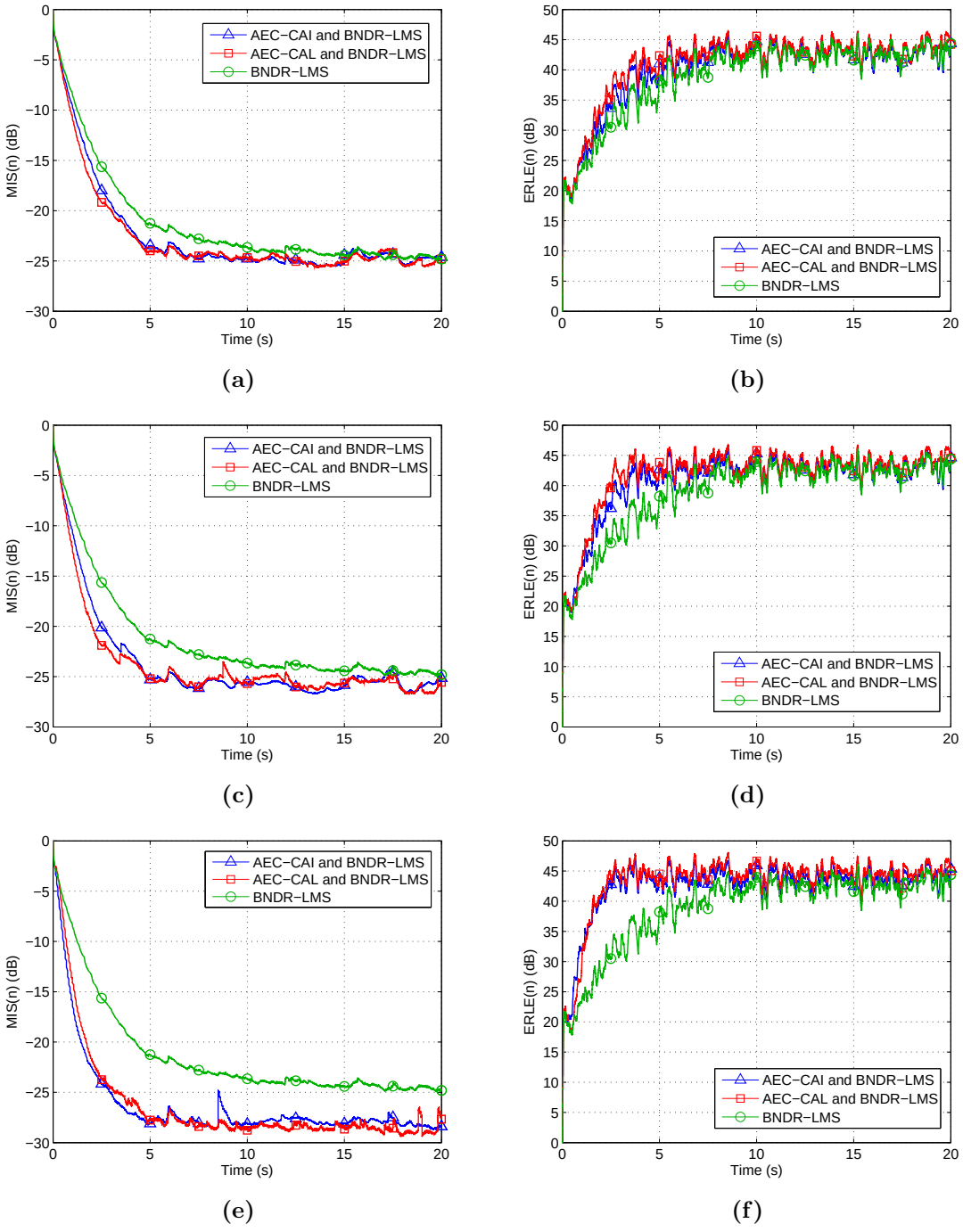


Figure 6.16: Performance comparison between the BNDR-LMS and AEC methods based on cepstral analysis and BNDR-LMS for ENR = 40: (a),(c),(e) MIS(n); (b),(d),(f) ERLE(n); (a),(b) $L_{fr} = 8000$; (c),(d) $L_{fr} = 16000$; (e),(f) $L_{fr} = 80000$.

6.6 Conclusions

This chapter addressed the topic of acoustic echo cancellation in teleconference systems. Similar to the AFC approach, the AEC approach uses an adaptive filter to identify the acoustic echo path and estimate the echo signal that is subtracted from the microphone signal. During the last decades, the use of the traditional gradient-based and least-squares-based adaptive filtering algorithms has been established in AEC applications.

The cepstral analysis, which was successfully applied to the AFC problem in the previous chapters, is now applied to the AEC problem. The availability of the loudspeaker and microphone signals in the AEC application is exploited to develop a new AEC method based on cepstral analysis with no lag (AEC-CA). The AEC-CA method selects, as usually, frames containing the newest L_{fr} samples of the system signals. An improved version, called improved AEC-CA (AEC-CAI), aims to obtain a more accurate frame of the microphone signal by partially performing the inverse of the overlap-and-add method using the adaptive filter as an estimate of the echo path. The AEC based on cepstral analysis with lag (AEC-CAL) method aims to obtain a even more accurate frame of the microphone signal by completely performing the inverse of the overlap-and-add method. Its drawback is a insertion of a lag equal to L_F in the estimation process.

The results showed that, depending on L_{fr}, the proposed AEC methods based on cesptral analysis are able to outperform the NLMS and BNDR-LMS, adaptive algorithms widely used in practical applications, and thereby can be alternative solutions to the AEC applications. However, in general, the proposed AEC methods proved to be more competitive regarding MIS(n) than ERLE(n), where they presented a worse performance in the first seconds. This can be explained by the fact that the proposed AEC methods directly identify the impulse response $\mathbf{f}(n)$ of the echo path through the cepstral analysis. On the other hand, the NLMS and BNDR-LMS algorithms indirectly identify $\mathbf{f}(n)$ by minimizing the instantaneous squared error signal. Therefore, the proposed AEC methods focus on identifying $\mathbf{f}(n)$, whose accuracy is measured by the MIS(n), while the NLMS and BNDR-LMS focus on minimizing $e^2(n)$, whose value is measured by the ERLE(n).

Hence, to combine the strengths of both methodologies, hybrid AEC methods that combine the AEC-CAI or AEC-CAL methods with the NLMS or BNDR algorithms were also proposed. As the AEC-CAI or AEC-CAL methods provide an instantaneous estimate of $\mathbf{f}(n)$, the adaptive filter in the proposed hybrid AEC methods was updated through the NLMS or BNDR-LMS algorithms most of the time and the AEC-CAI or AEC-CAL methods were sporadically used to accelerate or straighten the learning process.

The results showed that the proposed hybrid AEC methods can outperform the individual methods with regard to both MIS(n) and ERLE(n). This means that the proposed AEC methods based on cepstral analysis can be used alone or to improve the performance of the traditional adaptive filtering algorithms in AEC applications.

Chapter 7

Multi-channel Acoustic Echo Cancellation

7.1 Introduction

Chapter 6 dealt with the problem of acoustic echo in mono-channel teleconference systems. In recent years, these systems have evolved to provide a more realistic meeting experience. As regards sound, this is accomplished by using two or more independent audio channels through a configuration of two or more loudspeakers and microphones in each acoustic environment in order to enhance the sound realism in terms of spatiality. The acoustic coupling between the loudspeakers and microphones result in several acoustic echo paths. And, since the audio channels are independent in these systems, one adaptive filter is required to cancel each echo path.

Adaptive filters work quite well in mono-channel teleconference systems as discussed in Chapter 6, achieving good echo cancellation and low misalignment. But in a multi-channel system, a bias will be introduced in the impulse responses of the adaptive filters because of the strong correlation between the loudspeaker signals if they are originated from the same sound source. This will result in large misalignment between the adaptive filters and the echo paths. As a consequence, although it is possible to have good echo cancellation, the echo cancellation will worsen if the position of the speaker in the transmission room changes. In order to overcome the bias problem, the correlation between the loudspeaker signals should be reduced before feeding them to the adaptive filters.

During the past years, several decorrelation methods have been developed to overcome the bias problem in SAEC and an overview of them is presented in this chapter. Moreover, two sub-band hybrid methods based on FS will be proposed to decorrelate the loudspeaker signals in SAEC systems. The evaluation of the proposed methods is carried out in a simulated environment. Their ability to decrease the cross-correlation between the loudspeaker signals and thereby improve the performance of the SAEC system are measured as well as the audible distortion introduced in the processed loudspeaker signals.

7.2 Stereophonic Acoustic Echo Cancellation

In recent years, teleconference systems have evolved to telepresence systems which enable a more realistic meeting experience. This superior level of service is commonly accomplished through high-quality video and multi-channel audio. A multi-channel audio system uses two or more independent audio channels through a configuration of two or more loudspeakers and microphones in each acoustic environment in order to create the impression of sound heard from various directions, as in natural hearing.

Similarly to the mono-channel, the multi-channel acoustic echo cancellation uses adaptive filters to identify and track the echo paths. Such a scheme is depicted in Figure 7.1 for the stereophonic case, where the adaptive filters $H_{11}(q,n)$ and $H_{21}(q,n)$ model the echo paths $F_{11}(q,n)$ and $F_{21}(q,n)$, respectively. For now, disregard the pre-processing block so that $x'_k(n) = x_k(n),\ k = 1,2$. Then, estimates of the echo signals $\mathbf{f}_{11}(n) * x_1(n)$ and $\mathbf{f}_{21}(n) * x_2(n)$ are calculated as $\mathbf{h}_{11}(n) * x_1(n)$ and $\mathbf{h}_{21}(n) * x_2(n)$, respectively, and subtracted from the microphone signal $y_1(n)$, generating the error signal

$$e_1(n) = u_n(n) + [\mathbf{f}_{11}(n) - \mathbf{h}_{11}(n)] * x_1(n) + [\mathbf{f}_{21}(n) - \mathbf{h}_{21}(n)] * x_2(n), \tag{7.1}$$

which is the signal effectively sent to the transmission room.

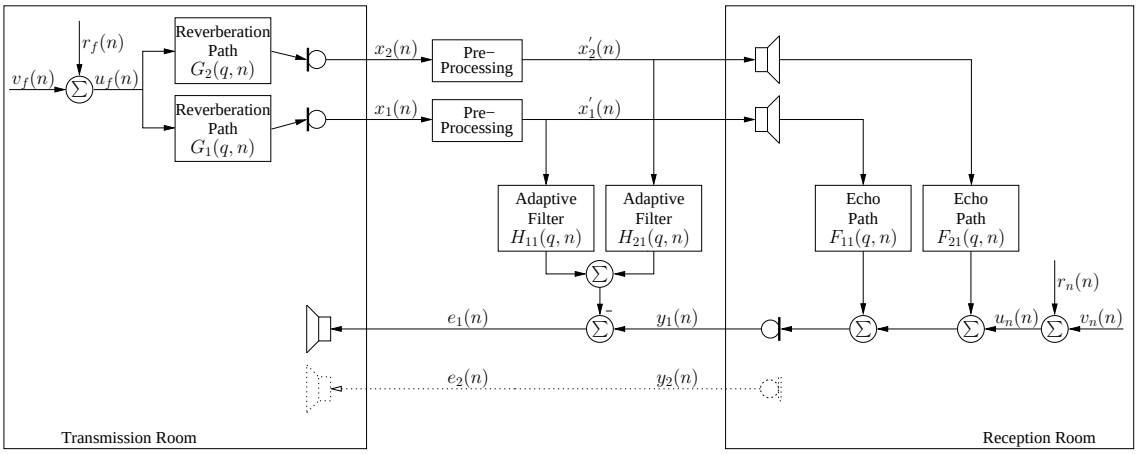


Figure 7.1: Stereophonic acoustic echo cancellation.

Defining $\tilde{\mathbf{f}}_{k1}(n) = \mathbf{f}_{k1}(n) - \mathbf{h}_{k1}(n)$, the mismatch between the impulse responses of the adaptive filter $H_{k1}(q,n)$ and echo path $F_{k1}(q,n)$, (7.1) can be written as

$$e_1(n) = u_n(n) + \tilde{\mathbf{f}}_{11}(n) * x_1(n) + \tilde{\mathbf{f}}_{21}(n) * x_2(n). \tag{7.2}$$

Therefore, the overall acoustic echo will be completed removed if

$$\tilde{\mathbf{f}}_{11}(n) * x_1(n) + \tilde{\mathbf{f}}_{21}(n) * x_2(n) = 0. \tag{7.3}$$

7.3 The Non-Uniqueness (Bias) Problem in Misalignment

The loudspeaker signals are defined as

$$\begin{aligned} x_1(n) &= u_f(n) * \mathbf{g}_1(n) \\ x_2(n) &= u_f(n) * \mathbf{g}_2(n), \end{aligned} \tag{7.4}$$

where $u_f(n)$ is the far-end speaker signal.

Replacing (7.4) in (7.3), the overall acoustic echo will be completed removed if

$$\left[\tilde{\mathbf{f}}_{11}(n) * \mathbf{g}_1(n) + \tilde{\mathbf{f}}_{21}(n) * \mathbf{g}_{21}(n)\right] * u_f(n) = 0 \tag{7.5}$$

or, in the frequency domain, if

$$\left[\tilde{F}_{11}(e^{j\omega}, n)G_1(e^{j\omega}, n) + \tilde{F}_{21}(e^{j\omega}, n)G_2(e^{j\omega}, n)\right] U_f(e^{j\omega}, n) = 0. \tag{7.6}$$

Regardless of $U_f(e^{j\omega}, n)$, the spectrum of the far-end speaker signal, the overall acoustic echo will be completely removed if

$$\tilde{F}_{11}(e^{j\omega}, n)G_1(e^{j\omega}, n) + \tilde{F}_{21}(e^{j\omega}, n)G_2(e^{j\omega}, n) = 0. \tag{7.7}$$

The problem of multi-channel AEC is that (7.7) has infinite solutions and they do not necessarily imply $\tilde{F}_{11}(e^{j\omega}, n) = \tilde{F}_{21}(e^{j\omega}, n) = 0$, which is the condition of complete alignment [5]. As a consequence, even if the impulse responses $\mathbf{f}_{11}(n)$ and $\mathbf{f}_{21}(n)$ of the echo paths are fixed, any variation in $G_1(e^{j\omega}, n)$ or $G_2(e^{j\omega}, n)$ requires adjustments of $\tilde{F}_{11}(e^{j\omega}, n)$ and $\tilde{F}_{21}(e^{j\omega}, n)$, except in the unlikely condition $\tilde{F}_{11}(e^{j\omega}, n) = \tilde{F}_{21}(e^{j\omega}, n) = 0$ [5].

Therefore, in order to completely remove the acoustic echo, the adaptive filters $H_{11}(q, n)$ and $H_{21}(q, n)$ must not only track the changes in the echo paths $F_{11}(q, n)$ and $F_{21}(q, n)$ in the reception room but also the changes in the reverberation paths $G_1(q, n)$ and $G_2(q, n)$ in the transmission room [5]. Apart from being undesirable, the latter changes are particularly hard to track because, if one speaker stops talking and another starts talking at a different place in the transmission room, the impulse responses $\mathbf{g}_1(n)$ and $\mathbf{g}_2(n)$ of the reverberation paths may change abruptly and by very large amounts [5].

Consider now two simultaneous far-end speakers, where the speech signal of the additional speaker is picked up by the microphones after going through the reverberation paths $G_3(q, n)$ and $G_4(q, n)$. The acoustic echo will be completely removed if

$$\begin{cases} \tilde{F}_{11}(e^{j\omega}, n)G_1(e^{j\omega}, n) + \tilde{F}_{21}(e^{j\omega}, n)G_2(e^{j\omega}, n) = 0 \\ \tilde{F}_{11}(e^{j\omega}, n)G_3(e^{j\omega}, n) + \tilde{F}_{21}(e^{j\omega}, n)G_4(e^{j\omega}, n) = 0. \end{cases} \tag{7.8}$$

The first condition in (7.8) is precisely (7.7). If $G_3(q, n)$ and $G_4(q, n)$ are linear independent from $G_1(q, n)$ and $G_2(q, n)$, (7.8) is only satisfied if $\tilde{F}_{11}(e^{j\omega}, n) = \tilde{F}_{21}(e^{j\omega}, n) = 0$.

Therefore, if two or more independent and spatially separated sources are active in the transmission room, the non-uniqueness problem essentially disappears because (7.7) cannot be simultaneously satisfied for two or more linear independent pairs of reverberation paths unless $\tilde{F}_{11}(e^{j\omega}, n) = \tilde{F}_{21}(e^{j\omega}, n) = 0$ [5].

A more refined analysis of the non-uniqueness problem in stereophonic AEC (SAEC) is provided in [6, 22]. It takes into account the lengths of the impulse responses of the reverberation paths, echo paths and adaptive filters, and proves that these lengths play a key role in SAEC. Considering $L_{G_1} = L_{G_2} = L_G$, $L_{F_{11}} = L_{F_{21}} = L_F$ and $L_{H_{11}} = L_{H_{21}} = L_H$, three possible scenarios can be described [6, 22]:

- $L_H \geq L_G$: the system has infinite solutions to the impulse responses $\mathbf{h}_{11}(n)$ and $\mathbf{h}_{21}(n)$ of the adaptive filters and all of them are undesirably dependent on the impulse responses $\mathbf{g}_1(n)$ and $\mathbf{g}_2(n)$ of the reverberation paths;

- $L_H < L_G$ and $L_H \geq L_F$: the system has unique solutions to the impulse responses $\mathbf{h}_{11}(n)$ and $\mathbf{h}_{21}(n)$ of the adaptive filters and the minimum value of the misalignment is zero, as desired;

- $L_H < L_G$ and $L_H < L_F$: the system has unique solutions to the impulse responses $\mathbf{h}_{11}(n)$ and $\mathbf{h}_{21}(n)$ of the adaptive filters but these solutions have a bias because of the strong correlation between the loudspeaker signals $x_1(n)$ and $x_2(n)$ if they are originated from the same sound source.

The last scenario is the real one because, in theory, both reverberation and echo paths have infinite lengths. So, due to the tails of the impulse responses $\mathbf{g}_1(n)$ and $\mathbf{g}_2(n)$ of the reverberation paths in the transmission room, the SAEC system has unique solutions to the impulse responses $\mathbf{h}_{11}(n)$ and $\mathbf{h}_{21}(n)$ of the adaptive filters. However, due to the unmodeled tails of the impulse responses $\mathbf{f}_{11}(n)$ and $\mathbf{f}_{21}(n)$ of the echo paths in the reception room, the unique solutions to $\mathbf{h}_{11}(n)$ and $\mathbf{h}_{21}(n)$ have a bias.

Therefore, the adaptive filters $H_{11}(q, n)$ and $H_{21}(q, n)$ generally converge to solutions that do not correctly match the real echo paths $F_{11}(q, n)$ and $F_{21}(q, n)$, respectively, which results in high misalignment. And the high misalignment is due to the strong correlation between the loudspeaker signals $x_1(n)$ and $x_2(n)$ which, in turn, depends on the impulse responses $\mathbf{g}_1(n)$ and $\mathbf{g}_2(n)$ of the reverberation paths in the transmission room.

It should be understood that it is possible to have good echo cancellation even when misalignment is large [6, 22]. However, in this case, the cancellation will worsen if the impulse responses $\mathbf{g}_1(n)$ and $\mathbf{g}_2(n)$ of the reverberation paths change [6, 22]. There are two ways to improve the misalignment. The first way is to use very long adaptive filters $H_{11}(q, n)$ and $H_{21}(q, n)$, which causes the traditional adaptive filtering algorithms to have a very slow convergence and high computational complexity. The second way is to reduce the correlation between the loudspeaker signals $x_1(n)$ and $x_2(n)$. The latter is the solution commonly used to overcome the bias problem in SAEC systems.

7.4 Solutions to The Non-Uniqueness (Bias) Problem

To overcome the bias in the impulse responses $\mathbf{h}_{11}(n)$ and $\mathbf{h}_{21}(n)$ of the adaptive filters, a pre-processing block is built into the SAEC system to decorrelate the loudspeaker signals $x_1(n)$ and $x_2(n)$ before feeding them to the adaptive filters, as shown in Figure 7.1. The pre-processing method must not introduce audible degradation, including modifications in the spatial image of the sound source, while keeping complexity low to be applied in real-time systems. Therefore, the challenge is to develop efficient decorrelation methods that do not significantly affect the perceptual quality of the stereo sound.

Several decorrelation methods have been proposed to add uncorrelated signals $p_1(n)$ e $p_2(n)$ to the loudspeaker signals $x_1(n)$ and $x_2(n)$, respectively, according to

$$x'_k(n) = x_k(n) + p_k(n), \quad k = 1, 2, \tag{7.9}$$

It is worth mentioning that the added signals $p_1(n)$ e $p_2(n)$ are the ones that would update the adaptive filters toward alignment while the loudspeaker signals $x_1(n)$ and $x_2(n)$ tend to disrupt the adaptation.

In [5], the added signals $p_k(n)$ were independent white noise signals. In [93], the added signals $p_k(n)$ were the loudspeaker signals modulated with independent white noise signals $w_k(n)$ as follows

$$p_k(n) = \varepsilon_k(n)x_k(n), \quad k = 1, 2, \tag{7.10}$$

where

$$\varepsilon_k(n) = \alpha\varepsilon_k(n-1) + (1-\alpha)w_k(n), \quad k = 1, 2. \tag{7.11}$$

In [94], the added signals $p_k(n)$ were noise signals shaped according to the human psychoacoustic model in order to mask the inserted distortion. A similar approach was proposed in [95] by applying perceptual audio coding/decoding (MPEG-1 Layer III) to the loudspeaker signals $x_1(n)$ and $x_2(n)$ such that $p_k(n)$ are uncorrelated quantization noise signals.

In [6, 22], instead of external noise, it was proposed to add to the loudspeaker signals $x_1(n)$ and $x_2(n)$ nonlinearly processed version of themselves. Then, the added signals were defined as

$$p_k(n) = \alpha f\left[x_k(n)\right], \quad k = 1, 2, \tag{7.12}$$

where $f\{\cdot\}$ must be a nonlinear function to reduce the linear relationship between the resulting signals $x'_1(n)$ and $x'_2(n)$, and α is the parameter that controls the amount of added nonlinearity.

In [6, 22], a half-wave rectifier (HWR) function was proposed such that

$$p_k(n) = \alpha\left(\frac{x_k(n) + |x_k(n)|}{2}\right), \quad k = 1, 2. \tag{7.13}$$

The stereo perception is not affected even with $\alpha = 0.5$ and the distortion introduced is hardly audible because of the nature of speech signals and psychoacoustic masking effects [6, 22].

Besides the half-wave rectifier, several nonlinear functions, such as square-law, square-sign, cubic, sign and full-wave rectifier, were evaluated in [96]. It was concluded that, for a roughly comparable decorrelation, the half-wave rectifier affects less the sound quality of speech signals. However, as the loudspeaker signals $x_1(n)$ and $x_2(n)$ are similar (or even the same), it is important to use different nonlinear functions for each [97]. Hence, in [97], a positive and a negative half-wave rectifier were used as follows

$$\begin{aligned} p_1(n) &= \alpha \left(\frac{x_1(n) + |x_1(n)|}{2} \right), \\ p_2(n) &= \alpha \left(\frac{x_2(n) - |x_2(n)|}{2} \right). \end{aligned} \tag{7.14}$$

The use of half-wave rectifiers changes the DC levels and the energies of $x_1'(n)$ and $x_2'(n)$ in relation to $x_1(n)$ and $x_2(n)$, respectively. Thus, it is customary to remove the added DC level and equalize the energies. The former can be performed by a highpass filter. The latter can approximately achieved by normalizing (7.13) or (7.14) with $\sqrt{1 + \alpha + \alpha^2}$.

The combination of the HWR at the frequency components (below 1 kHz) and comb filtering at the remaining frequency components (above 1 kHz) was proposed in [98]. However, it may lead to unacceptable degradation in the spatial image perception [99].

Another solution applies a time-varying filter to the loudspeaker signals $x_1(n)$ and $x_2(n)$ according to [100, 101, 102, 103]

$$x_k'(n) = x_k(n) * c_k(n) + x_k(n-1) * [1 - c_k(n)], \quad k = 1, 2, \tag{7.15}$$

where $0 \leq c_k(n) \leq 1$ is a periodic function with period Q.

In [100, 101, 102], the method was applied to only one loudspeaker signal. The preliminary idea was to make $x_1'(n) = x_1(n)$ by means of $c_1(n) = 1$ for the first $Q/2$ iterations and then to make $x_1'(n) = x_1(n-1)$ by means of $c_1(n) = 0$ for the following $Q/2$ iterations. However, the instantaneous change of $c_1(n)$ from 0 to 1 generates audible distortion that can be avoided by smoothly varying $c_1(n)$ between 0 and 1 over $L < Q/2$ samples [100, 101, 102]. The same occurs when $c_1(n)$ varies from 1 to 0. In [103], the method was applied simultaneously to the $x_1(n)$ and $x_2(n)$ using periodic functions $c_k(n)$ with different phases, which improved the performance of the adaptive filters and the sound quality.

The reference [104] proposed the use of time-varying all-pass filters $A_k(q, n)$ to modify the phase responses of the loudspeaker signals without affecting the magnitude responses. This was performed by making

$$X_k'(e^{j\omega}, n) = X_k(e^{j\omega}, n) A_k(e^{j\omega}, n), \quad k = 1, 2, \tag{7.16}$$

where

$$A_k(e^{j\omega}, n) = \frac{e^{-j\omega} - \alpha_k(n)}{1 - \alpha_k(n)e^{-j\omega}}. \tag{7.17}$$

The parameter $\alpha_k(n)$ is defined as

$$\alpha_k(n+1) = \alpha_k(n) + w_k(n), \tag{7.18}$$

where $w_k(n)$ are independent and identically distributed random variables that have a uniform probability distribution function over a specific interval. In order to ensure stability, $|\alpha_k(n)| < 1$. But $-0.9 \leq \alpha_k(n) \leq 0$ in order to not affect the stereo perception [104].

Another method based on phase modification of the loudspeaker signals proposed a sub-band phase modulation that uses a sine wave modulator function defined as [99]

$$\varphi(n, s) = \alpha(s) \sin(2\pi f_m n), \tag{7.19}$$

with constant frequency $f_m = 0.75$ Hz but amplitude $\alpha(s)$ dependent on the sub-band s. The amplitude $\alpha(s)$ started with 10 degrees and increased slowly to reach 90 degrees for frequencies above 2.5 kHz. The modulator function was applied in a conjugate complex way as follows

$$\begin{aligned} X_1'(e^{j\omega}, n) &= X_1(e^{j\omega}, n)e^{j\varphi(n,s)}, \\ X_2'(e^{j\omega}, n) &= X_2(e^{j\omega}, n)e^{-j\varphi(n,s)}. \end{aligned} \tag{7.20}$$

This phase modulation method can achieve superior perceptual quality of the stereo sound with similar misalignment performance compared with the HWR method [99]. The drawback is that, due to a low-intensity modulation at low frequencies, only a small decorrelation may be achieved in this frequency range [105, 106].

The reference [105] proposed a method based on the missing fundamental effect. This is a psychoacoustic phenomenon that, when the fundamental frequency is removed from a set of harmonics, causes the perception of pitch (fundamental frequency) not to change, although there is a slight change of timbre due to the number of harmonics reproduced [105]. This phenomenon has been explained as a human brain capability to process the information present in the overtones to calculate the missing fundamental frequency. As a consequence, the sound perceived is almost unchanged [105].

Hence, this method adaptively tracks and removes the pitch of only one of the loudspeaker signals by means of a notch filter. Being applied to the channel 1, the method aims to create a processed signal $x_1'(n)$ that is almost perceived as the original $x_1(n)$ while hopes that the modifications in $x_1'(n)$ reduce the correlation between $x_1'(n)$ and $x_2'(n)$. However, since the pitch of speech signals is usually located at low frequency components, the method may only decorrelate the loudspeaker signals $x_1(n)$ and $x_2(n)$ at the this

frequency range, thereby achieving only a partial decorrelation. When applied to the frequency range of $0 - 500$ Hz, this method achieves better sound quality and misalignment performance compared with a masked noise approach [105].

In [106], the missing fundamental approach and the sub-band phase modulation methods were combined. The former was applied at the low frequency components (0-500 Hz) and the latter in the remaining spectrum. In comparison with the phase modulation method, the combined method is able to improve the misalignment but degrades the sound quality [106].

7.5 Hybrid Pre-Processor Based on Frequency Shifting

As explained in Chapter 2, frequency shifting (FS) was initially proposed to increase the stability margin of PA systems. The idea is to shift, at each loop, the spectrum of the microphone signal by a few Hz so that its spectral peaks, including the frequency component that is responsible for the howling, fall into spectral valleys of the feedback path. In general, the use of FS smoothes the gain of the open-loop transfer function.

Later, it was observed that the use of FS to smooth the open-loop gain in PA systems, as originally proposed, also reduced the correlation between the loudspeaker and system input signals. Then, FS was also proposed as a decorrelation method in AFC systems in order to reduce the bias in the estimate of the feedback path provided by adaptive filters [2, 52]. It is noteworthy that a benefitial effect of using FS as a decorrelation method in AFC is that it also stabilizes the closed-loop system by smoothing the open-loop gain.

In SAEC, FS was already evaluated as a decorrelation method in [5], where the entire spectrum of one of the loudspeaker signals was shifted relative to the other [5]. And it was stated that this caused a total destruction of the stereo perception of the signals. Preliminary listening tests confirmed this effect since the position of the sound source appeared to oscillate proportionally to the applied frequency shift. However, the ability of this technique to decorrelate the loudspeaker signals was found to be quite high, thereby stimulating our attention and analysis.

It was understood that a frequency shift is critically perceived at the low frequencies of stereophonic images because, in this range, the human perception of the azimuthal position of sound sources is highly dependent on the interaural time difference [107]. And this dependence gradually reduces with increasing frequency until it vanishes [99, 107]. Therefore, in order to efficiently apply FS as a decorrelation method in SAEC so that stereo perception of the sound signal is not affected, the value of the frequency shift must be properly chosen as a function of the frequency range where it will be applied. To this purpose, a sub-band frequency shifting method should be developed.

Informal tests showed that a considerable frequency shift at high frequencies is difficult to be perceptually detected and may produce a great decorrelation between the loudspeaker signals in the frequency range where it is applied. On the other hand, a

small frequency shift at low frequencies (< 2 kHz) is easily perceived, which practically precludes its application in this frequency range. As a consequence, for SAEC, a decorrelation method based solely on FS should not decorrelate the loudspeaker signals at low frequencies, which certainly limits its misalignment correction performance.

Therefore, in a sub-band approach, some other decorrelation method should be applied at the low frequencies (< 2 kHz) to improve the misalignment correction performance. As discussed in the previous section, the phase modulation method can achieve only a small decorrelation in this frequency range. The method based on the fundamental missing problem can be applied only between 0 and 500 Hz and thus the frequency components between 500 and 2000 Hz would remain correlated. For speech signals, the methods based on perceptual coding/decoding and HWR similarly decorrelate the loudspeaker signals in this frequency range [95], and present similar performances both in misalignment and sound quality when applied to the full-band [99]. Then, because of its simple implementation, the HWR method was chosen for the low frequency components.

Coincidentally, preliminary tests showed that the widely used HWR method may achieve a considerable decorrelation at low frequencies but not at high frequencies. Therefore, the new hybrid method combines the strengths of both solutions: FS and HWR. Among many possible combinations, two hybrid configurations, called Hybrid1 and Hybrid2, were chosen to face the bias problem in SAEC. Considering 8 kHz band-limited speech signals, the hybrid methods and their configurations are summarized in Table 7.1.

Table 7.1: Configuration of the hybrid methods.

	Spectrum band		
	0-2 kHz	2-4 kHz	4-8 kHz
HWR	HWR: $\alpha = 0.5$	HWR: $\alpha = 0.5$	HWR: $\alpha = 0.5$
Hybrid1	HWR: $\alpha = 0.5$	HWR: $\alpha = 0.5$	FS: $\omega_0 = 5$ Hz
Hybrid2	HWR: $\alpha = 0.5$	FS: $\omega_0 = 1$ Hz	FS: $\omega_0 = 5$ Hz

The FS was applied by means of the implementation described in Section 2.3.1, where ω_0 is the value of the desired frequency shift. It is evident that the efficiency of this implementation depends on the length of the Hilbert filter: higher values of N_{hil} provide more accurate solutions but, at the same time, insert longer delays in the output signal. Fortunately, as the more $|m|$ increases the more the filter coefficients tend to zero, N_{hil} values do not need to be very large to have an accurate solution. The FS method applied a positive frequency shift in one channel and a negative in the other, and N_{hil} was equivalent to 20 ms. The HWR method was applied according to (7.9) and (7.14). Due to the intrinsic delay of the FS implementation, in the sub-bands of the hybrid methods where the HWR were applied, the signals had to be properly delayed.

7.5.1 Filter Bank

The hybrid methods used an orthogonal two-channel filter bank, which allows a perfect reconstruction, to split the spectra of the loudspeaker signals $x_1(n)$ and $x_2(n)$. The passband edge frequency of the lowpass filters was 0.48π, the passband edge frequency of the highpass filters was 0.52π and the maximum stopband ripple of the analysis filters was 60 dB. The frequency responses of the analysis and synthesis filters are shown in Figure 7.2.

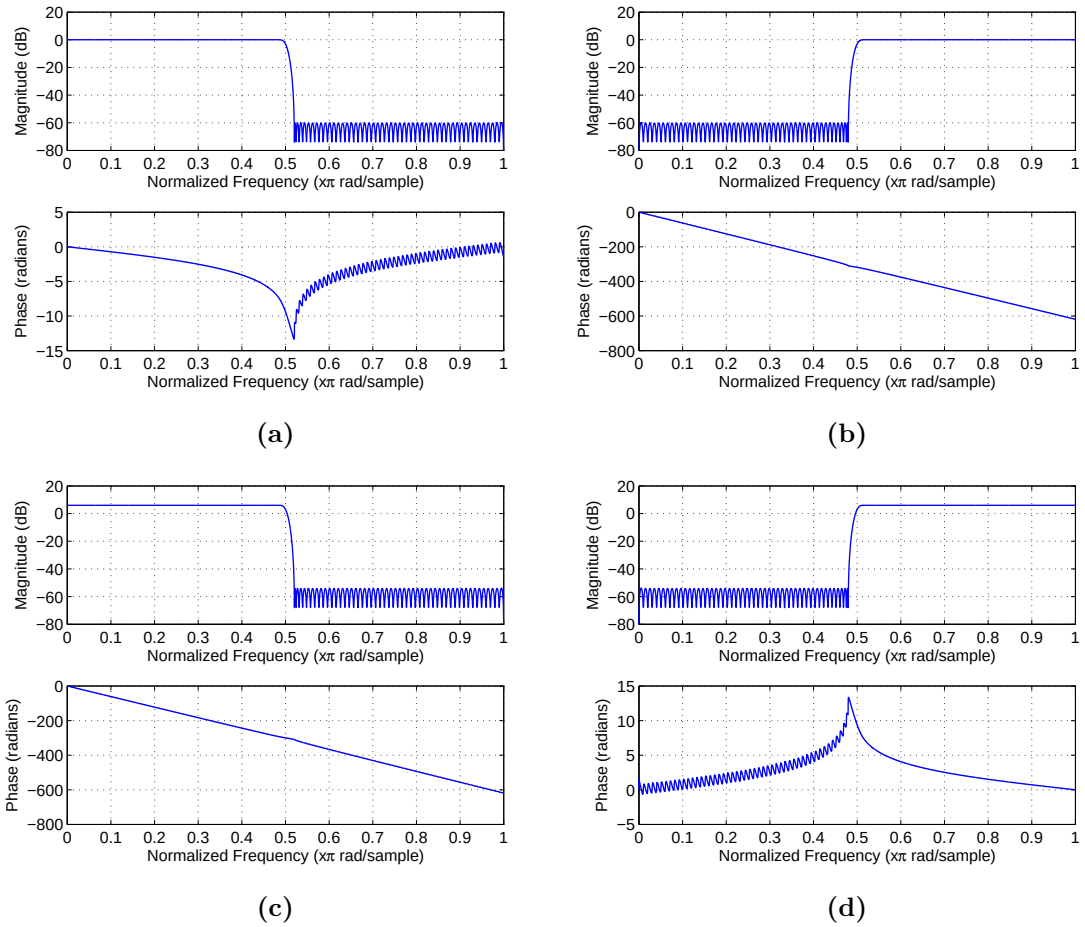


Figure 7.2: Frequency responses of the orthogonal filter bank: (a),(b) analysis filters; (c),(d) synthesis filters.

7.6 Simulation Configurations

With the aim to assess the relative performances of proposed hybrid methods, two experiments were carried in a simulated environment. In the first, the impulse responses of the transmission room were fixed throughout the simulation and the decorrelation methods were evaluated regarding their ability to decrease the cross-correlation between the loudspeaker signals and thereby improve the performance of the SAEC system. Moreover, the

audible distortion introduced by the methods were measured through a standardized subjective test. In the second, the impulse responses of the transmission room were changed during the simulation time in order to evaluate the ability of the decorrelation methods to make the performance of the SAEC system independent of transmission room. To these purposes, the following configuration was used.

7.6.1 Simulated Environment

To simulate a stereophonic teleconference system, two measured room impulse responses from [108] were used as the impulse responses $\mathbf{g}_1(n)$ and $\mathbf{g}_2(n)$ of the reverberation paths in the transmission room and two measured room impulse responses from [60] were used as the impulse responses $\mathbf{f}_1(n)$ and $\mathbf{f}_2(n)$ of the echo paths in the reception room. Consequently, $\mathbf{g}_k(n) = \mathbf{g}_k$ and $\mathbf{f}_k(n) = \mathbf{f}_k$, where $k = 1, 2$. The impulse responses were downsampled to $f_s = 16$ kHz and then truncated to lengths $L_G = L_F = 4000$ samples, and are illustrated in Figure 7.3. It is noteworthy that $\mathbf{g}_1$ and $\mathbf{g}_2$ had to be concatenated with very low-intensity white noise so that $L_G = 4000$.

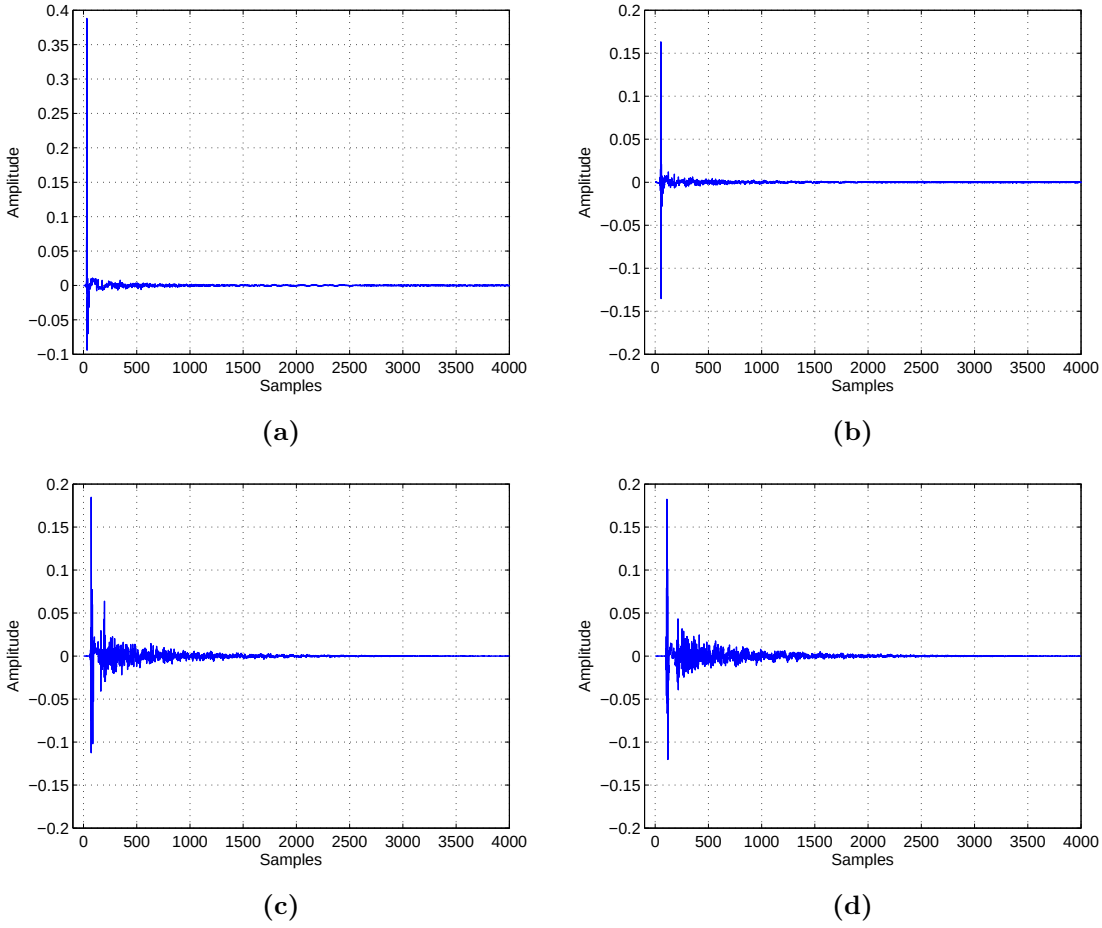


Figure 7.3: Impulse responses of the reverberation and echo paths: a) $\mathbf{g}_1$, b) $\mathbf{g}_2$, c) $\mathbf{f}_1$, d) $\mathbf{f}_2$.

In a first experiment, the impulse responses $\mathbf{g}_1$ and $\mathbf{g}_2$ of the transmission room were fixed throughout the simulation. But in a second experiment, $\mathbf{g}_1$ and $\mathbf{g}_2$ were changed at $t = 20$ s in order to verify the ability of the decorrelation methods to make the impulse responses $\mathbf{h}_1(n)$ and $\mathbf{h}_2(n)$ of the adaptive filters independent of them, as desired. To this end, $\mathbf{g}_1$ and $\mathbf{g}_2$ were changed to

$$\mathbf{g}_1' = \begin{bmatrix} \mathbf{0}_{47\times 1} \\ 1.2\overline{\mathbf{g}_2}_{L_G-47} \end{bmatrix} \tag{7.21}$$

and

$$\mathbf{g}_2' = \begin{bmatrix} \mathbf{0}_{25\times 1} \\ 0.8\overline{\mathbf{g}_1}_{L_G-25} \end{bmatrix}, \tag{7.22}$$

where $\overline{\mathbf{a}}_N$ denotes the N first samples of the vector $\mathbf{a}$.

The ambient noise condition of the reception room was close to real-world where $r_1(n) \neq 0$ such that the echo-to-noise ratio ENR = 30 dB.

7.6.2 Coherence Function

A very common metric in evaluating the efficiency of decorrelation methods is the coherence (COH). The COH is related to the conditioning of the covariance matrix and, in practice, is used to measure the cross-correlation between two signals in the frequency domain [6]. In this work, the performance of the decorrelation methods was evaluated through the COH function defined as [6]

$$\mathrm{COH}(e^{j\omega}, n) = \frac{S_{x_1'x_2'}(e^{j\omega}, n)}{\sqrt{S_{x_1'x_1'}(e^{j\omega}, n) S_{x_2'x_2'}(e^{j\omega}, n)}}, \tag{7.23}$$

where $S_{x_1'x_2'}(e^{j\omega}, n)$ is the short-term cross-power spectral density of the processed signals $x_1'(n)$ and $x_2'(n)$. The short-term cross-power spectral densities were computed using frames of 2000 samples taken with 50% overlap and an N_{FFT}-point FFT, where $N_{FFT} = 320000$ in order to achieve a fine resolution so that small values of ω_0 could be evaluated. The time average of (7.23) was denoted as $\mathrm{COH}(e^{j\omega})$.

7.6.3 Misalignment

The main goal of any decorrelation method in a SAEC system is to improve (decrease) the misalignment (MIS). The MIS measures the distance between the impulse responses of the adaptive filter and echo path, as discussed in Section 3.6.3, and has a bias in SAEC.

In this work, the performance of the SAEC system with the decorrelation methods was evaluated through the normalized MIS that, in the stereo case, is defined as

$$\mathrm{MIS}(n) = \sum_{k=1}^{2} \frac{\|\mathbf{f}_k(n) - \mathbf{h}_k(n)\|}{\|\mathbf{f}_k(n)\|}. \tag{7.24}$$

7.6.4 Echo Return Loss Enhancement

As previously explained, in SAEC, it is possible to have good echo cancellation even with high misalignment. However, the cancellation will worsen if the impulse responses $\mathbf{g}_1$ and $\mathbf{g}_2$ of the reverberation paths change. Therefore, in order to verify the ability of the decorrelation methods to keep good echo cancellation with changes in $\mathbf{g}_1$ and $\mathbf{g}_2$, the Echo Return Loss Enhancement (ERLE) metric was used. The ERLE measures the attenuation of the echo signal provided by the echo canceller as discussed in Section 6.4.4.3.

In this work, the performance of the SAEC system with the decorrelation methods was also measured through the normalized ERLE defined as

$$\mathrm{ERLE}(n) = \frac{\mathrm{LPF}\{\sum_{k=1}^{2} [y_k(n) - r_k(n)]^2\}}{\mathrm{LPF}\{\sum_{k=1}^{2} [e_k(n) - r_k(n)]^2\}}, \tag{7.25}$$

where $\mathrm{LPF}\{\cdot\}$ denotes a low-pass filter with a single pole at 0.999. As discussed in Section 6.4.4.3, the use of the low-pass filter is a common practice in AEC to smooth the curve $\mathrm{ERLE}(n)$ by removing the high frequency components without significantly affecting the convergence behavior.

7.6.5 MUSHRA

The perceived quality of the processed stereo signals was evaluated through the standardized subjective listening test called Multi Stimulus test with Hidden Reference and Anchor (MUSHRA) [109].

In MUSHRA, the evaluators assess the sound quality of the processed signal, one hidden reference signal and one hidden anchor signal (3.5 kHz band-limited reference signal) in comparison with the known reference signal (original unprocessed signal). The evaluators have access to all the signals, including the reference signal, at the same time so that they can carry out any comparison between them and hear all the signals at will. The sound quality of the signals is quantified from 0 (very bad quality) to 100 (indistinguishable from original) according to the continuous quality scale (CQS), which is shown in Figure 7.4.

In this case, the reference signals were the stereo signals formed by the unprocessed loudspeaker signals $x_1(n)$ and $x_2(n)$ while the processed signals were the stereo signals formed by the processed loudspeaker signals $x'_1(n)$ and $x'_2(n)$. The hidden reference signal and hidden anchor signal were used to recognize listeners as outliers.

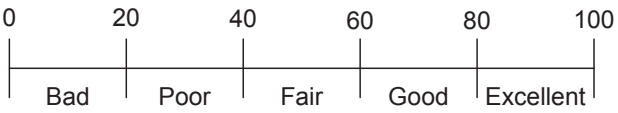


Figure 7.4: Grading scale of the MUSHRA test.

Rejecting the listeners classified as outliers, the listening test was performed by 10 evaluators where half of them were experienced listeners, i.e., that have experience in listening to sound in a critical way. The quality and stereo perception of the signals were considered together in the grading procedure. Due to the time consumption of the subjective quality tests, only 5 of the signals recorded in English were assessed.

7.6.6 Signal Database

The signal database was formed by the same 10 speech signals used in Chapters 3, 4 and 5. In the first experiment, where the impulse responses of the transmission room were fixed, the signals had a duration of 20 s as in Chapter 1. But in the second experiment, where the impulse responses of the transmission room changed at $t = 20$ s, the signals had a duration of 40 s. A detailed description can be found in Section 3.6.6.

7.7 Simulation Results

This section presents and discusses the performance of the proposed hybrid pre-processors based on frequency shifting, Hybrid1 and Hybrid2, using the configuration of the teleconference system, the evaluation metrics and the signals described in Section 7.6.

In order to analyze the performance of the decorrelation methods in the SAEC system, the adaptive filters $H_1(q,n)$ and $H_2(q,n)$ were updated using the Gauss-Seidel Fast Affine Projection (GSFAP) algorithm [110] with 20 projections and $L_H = 2000$ samples. Their stepsize μ and normalization parameter δ were optimized for each signal. From a predefined range for each one, the values of μ and δ were chosen empirically in order to optimize the curve $\mathrm{MIS}(n)$ with regard to minimum mean value within the simulation time. The optimal curve for the kth signal was denoted as $\mathrm{MIS}_k(n)$ while the $\mathrm{COH}(e^{j\omega})$ and $\mathrm{ERLE}(n)$ curves obtained with the same values of μ, δ and L_H were denoted as $\mathrm{COH}_k(e^{j\omega})$ and $\mathrm{ERLE}_k(n)$, respectively. The MUSHRA grade for the corresponding processed stereo signal given by the ith listener was defined as $\mathrm{MUSHRA}_{k,i}$.

Then, the mean curves MIS(n), COH($e^{j\omega}$) and ERLE(n) were obtained by averaging the curves of each signal according to

$$\begin{aligned} \mathrm{MIS}(n) &= \frac{1}{10}\sum_{k=1}^{10}\mathrm{MIS}_k(n), \\ \mathrm{COH}(e^{j\omega}) &= \frac{1}{10}\sum_{k=1}^{10}\mathrm{COH}_k(e^{j\omega}), \\ \mathrm{ERLE}(n) &= \frac{1}{10}\sum_{k=1}^{10}\mathrm{ERLE}_k(n). \end{aligned} \tag{7.26}$$

And their respective mean values were defined as

$$\begin{aligned} \overline{\mathrm{MIS}} &= \frac{1}{N_T}\sum_{n=1}^{N_T}\mathrm{MIS}(n), \\ \overline{\mathrm{COH}} &= \frac{1}{2\pi}\sum_{\omega=0}^{2\pi}\mathrm{COH}(e^{j\omega}), \\ \overline{\mathrm{ERLE}} &= \frac{1}{N_T}\sum_{n=1}^{N_T}\mathrm{ERLE}(n), \end{aligned} \tag{7.27}$$

where N_T is the number of samples relating to the simulation time. In addition to the mean coherence value considering the entire spectrum as defined in (7.27), mean coherence values considering only spectrum sub-bands were also calculated. Moreover, the asymptotic values of MIS(n) and ERLE(n) were defined as $\overrightarrow{\mathrm{MIS}}$ and $\overrightarrow{\mathrm{ERLE}}$, respectively, and were estimated only by graphically inspecting the curves.

The mean MUSHRA grade for the kth signal was calculated by averaging the grades of each listener as follows

$$\mathrm{MUSHRA}_k = \frac{1}{10}\sum_{i=1}^{10}\mathrm{MUSHRA}_{k,i} \tag{7.28}$$

and the overall MUSHRA grade of a decorrelation method was defined as

$$\overline{\mathrm{MUSHRA}} = \frac{1}{5}\sum_{k=1}^{5}\mathrm{MUSHRA}_k. \tag{7.29}$$

Note that the numbers 10 and 5 in (7.28) and (7.29) refer to the number of listeners and assessed speech signals, respectively.

7.7.1 First Experiment

In the first experiment, the impulse responses $\mathbf{g}_1$ and $\mathbf{g}_2$ of the transmission room were fixed throughout the simulation. In this experiment, the performance of the decorrelation

methods was analyzed regarding cross-correlation between the processed signals (COH), misalignment (MIS), echo chancellation (ERLE) and sound quality (MUSHRA).

Figure 7.5 shows the $\mathrm{COH}(e^{j\omega})$ between the processed signals $x'_1(n)$ and $x'_2(n)$ obtained by the HWR, Hybrid1 and Hybrid2 methods. In order to illustrate the bias problem in SAEC, the $\mathrm{COH}(e^{j\omega})$ achieved with no decorrelation method, i.e., when $x'_1(n) = x_1(n)$ and $x'_2(n) = x_2(n)$, is also considered. Figure 7.5a makes clear the strong correlation between the loudspeaker signals $x_1(n)$ and $x_2(n)$ in a stereophonic teleconference system where $\mathrm{COH}(e^{j\omega}) \approx 1$ in the entire spectrum. The HWR method obtained $\overline{\mathrm{COH}} = 0.85, 0.9$ and 0.92 in the low, middle and high sub-band, respectively, demonstrating the lower efficiency of the HWR method in the high frequencies as can be observed in Figure 7.5b. In Figure 7.5c, the good effect of the FS technique can already be noticed in the high sub-band (above 4 kHz) where it achieved $\overline{\mathrm{COH}} = 0.44$, less than half of the value obtained by the HWR. In the Hybrid2 method, the superiority of the FS technique with respect to decorrelation is extended to the middle sub-band (2-4 kHz), as illustrated in Figure 7.5d, where it achieved $\overline{\mathrm{COH}} = 0.46$. Therefore, because of their greater decorrelation capacity, it is expected that the proposed hybrid methods outperform the HWR method with regard to misalignment with an advantage for the Hybrid2 method.

Figure 7.6 shows the $\mathrm{MIS}(n)$ and $\mathrm{ERLE}(n)$ obtained by the SAEC system with the decorrelation methods under evaluation. The problem in SAEC is evident in the results obtained with no decorrelation method where good echo cancellation (high ERLE) is achieved even with high MIS. In fact, when using decorrelation methods, the performance of the SAEC system practically does not change regarding ERLE, as can be observed in Figure 7.6b, but greatly improves regarding MIS, as can be observed in Figure 7.6a. With no decorrelation method, the SAEC system achieved $\overline{\mathrm{MIS}} = -3.1$ dB, $\overrightarrow{\overline{\mathrm{MIS}}} \approx -3.4$ dB, $\overline{\mathrm{ERLE}} = -30.4$ dB and $\overrightarrow{\overline{\mathrm{ERLE}}} \approx -32$ dB . It can be observed that both proposed hybrid methods outperformed the HWR method with a advantage for Hybrid2 method. The Hybrid2 method achieved $\overline{\mathrm{MIS}} = -10.8$ dB and $\overrightarrow{\overline{\mathrm{MIS}}} \approx -13$ dB, outscoring respectively the HWR by 3.6 dB and 4 dB, and the Hybrid1 by 1.0 dB and 0.9 dB. These results of $\mathrm{MIS}(n)$ and $\mathrm{ERLE}(n)$ confirm the results of $\mathrm{COH}(e^{j\omega})$ previously presented.

With respect to the sound quality, Figure 7.7 shows, for each decorrelation method, the MUSHRA grades for each signal (MUSHRA_k) and the overall MUSHRA grades ($\overline{\mathrm{MUSHRA}}$) with a 95% confidence interval. The grades for the hidden references and anchors are also included. The results showed that, in general, the HWR method produces processed stereo signals with low degradation as widely recognized in the literature. And it also demonstrated that both proposed hybrid methods outperformed the HWR method with a slight average advantage for the Hybrid2. The Hybrid2 method achieved $\overline{\mathrm{MUSHRA}} = 87.2$, outscoring the HWR and Hybrid1 methods by 9.4 and 1.8, respectively. As the difference between the processed stereo signals resides only in the frequencies higher than 2 kHz, it can be concluded that the distortion introduced by the HWR method in this frequency range are more audible than those introduced by the frequency shifts.

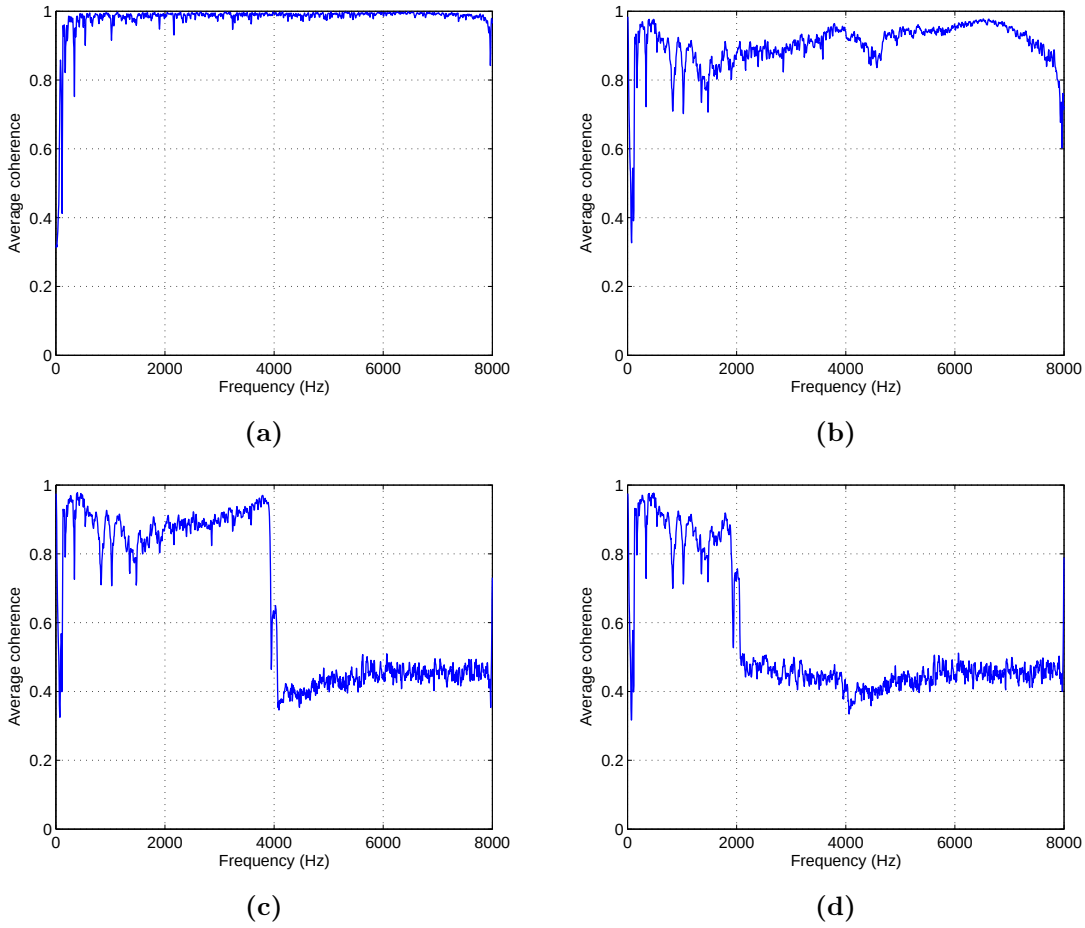


Figure 7.5: Average coherence function between the processed loudspeakers signals using: (a) no decorrelation method; (b) HWR; (c) Hybrid1; (d) Hybrid2.

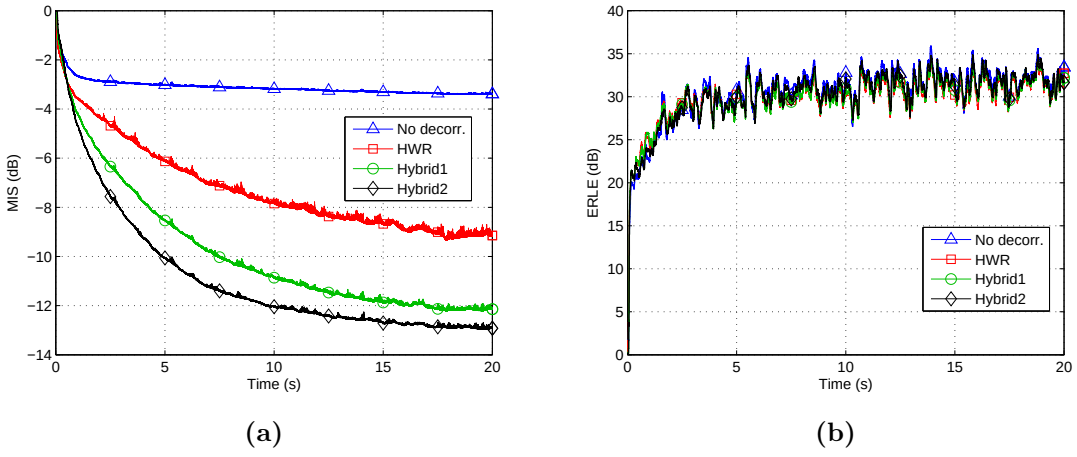


Figure 7.6: Average results of the SAEC system with the decorrelation methods: (a) MIS(n); (b) ERLE$_(n)$.

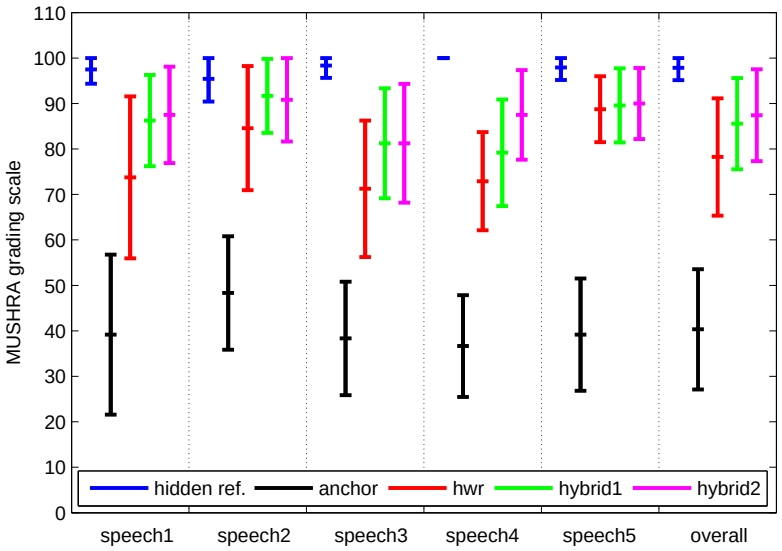


Figure 7.7: Average MUSHRA grades using the decorrelation methods.

In some of the depicted cases, the size of the 95% confidence interval is greater than desired. This was due to the subjective nature of the test and to the restricted number of evaluators. Moreover, the use of non-expert listeners usually tends to increase the variance of the results. But even so, the results are quite significant because, for all the signals, the new proposed hybrid methods presented an average perceptual quality superior to the widely used HWR method.

In conclusion, the results proved that FS can decorrelate stereo speech signals with small degradation in the global perceptual quality. To this purpose, the value ω_0 of the frequency shift must be chosen appropriately according to the spectrum sub-bands and not equally in the entire spectrum as did in [5]. However, the use of FS at the lower frequencies (< 2 kHz) is prohibitive and thus other decorrelation method should be used in this frequency range. In this work, the HWR method was used. The proposed Hybrid2 method caused the SAEC system to estimate the impulse responses of the echo paths with an MIS of -13 dB, outperforming the Hybrid1 and HWR by 0.9 and 4 dB, respectively. Moreover, the Hybrid2 method produced processed stereo signals with a MUSHRA grade of 87.2, outscoring the Hybrid1 and HWR methods by 1.8 and 9.4, respectively. Table 7.2 summarizes the results obtained by all the decorrelation methods evaluated.

7.7.2 Second Experiment

In the second experiment, the impulse responses $\mathbf{g}_1$ and $\mathbf{g}_2$ of the reverberation paths in the transmission room were changed at $t = 20$ s. In this experiment, the performance of the decorrelation methods was analyzed only regarding MIS and ERLE.

Table 7.2: Summary of the results obtained by the HWR, Hybrid1 and Hybrid2 methods.

	$\overline{\text{COH}}$				$\overline{\text{MIS}}$	$\overrightarrow{\overline{\text{MIS}}}$	$\overline{\text{ERLE}}$	$\overrightarrow{\overline{\text{ERLE}}}$	$\overline{\text{MUSHRA}}$
	0-2	2-4	4-8	0-8 kHz					
No decorr.	0.95	0.99	0.99	0.98	-3.1	-3.4	30.4	32.3	97.9
HWR	0.85	0.9	0.92	0.9	-7.2	-9	29.8	31.7	77.8
Hybrid1	0.85	0.9	0.44	0.66	-9.8	-12.1	29.9	31.9	85.4
Hybrid2	0.85	0.46	0.44	0.55	-10.8	-13	30.1	32.1	87.2

Figure 7.8 shows the MIS(n) and ERLE(n) obtained by the SAEC system with the HWR and Hybrid2 methods. The results obtained by the Hybrid1 method are not shown to make easier the visualization of the details of Figures 7.8b and 7.8c. The influence of the impulse responses $\mathbf{g}_1(n)$ and $\mathbf{g}_2(n)$ of the reverberation paths on SAEC is evident in Figure 7.8b, where the echo cancellation worsens when they were changed. As discussed in Section 7.3, this worsening in ERLE is directly related to the magnitude of MIS. It was proved in the first experiment that the proposed Hybrid2 causes the SAEC system to achieve the lowest MIS. The same occured in this experiment as shown in Figure 7.8a. Consequently, the Hybrid2 method causes the SAEC system to be less sensitive, with regard to echo cancellation, to variations in the impulse responses $\mathbf{g}_1$ and $\mathbf{g}_2$ of the reverberation paths, as can be observed in detail in Figure 7.8c.

7.8 Conclusions

The use of adaptive filters works quite well in a mono-channel teleconference system as discussed in Chapter 6. But in a multi-channel system, a bias is introduced in the impulse responses of the adaptive filters because of the strong correlation between the loudspeaker signals if they are originated from the same sound source. This results in high misalignment between the impulse responses of the adaptive filters and echo paths. As a consequence, although it is possible to have good echo cancellation, the echo cancellation will worsen if the impulse responses of the reverberation paths change.

To overcome this bias problem, pre-processing blocks are usually built into the multi-channel system to decorrelate the loudspeaker signals before feeding them to the adaptive filters. Nevertheless, the pre-processing methods must not introduce audible degradation, including modifications in the spatial image of the sound source, while keeping complexity low to be applied in real-time systems. Therefore, the challenge is to develop efficient decorrelation methods that do not affect the perceptual quality of the multi-channel sound.

In SAEC, the FS technique was already used as a decorrelation method such that the entire spectrum of one of the loudspeaker signals was shifted relative to the other but this caused a total destruction of the stereo perception of the signals. In this work, it was understood that a frequency shift is critically perceived at the low frequencies

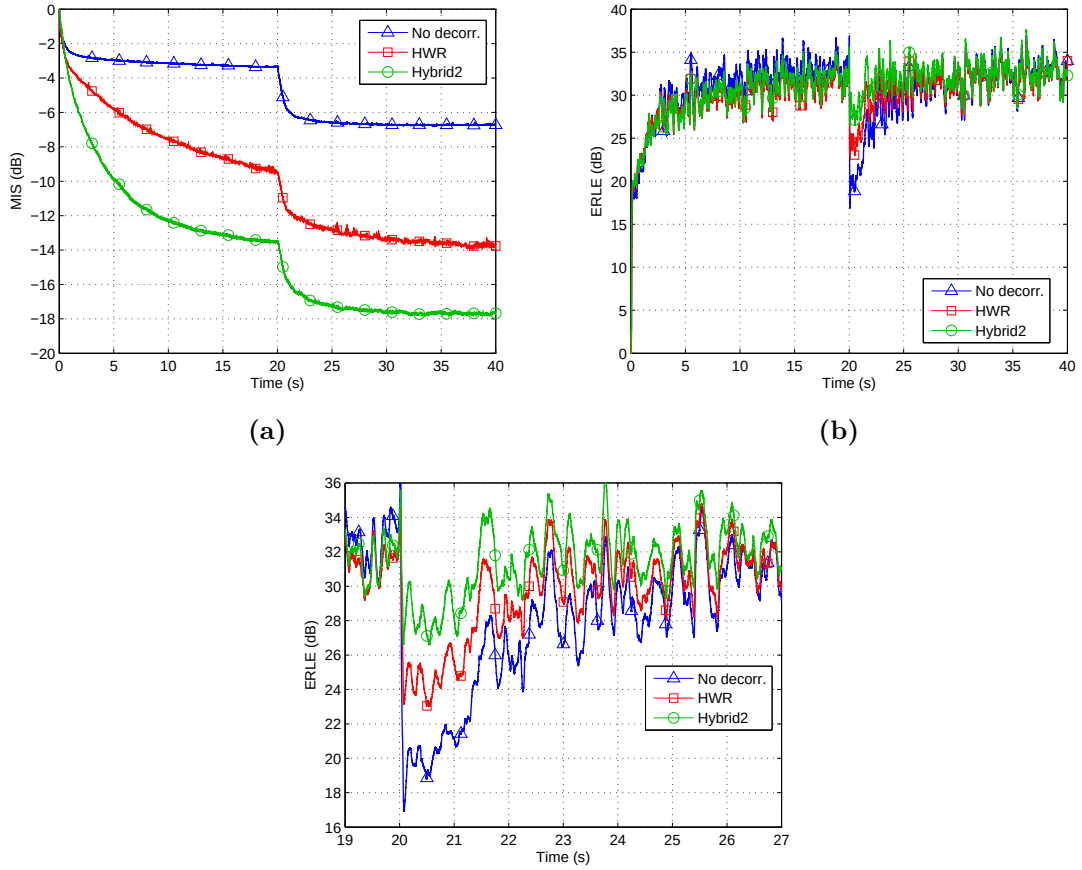


Figure 7.8: Average results of the SAEC system with the decorrelation methods when the impulse responses of the reverberation paths are changed at $t = 20$ s: (a) MIS(n), (b) ERLE(n); (c) zoom in ERLE(n).

of stereophonic images because, in this range, the human perception of the azimuthal position of sound sources is highly dependent on the interaural time difference. And this dependence gradually reduces with increasing frequency until it vanishes. Hence, in order to efficiently apply FS as a decorrelation method in SAEC so that the stereo perception of the sound signal is not significantly affected, a sub-band FS method was developed.

The application of frequency shifts at low frequencies is practically prohibited because it introduces audible distortion. On the other hand, the widely used half-wave rectifier method presents, at low frequencies, a good trade-off between reduction in the cross-correlation and introduction of audible degradation. Thus, two hybrid pre-processor methods, Hybrid1 and Hybrid2, that combine frequency shifting and half-wave rectifying were proposed. Considering 8 kHz band-limited speech signals, the Hybrid1 method applies a frequency shift of 5 Hz to the frequency components higher than 4 kHz and a half-wave rectifier function with $\alpha = 0.5$ to the remaining spectrum. The Hybrid2 method

applies a frequency shift of 5 Hz to the frequency components higher than 4 kHz, a frequency shift of 1 Hz to the frequency components in the range $2-4$ kHz and a half-wave rectifier function with $\alpha = 0.5$ to the remaining spectrum.

Simulation results demonstrated that the proposed Hybrid2 method caused the SAEC system to estimate the impulse responses of the echo paths with an MIS of -13 dB, outperforming the Hybrid1 and HWR by 0.9 and 4 dB, respectively. Consequently, Hybrid2 method caused the SAEC system to be less sensitive, with regard to echo cancellation, to variations in the impulse responses of the reverberation paths. Moreover, the Hybrid2 method produced processed stereo signals with a MUSHRA grade of 87.2, outscoring the Hybrid1 and HWR methods by 1.8 and 9.4, respectively. It may be concluded that the proposed hybrid methods cause the SAEC system to achieve a better estimate of the real echo paths and processed stereo signals with less perceptible degradation in comparison with the HWR method widely used in practical systems. The drawback is a small increase in the delay of the transmission channel due to the filterbank.

Chapter 8

Conclusion and Future Work

Communication is a necessity of human beings. With current technologies, communication systems have been developed in order to fulfill this need and make life easier. Inevitably, the communication systems use microphones and loudspeakers to pick up and play back the voice signal, respectively. The acoustic couplings from loudspeakers to microphones, that occur in the environment where these devices operate, may cause the signal played back by the loudspeakers to be picked up by the microphones and return into the communication system. The existence of the acoustic feedback is inevitable and may generate annoying effects that disturb the communication or even make it impossible.

This work investigated techniques to cancel the effects of the acoustic feedback in two different communication systems: public address (or reinforcement) and teleconference (or hands-free communication). In a PA system, a speaker employs microphone(s) and loudspeaker(s) along with an amplification system to apply a gain on his/her voice signal aiming to be heard by a large audience in the same acoustic environment. The acoustic feedback limits the system performance in two ways: first and more important, it causes the system to have a closed-loop transfer function that, depending on the amplification gain, may become unstable and, therefore, the MSG of the PA system has an upper limit; second, even if the MSG is not exceeded, the sound quality is affected by excessive reverberation. In a teleconference system, individuals employ microphone(s) and loudspeaker(s) along with a VoIP system to communicate remotely. It is considered that there is no closed-loop system, although it may exist, and thereby the acoustic feedback limits the system performance only with regard to sound quality, which is affected by echoes.

Primarily concerned with PA systems, this work detailed a cepstral analysis of a typical PA system. It was proved that the cepstrum of the microphone signal contains time domain information about the system, including its open-loop impulse response, if the NGC of the PA system is fulfilled. This work used these system information contained in the cepstrum of the microphone to update an adaptive filter in a typical AFC system, where an adaptive filter estimates the feedback path and subtracts an estimate of the feedback signal from the microphone signal.

To this end, a cepstral analysis of an AFC system, where an error signal is created from the microphone signal, was also detailed. It was proved that, in an AFC system, the cepstrum of the microphone signal may also contain time domain information about the AFC system including its open-loop impulse response. Then, a new AFC method based on cepstral analysis of the microphone signal, called AFC-CM, was proposed to identify the acoustic feedback path and cancel its effects. The AFC-CM method computes the open-loop impulse response of the PA system from the cepstrum of the microphone signal and, hereupon, calculates an estimate of the impulse response of the acoustic feedback path that is used to update the adaptive filter. But for that, besides the fulfillment of the NGC of the AFC system, it is also required to fulfill a gain condition as a function of the frequency responses of the forward path and adaptive filter. A complete theoretical discussion of why this issue limits the use of the cepstrum of the microphone signal in an AFC system was presented and it was also demonstrated in practice by the proposed AFC-CM method through simulations performed with single and multiple feedback paths.

Moreover, in an AFC system, it was also proved that the cepstrum of the error signal may contain time domain information about the AFC system including its open-loop impulse response. But, as an advantage over the microphone signal, only the fulfillment of the NGC of the AFC system is required for that. Then, a new AFC method based on cepstral analysis of the error signal, called AFC-CE, was proposed to identify the acoustic feedback path and cancel its effects. The AFC-CE method computes the open-loop impulse response of the AFC system from the cepstrum of the error signal and, hereupon, calculates an estimate of the impulse response of the acoustic feedback path that is used to update the adaptive filter. Improvements in performance of the AFC-CM and AFC-CE methods by the use of smoothing windows and highpass filtering were also proposed. Several simulations carried out considering single and multiple acoustic feedback paths demonstrated the effectiveness of the proposed AFC-CE method.

With regard to teleconference systems, the cepstral analysis, which is the basis of the proposed AFC methods, was applied in a different way to develop a new approach for mono-channel AEC. As a result, we proposed three new AEC methods: the AEC method based on cepstral analysis with no lag (AEC-CA), the improved AEC-CA (AEC-CAI) and the AEC method based on cepstral analysis with lag (AEC-CAL). The AEC-CAI and AEC-CAL methods may estimate more accurately the echo path impulse response by performing partially and completely, respectively, the inverse of the overlap-and-add method in the computation of the frame of the microphone signal. The drawback of the AEC-CAL is an estimation lag equal to the length of the echo path impulse response. Simulation results demonstrated that the proposed methods may be more competitive regarding misalignment than echo cancellation, where they presented a worse performance in the first seconds. Then, in order to overcome this weakness, hybrid AEC methods that combine the AEC-CAI and AEC-CAL with some traditional adaptive filtering algorithms were developed and evaluated.

In SAEC, additional processing is required to decorrelate the loudspeaker signals before feeding them to the adaptive filters but it must not insert audible degradations, including modifications in the spatial image of the sound source. The application of frequency shift in the entire spectrum of the loudspeakers signals was already tried as a decorrelation method but it destroyed the stereophonic effect. We understood that a frequency shift is critically perceived at the low frequencies of stereophonic images and this effect gradually reduces with increasing frequency until it vanishes. Therefore, a sub-band FS was proposed. Since frequency shifts in the low frequencies are prohibited, the traditional HWR method was applied below 2 kHz resulting in two new hybrid pre-processors. Results demonstrated that the proposed hybrid methods cause the SAEC system to achieve a better estimate of the echo paths and pre-processed stereo signals with less perceptible degradations in comparison with the HWR method.

8.1 Outlook for Future Work

With regard to the main theme of the work, AFC in PA and reinforcement systems, it would be interesting to pursue, in future work, the following research lines:

- Despite the experimental tests carried out in this work, it was not possible to validate the developed AFC-CM and AFC-CE methods, discussed in Chapter 3, in real-time. This validation should be tackled in future studies primarily through a personal computer and subsequently a digital signal processor.

- A combination of the developed AFC methods with other techniques to control the Larsen effect should be addressed. In particular, it would be very interesting to explore the application of an NHS method to the error signal, after the adaptive filtering, aiming to smooth the feedback path frequency response that was not modeled by the adaptive filter and further increase the MSG of the system. Moreover, the NHS approach has already proved to be competitive when the feedback path impulse response is quickly shifted. Therefore, it would be a very valuable task.

- Another avenue that can be explored is the use of two adaptive filters: foreground and background. The former would have a small convergence speed and would be responsible for the conservative solution of the system. The latter would have a fast convergence speed and would be responsible for tracking changes in the feedback path. Then, a control mechanism should be developed to decide over time which filter is most appropriate to be applied to the system. To this end, a comparison between the energies of the error signals, foreground and background, could be used. In addition, a comparison between the misalignments present in the cepstra of the error signals, foreground and background, could also be very helpful. Similar effect could also be achieved by making the parameter that controls the trade-off between robustness and tracking rate of the adaptive filter time-varying.

- The acoustic feedback in hearing aids is another major research topic that deserves further attention. Its difference from the problem tackled in this work is twofold: feedback path with very short length and great limitation on the computational power. The lack of the tail of the feedback path impulse response would improve the performance of the developed AFC methods due to their difficulty in estimating it, as discussed in Chapter 3. The possibility of an adaptive filter with very short length will greatly decrease the computational complexity required by the developed methods. Therefore, we believe that the AFC-CM and AFC-CE methods, developed in this work, are well placed to cope with this problem. Nevertheless, further research should be carried out to evaluate their performance under this scenario.

With regard to the second theme of the work, AEC in teleconference and hands-free communication systems, it would be interesting to pursue, in future work, the following research ideas:

- Despite the experimental tests carried out in this work, it was not possible to validate the developed mono-channel AEC methods, discussed in Chapter 6, as well as the develop pre-processor for SAEC, discussed in Chapter 7, in real-time. This validation should be tackled in future studies through a personal computer.

- Due to the constraint that the microphone signal must contain only the echo signal, the methodology based on cepstral analysis employed in Chapter 6 makes the AEC system to be sensitive to the ambient noise conditions. Hence, it would be pertinent to first apply noise reduction techniques, which are widely available in the literature, to the microphone signal aiming to overcome this limitation and improve the performance of the developed AEC-CA, AEC-CAI and AEC-CAL methods.

- Finally, the pre-processor based on frequency shifting for SAEC proved to efficiently decorrelate the high frequency components. However, its use at the low frequencies (< 2 kHz) is prohibited because it inserts audible degradations in the spatial image of the sound source. Therefore, further investigation is necessary to develop a technique able to extend such efficiency to the low frequency components (< 2 kHz) without affecting the perceptual quality of the stereo signals.

References

[1] K. R. Scherer, "Vocal communication of emotion: a review of research paradigms," *Speech Communication*, vol. 40, no. 1-2, pp. 227–256, April 2003.

[2] T. van Waterschoot and M. Moonen, "Fifty years of acoustic feedback control: state of the art and future challenges," *Proceedings of the IEEE*, vol. 99, no. 2, pp. 288–327, February 2011.

[3] G. Rombouts, T. van Waterschoot, K. Struyve, and M. Moonen, "Acoustic feedback cancellation for long acoustic paths using a nonstationary source model," *IEEE Transactions on Signal Processing*, vol. 54, no. 9, pp. 3426–3434, September 2006.

[4] A. Spriet, I. Proudler, M. Moonen, and J. Wouters, "Adaptive feedback cancellation in hearing aids with linear prediction of the desired signal," *IEEE Transactions on Signal Processing*, vol. 53, no. 10, pp. 3749–3763, October 2005.

[5] M. M. Sondhi and D. R. Morgan, "Stereophonic acoustic echo cancellation - an overview of the fundamental problem," *IEEE Signal Processing Letters*, vol. 2, no. 8, pp. 148–151, August 1995.

[6] J. Benesty, D. R. Morgan, and M. M. Sondhi, "A better understanding and an improved solution to the specific problems of stereophonic acoustic echo cancellation," *IEEE Transactions on Speech and Audio Processing*, vol. 6, no. 2, pp. 156–165, March 1998.

[7] M. R. Schroeder, "Stop feedback in public address systems," *Radio Electronics*, vol. 31, pp. 40–42, February 1960.

[8] ——, "Improvement of feedback stability of public address systems by frequency shifiting," in *Preprints of AES 13rd Convention*, New York, USA, October 1961.

[9] ——, "Improvement of feedback satbility of public address systems by frequency shifting," *Journal of Audio Engineering Society*, vol. 10, no. 2, pp. 108–109, April 1962.

[10] ——, "Improvement of acoustic-feedback stability by frequency shifting," *Journal of the Acoustical Society of America*, vol. 36, no. 9, pp. 1718–1724, 1964.

[11] M. D. Burkhard, "A simplified frequency shifter for improve acoustic feedback stability," *Journal of Audio Engineering Society*, vol. 11, no. 3, pp. 234–237, July 1963.

[12] E. Hänsler and G. Schmidt, *Acoustic Echo and Noise Control.* Hoboken, New Jersey: John Wiley & Sons, 2004.

[13] J. L. Nielsen and U. P. Svensson, "Performance of some linear time-varying systems in control of acoustic feedback," *Journal of the Acoustical Society of America*, vol. 106, no. 1, pp. 240–1254, July 1999.

[14] J. L. Nielsen, "Control of stability and coloration in electroacoustic systems in rooms," Ph.D. dissertation, Norges Tekniske Høgskole, 1996.

[15] T. van Waterschoot and M. Moonen, "Comparative evaluation of howling detection criteria in notch-filte-based howling suppression," in *Preprints of AES 126th Convention*, Munich, Germany, May 2009.

[16] ——, "Comparative evaluation of howling detection criteria in notch-filte-based howling suppression," *Journal of the Audio Engineering Society*, vol. 58, no. 11, pp. 923–949, November 2010.

[17] M. G. Siqueira and A. Alwan, "Bias analysis in continuous adaptation systems for hearing aids," in *Proceedings of IEEE Conference on Acoustics, Speech, and Signal Processing*, Phoenix, USA, March 1999, pp. 925–928.

[18] ——, "Steady-state analysis of continuous adaptation in acoustic feedback reduction systems for hearing-aids," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 4, pp. 443–453, July 2000.

[19] T. van Waterschoot, G. Rombouts, and M. Moonen, "On the performance of decorrelation by prefiltering for adaptive feedback cancellation in public address system," in *Processing of the 4th IEEE Benelux Signal Processing Symposium*, Hilvarenbeek, The Netherlands, April 2004, pp. 167–170.

[20] ITU-T G.164, "Echo suppressors," International Telecommunications Union, Geneva, Switzerland 1988.

[21] ITU-T G.165, "Echo cancellers," International Telecommunications Union, Geneva, Switzerland 1993.

[22] J. Benesty, D. R. Morgan, and M. M. Sondhi, "A better understanding and an improved solution to the specific problems of stereophonic acoustic echo cancellation," in *Proceedings of the IEEE International Conference on Accoustics, Speech, and Signal Processing*, Munich, Germany, April 1997, pp. 303–306.

[23] L. Ljung, *System Identification: Theory for the User.* Englewood Cliffs, New Jersey: Prentice-Hall, 1987.

[24] H. Nyquist, "Regeneration theory," *Bell System Techincal Journal*, vol. 11, pp. 126–147, 1963.

[25] A. J. Prestigiacomo and D. J. MacLean, "A frequency shifter for improving acoustic feedback stability," in *Preprints of AES 13rd Convention*, New York, USA, October 1961.

[26] C. V. Deutschbein, "Digital frequency shifting for electroacoustic feedback suppression," in *Preprints of AES 118th Convention*, Barcelona, Spain, May 2005.

[27] E. Berdahl and D. Harris, "Frequency shifting for acoustic howling supression," in *Proceedings of the 13th International Conference on Digital Audio Effects*, Graz, Austria, September 2010, pp. 1–4.

[28] M. A. Poletti, "The stability of multichannel sound systems with frequency shifting," *Journal of the Acoustical Society of America*, vol. 116, no. 2, pp. 853–871, August 2004.

[29] E. T. Patronis Jr., "Electronic detection of acoustic feedback and automatic sound system gain control," *Journal of Audio Engineering Society*, vol. 26, no. 5, pp. 323–326, May 1978.

[30] N. Osmanovic, V. E. Clarke, and E. Velandia, "An in-flight low latency acoustic feedback cancellation algorithm," in *Preprints of AES 123rd Convention*, New York, USA, October 2007.

[31] A. F. Rocha and A. J. S. Ferreira, "An accurate method of detection and cancellation of multiple acoustic feedbacks," in *Preprints of AES 118th Convention*, Barcelona, Spain, May 2005.

[32] G. W. Elko and M. M. Goodwin, "Beam dithering: Acoustic feedback control using a modulated-directivity loudspeaker array," in *Preprints of AES 93rd Convention*, San Francisco, USA, October 1992.

[33] ——, "Beam dithering: Acoustic feedback control using a modulated-directivity loudspeaker array," in *Proceedings of IEEE Conference on Acoustics, Speech, and Signal Processing*, Minneapolis, USA, April 1993, pp. 173–176.

[34] K. Kobayashi, K. Furuya, and A. Kataoka, "An adaptive microphone array for howling cancellation," *Acoustical Science and Technology*, vol. 24, no. 1, pp. 45–47, January 2003.

[35] G. Rombouts, A. Spriet, and M. Moonen, "Generalized sidelobe canceller based acoustic feedback cancellation," in *Proceedings of European Signal Processing Conference*, Florence, Italy, September 2006.

[36] ——, "Generalized sidelobe canceller based combined acoustic feedback and noise cancellation," *Signal Processing*, vol. 88, no. 3, pp. 571–581, March 2008.

[37] M. Miyoshi, "Inverse filtering of room acoustics," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 36, no. 2, pp. 145–152, February 1988.

[38] A. Favrot and C. Faller, "Adaptive equalizer for acoustic feedback control," *Journal of Audio Enginerring Society*, vol. 61, no. 12, pp. 1015–1021, December 2013.

[39] J. Hellgren and U. Forssell, "Bias of feedback cancellation algorithms based on direct closed loop indentification," in *Proceedings of IEEE Conference on Acoustics, Speech, and Signal Processing*, Istanbul, Turkey, June 2000, pp. 869–872.

[40] ——, "Bias of feedback cancellation algorithms in hearing aids based on direct closed loop identification," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 7, pp. 906–913, November 2001.

[41] J. Hellgren, "Analysis of feedback cancellation in hearing aids with Filtered-X LMS and the direct method of closed loop identification," *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 2, pp. 119–131, February 2002.

[42] G. Rombouts, T. van Waterschoot, and M. Moonen, "Robust and efficient implementation of the PEM-AFROW algorithm for acoustic feedback cancellation," *Journal of the Audio Engineering Society*, vol. 55, no. 11, pp. 955–966, November 2007.

[43] T. A. C. M. Claasen and W. F. G. Mecklenbrauker, "On stationary linear time-varying systems," *IEEE Transactions on Circuits and System*, vol. 29, no. 3, pp. 169–184, March 1982.

[44] A. V. Oppenheim, A. S. Wilsky, and I. T. Young, *Signals and Systems*. Prentice Hall, 1983.

[45] S. L. Marple Jr., "Computing the discrete-time 'analytic' signal via FFT," in *Proceedings of the 31th Asilomar Conferece on Signals, Systems & Computers*, Pacific Grove, USA, November 1997, pp. 1322–1325.

[46] ——, "Computing the discrete-time 'analytic' signal via FFT," *IEEE Transactions on Signal Processing*, vol. 47, no. 9, pp. 2600–2603, September 1999.

[47] S. J. Orfanidis, *Introdution to Signal Processing*. Upper Saddle River, New Jersey: Prentice Hall, 1995.

[48] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Upper Saddle River, New Jersey: Prentice-Hall, 1998.

[49] T. van Waterschoot and M. Moonen, "Assessing the acoustic feedback control performance of adaptive feedback cancellation in sound reinforcement systems," in *Proceedings of 17th European Signal Processing Conference*, Glasgow, Scotland, August 2009, pp. 1997–2001.

[50] G. Schmidt and T. Haulick, "Signal processing for in-car communication systems," *Signal Processing*, vol. 86, no. 6, pp. 1307–1326, June 2006.

[51] M. Guo, S. H. Jensen, and J. Jensen, "Novel acoustic feedback cancellation approaches in hearing aid applications using probe noise and probe noise enhancement," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 9, pp. 2549–2563, November 2012.

[52] F. J. van der Meulen, S. Kamerling, and C. P. Janse, "A new way of acoustic feedback suppression," in *Preprints of AES 104th Convention*, Amsterdam, The Netherlands, May 1998.

[53] M. Guo, S. H. Jensen, J. Jensen, , and S. L. Grant, "On the use of a phase modulation method for decorrelation in acoustic feedback cancellation," in *Proceedings of the 20th European Signal Processing Conference*, Bucharest, Romania, August 2012, pp. 2000–2004.

[54] A. Ortega and 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.

[55] T. van Waterschoot and M. Moonen, "Adaptive feedback cancellation for audio applications," *Elsevier Signal Processing*, vol. 89, no. 11, pp. 2185–2201, November 2009.

[56] U. Forssel, "Closed-loop identifcation: methods, theory, and applications," Ph.D. dissertation, Linköpings Universitet, 1999.

[57] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*. Englewood Cliffs, New Jersey: Prentice-Hall, 1978.

[58] J. R. Deller Jr., J. H. L. Hansen, and J. G. Proakis, *Discrete-Time Processing of Speech Signals*. Piscataway, New Jersey: IEEE Press, 2000.

[59] R. P. Ramachandran and P. Kabal, "Pitch prediction filter in speech conding," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 37, no. 4, pp. 467–478, Abril 1989.

[60] M. Jeub, M. Schäfer, and P. Vary, "A binaural room impulse response database for the evaluation of dereverberation algorithms," in *Proceedings of the International Conference on Digital Signal Processing*, Santorini, Greece, July 2009.

[61] ANSI, "ANSI S3.5: American national standard methods for calculation of the speech intelligibility index," American National Standard Institute, 1997.

[62] ITU-T P.862, "Perceptual evaluation of speech quality (PESQ): objective method for end-to-end speech quality assessment of narrow band telephone networks and speech codecs," International Telecommunications Union, Geneva, Switzerland 2001.

[63] ITU-T P.862.2, "Wideband extension to recommendation P.862 for the assessment of wideband telephone networks and speech codecs," International Telecommunications Union, Geneva, Switzerland 2005.

[64] A. A. de Lima, F. P. Freeland, R. A. de Jesus, B. C. Bispo, L. W. P. Biscainho, S. L. Netto, A. Said, T. Kalker, R. W. Schafer, B. Lee, and M. Jam, "On the quality assessment of sound signals," in *Proceedings of the IEEE International Symposium on Circuits and Systems*, Seattle, USA, May 2008, pp. 416–419.

[65] B. C. Bispo, P. A. A. Esquef, L. W. P. Biscainho, A. A. de Lima, F. P. Freeland, R. A. de Jesus, A. Said, B. Lee, R. W. Schafer, and T. Kaller, "EW-PESQ: A quality assessment method for speech signals sampled at 48 kHz," *Journal of the Audio Engineering Society*, vol. 58, no. 4, pp. 251–268, April 2010.

[66] A. A. de Lima, F. P. Freeland, P. A. A. Esquef, L. W. P. Biscainho, B. C. Bispo, R. A. de Jesus, S. L. Netto, R. W. Schafer, A. Said, B. Lee, and T. Kalker, "Reverberation assessment in audioband speech signals for telepresence systems," in *Proceedings of International Conference on Signal Processing and Multimedia Applications*, Porto, Portugal, July 2008, pp. 257–262.

[67] A. A. de Lima, S. L. Netto, L. W. P. Biscainho, F. P. Freeland, B. C. Bispo, R. A. de Jesus, R. W. Schafer, A. Said, B. Lee, and T. Kalker, "Quality evaluation of reverberation in audioband speech signals," in *e-Business and Telecommunications - Communications in Computer and Information Science*, J. Filipe and M. S. Obaidat, Eds. Springer, 2009, vol. 48, pp. 384–396.

[68] ITU-T P.862.3, "Application guide for objective quality measurement based on recommendations P.862, P.862.1 and P.862.2," International Telecommunications Union, Geneva, Switzerland 2007.

[69] S. J. Hubbard, "A cepstrum-based acoustic echo cancellation technique for improving public address system performance," Ph.D. dissertation, Georgia Institute of Technology, August 1994.

[70] J. M. Tribolet, *Seismic Applications of Homomorphic Signal Processing.* Englewood Cliffs, New Jersey: Prentice-Hall, 1979.

[71] A. V. Oppenheim and R. W. Schafer, "From frequency to quefrency: A history of the cepstrum," *IEEE Signal Processing Magazine*, vol. 21, pp. 95–106, September 2004.

[72] P. S. R. Diniz, *Adaptive Filtering: Algorithms and Practical Implementation*, 2nd ed. Norwell, Massachusetts: Kluwer Academic Publishers, 2002.

[73] O. Vinyals, G. Friedland, and N. Mirghafori, "Revisiting a basic function on current CPUs: a fast logarithm implementation with adjustable accuracy," International Computer Science Institute, Tech. Rep., June 2007.

[74] S. Haykin, *Adaptive Filter Theory*, 3rd ed. Upper Saddle River, New Jersey: Prentice Hall, 1996.

[75] R. L. Das and M. Chakraborty, "Sparse adaptive filters - an overview and some new results," in *Proceedings of the 2012 IEEE International Symposium on Circuits and Systems*, Seoul, South Korea, May 2012, pp. 2745–2748.

[76] A. W. Khong and P. A. Naylor, "Efficient use of sparse adaptive filters," in *Proceedings of Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, USA, October 2006.

[77] D. L. Duttweiler, "Proportionate normalized least-mean-square adaptation in echo cencelers," *IEEE Transactions on Speech and Audio Processing*, vol. 8, pp. 508–518, September 2000.

[78] J. Benesty and S. L. Gay, "An improved PNLMS algorithm," in *Proceedings of the IEEE International Conference on Accoustics, Speech, and Signal Processing*, Orlando, USA, May 2002, pp. 1881–1884.

[79] C. Paleologu, J. Benesty, and S. Ciochina, "An improved proportionate NLMS algorithm based on the l0 norm," in *Proceedings of the IEEE International Conference on Accoustics, Speech, and Signal Processing*, Dallas, USA, March 2010, pp. 309–312.

[80] C. Paleologu, S. Ciochina, and J. Benesty, "An efficient proportionate affine projectionalgorithm for echo cancellation," *IEEE Signal Processing Letter*, vol. 17, pp. 165–168, February 2010.

[81] C. Paleologu, J. Benesty, F. Albu, and S. Ciochina, "An efficient variable step-size proportionate affine projection algorithm," in *Proceedings of the IEEE International Conference on Accoustics, Speech, and Signal Processing*, Prague, Czech Republic, May 2011, pp. 77–80.

[82] T. Gänsler, S. L. Gay, M. M. Sondhi, and J. Benesty, "Double-talk robust fast converging algorithms for network echo cancellation," *IEEE Transactions on Speech and Audio Processing*, vol. 8, pp. 656–663, November 2000.

[83] O. Hoshuyama, R. A. Goubran, and A. Sugiyama, "A generalized proportionate variable step-size algorithm for fast changing acoustic environments," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, Montreal, Canada, May 2004, pp. 161–164.

[84] B. Widrow and S. D. Stearns, *Adaptive Signal Processing.* Englewood Cliffs, New Jersey: Prentice Hall, 1985.

[85] D. L. Duttwieler, "A twelve-channel digital echo canceler," *IEEE Transactions on Communication*, vol. 26, no. 5, pp. 647–653, May 1978.

[86] H. Ye and B. X. Wu, "A new double-talk detection algorithm based on orthogonality theorem," *IEEE Transactions on Communication*, vol. 39, pp. 1542–1545, November 1991.

[87] J. H. Cho, D. R. Morgan, and J. Benesty, "An objective technique for evaluating doubletalk detectors in acoustic echo cancelers," *IEEE Transactions on Seepch and Audio Processing*, vol. 7, no. 6, pp. 718–724, November 1999.

[88] J. Benesty, D. R. Morgan, and J. H. Cho, "A new class of double-talk detectors based on cross-correlation," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 2, pp. 168–172, March 2000.

[89] M. A. Iqbal, J. W. Stokes, and S. L. Grant, "Normalized double-talk detection based on microphone and aec error cross-correlation," in *Proceedings of IEEE International Conference on Multimedia and Expo*, July 2007, pp. 360–363.

[90] ITU-T G.168, "Digital network echo cancellers," International Telecommunications Union, Geneva, Switzerland 2012.

[91] M. L. R. de Campos, P. S. R. Diniz, and J. A. Apolinário Jr., "On normalized data-resusing and affine-projections algorithms," in *6th IEEE Internacional Conference on Electronics, Circuits and Systems*, 1999, pp. 843–846.

[92] J. A. Apolinário Jr., M. L. R. de Campos, and P. S. R. Diniz, "Convergence analysis of the binormilized data-reusing LMS algorithm," *IEEE Transactions on Signal Processing*, vol. 48, no. 11, pp. 3235–3242, November 2000.

[93] S. Shimauchi and S. Makino, "Stereo projetction echo canceller with true echo path estimation," in *Proceedings of the IEEE International Conference on Accoustics, Speech, and Signal Processing*, Detroit, USA, May 1995, pp. 3059–3062.

[94] A. Gilloire and V. Turbin, "Using auditory properties to improve the behaviour of stereophonic acoustic echo cancellers," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Seatle, USA, May 1998, pp. 3681–3684.

[95] T. Gänsler and P. Eneroth, "Influence of audio coding on stereophonic acoustic echo cancellation," in *Proc. IEEE ICASSP*, Seatle, USA, May 1998, pp. 3649–3652.

[96] D. R. Morgan, J. L. Hall, and J. Benesty, "Investigation of several types of nonlinearities for use in stereo acoustic echo cancellation," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 6, pp. 686–696, September 2001.

[97] J. Benesty, D. R. Morgan, J. L. Hall, and M. M. Sondhi, "Synthesized stereo combined with acoustic echo cancellation for desktop conferencing," in *Proceedings of IEEE Conference on Acoustics, Speech, and Signal Processing*, Phoenix, USA, March 1999, pp. 853–856.

[98] ——, "Sterophonic acoustic echo cancellation using nonlinear trasformations and comb filtering," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Seatle, USA, May 1998, pp. 3673–3676.

[99] J. Herre, H. Buchner, and W. Kellermann, "Acoustic echo cancellation for surround sound using perceptually motivated convergence enhancement," in *Proceedings of the IEEE International Conference on Accoustics, Speech, and Signal Processing*, Honolulu, Hawaii, USA, April 2007, pp. 17–20.

[100] Y. Joncour and A. Sugiyama, "A stereo echo canceler with pre-processing for correct echo-path indentification," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Seatle, USA, May 1998, pp. 3677–3680.

[101] Y. Joncour, A. Sugiyama, and A. Hirano, "Dsp implementations and performance evaluation of a stereo echo canceller with pre-processing," in *Proceedings of the 9th European Signal Processing Conference*, Rhodos, Greece, September 1998, pp. 981–984.

[102] A. Sugiyama, Y. Joncour, and A. Hirano, "A stereo echo canceler with correct echo-path identification based on an input-sliding technique," *IEEE Transactions on Signal Processing*, vol. 49, no. 11, pp. 2577–2587, November 2001.

[103] A. Sugiyama, Y. Mizuno, L. Kazdaghli, A. Hirano, , and K. Nakayama, "A stereo echo canceller with simultaneuos 2-channel input slides for fast convergence and good sound localization," in *Proceedings of the 17th European Signal Processing Conference*, Glasgow, Scotland, August 2009, pp. 1992–1996.

[104] M. Ali, "Stereophonic acoustic echo cancellation system using time-varying all-pass filtering for signal decorrelation," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Seatle, USA, May 1998, pp. 3689–3692.

[105] L. Romoli, S. Cecchi, L. Palestini, P. Peretti, and F. Piazza, "A novel approach to channel decorrelation for stereo acoustic echo cancellation based on missing fundamental theory," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Dallas, USA, March 2010, pp. 329–332.

[106] S. Cecchi, L. Romoli, P. Peretti, and F. Piazza, "A combined psychoacoustic approach for stereo acoustic echo cancellation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 6, pp. 1530–1539, August 2011.

[107] J. Blauert, *Spatial Hearing*, 2nd ed. Cambridge: MIT Press, 1983.

[108] ITU-T G.191, "Software tools for speech and audio coding standardization," International Telecommunications Union, Geneva, Switzerland 2010.

[109] ITU-R BS.1534-1, "Method for the subjective assessment of intermediate quality level of coding systems," International Telecommunications Union, Geneva, Switzerland 2003.

[110] F. Albu, J. Kadlec, N. Coleman, and A. Fagan, "The gauss-seidel fast affine projection algorithm," in *IEEE Workshop on Signal Processing Systems*, San Diego, USA, October 2002, pp. 109–114.