ANALYSIS OF ADAPTIVE FEEDBACK AND ECHO CANCELATION ALGORITHMS IN A GENERAL MULTIPLE-MICROPHONE AND SINGLE-LOUDSPEAKER SYSTEM

Meng Guo1,2, Thomas Bo Elmedyby1, Søren Holdt Jensen2, and Jesper Jensen1

1Oticon A/S, Denmark, 2Department of Electronic Systems, Aalborg University, Denmark
emails: {guo, tbl, jsj}@oticon.dk, shj@es.aau.dk

ABSTRACT

In this paper, we analyze a general multiple-microphone and single-loudspeaker system, where an adaptive algorithm is used to cancel acoustic feedback/echo and a beamformer processes the feedback/echo canceled signals. This system can be viewed as part of a typical hearing aid system and/or a traditional acoustic echo cancelation system. We introduce and derive an approximation of a useful frequency domain measure — the power transfer function — and show how to predict the system stability bound, convergence rate and the steady-state behavior across time and frequency. Furthermore, we show how the derived expressions can be used to determine e.g. the step size parameter in the adaptive algorithms to achieve a desired system property e.g. convergence rate at a specific frequency.

Index Terms— Adaptive filters, acoustic feedback cancelation, convergence rate, steady-state behavior, step size parameter.

1. INTRODUCTION

Generally, the problem of acoustic feedback/echo arises in audio systems where recording and playback occur simultaneously, e.g. in public address systems, audio conference systems and hearing aids. Many solutions have been proposed over the years, see e.g. [1, 2] and references therein. A widely used solution is an adaptive filter, which provides an approximation of the acoustic feedback/echo path, in a system identification configuration, see e.g. [3]. The terminologies feedback and echo are often used in closed-loop and open-loop systems, respectively. In this work, we refer to both as feedback.

Many existing studies analyze, characterize and improve adaptive algorithms in terms of robustness, stability bounds, convergence rate and steady-state behavior. Often, these studies focus on criteria mean-square error of the error signal, mean-square deviation of the estimation error vector and variants of these, see e.g. [3, 4]. Although they provide useful information about the behavior of the adaptive systems, they do not reveal frequency domain behavior, which is more prevailing in areas such as acoustic feedback cancelation, because electro-acoustic properties of feedback paths are easier described in the frequency domain, and frequency domain measures are more directly related to auditory perception. Some examples of frequency domain measures can be found in e.g. [5].

Fig. 1(a) shows a system with an acoustic feedback path represented by the transfer function (TF) \( H(\omega, n) \), where \( \omega \) and \( n \) denote the normalized frequency and time index, respectively. An estimate \( \hat{H}(\omega, n) \) is typically computed by means of an adaptive algorithm in order to cancel the effect of \( H(\omega, n) \). The TF \( F(\omega, n) \) denotes a forward path, which is found in applications with a closed-loop setup e.g. to implement amplification in a hearing aid. In the area of acoustic echo cancelation (AEC), \( F(\omega, n) \) represents a far-end TF and is often ignored, resulting in an open-loop setup.

Fig. 1(b) shows a system with acoustic feedback coupling and cancelation. (a) A basic system. (b) Multiple-microphone and single-loudspeaker system with a beamformer, where \( i = 1, \ldots, P \).

In both setups, the measure \( D = |H(\omega, n) - \hat{H}(\omega, n)|^2 \) provides useful frequency-wise information about the cancelation performance. Especially in closed-loop setups, \( D \) is directly related to system stability and the capability of amplification. In practice, however, \( D \) cannot be computed directly since \( H(\omega, n) \) is unknown.

This work is inspired by the study in [6] of tracking characteristics in terms of frequency domain mean-square errors (FDMSE) \( E[|H(\omega, n) - \hat{H}(\omega, n)|^2] \) for a single microphone and single-loudspeaker (SMSL) system. We generalize this study and work on a multiple-microphone and single-loudspeaker (MMSL) system, while a beamformer processes the feedback-compensated signals as shown in Fig. 1(b). This system is general and can be considered as part of a hearing aid as described in [7], where a binaural hearing aid system was studied, and/or it can be viewed as an AEC system.

Our goal is to derive simple expressions which describe stability bound, convergence rate and steady-state behavior as a function of frequency of the system in Fig. 1(b). Our derivations are based on an open-loop setup, i.e. \( F(\omega, n) \) is omitted in Fig. 1, which corresponds to a traditional AEC setup. However, preliminary simulation results have shown that the derived results provide accurate approximations in a closed-loop setup as well.

The derivations in this paper are based on the Least Mean Square (LMS) [3] as an example of an adaptation algorithm. With this choice, we derive expressions for the system stability bound, convergence rate and the steady-state behavior. Furthermore, we show how to choose the step size parameter in the adaptive algorithm, for a given desired convergence rate and/or steady-state behavior. Other adaptive algorithms can be analyzed using the same methodology.

In this paper, column vectors and matrices are emphasized using lower and upper letters in bold, respectively. Transposition, Hermitian transposition and complex conjugation are denoted by the superscripts \( T \), \( H \) and \( * \), respectively.

2. SYSTEM DESCRIPTION

The MMSL system under analysis has \( P \) microphones, a single loudspeaker and a beamformer as shown between points A and B in Fig.
Fig. 2. A detailed view of the system under analysis.

The feedback path from the loudspeaker to the i’th microphone is modeled by a finite impulse response (FIR) of order \( L - 1 \),

\[
h_i(n) = [h_i(0, n), \ldots, h_i(L - 1, n)]^T.
\]

The frequency response of \( h_i(n) \) is expressed by

\[
H_i(\omega, n) = \sum_{k=0}^{L-1} h_i(k, n)e^{-j\omega k}.
\]

We model the time variations of the true feedback paths using the random walk model

\[
H_i(\omega, n) = H_i(\omega, n - 1) + H_i(\omega, n),
\]

where \( H_i(\omega, n) \in \mathbb{C} \) is a sample from a zero-mean Gaussian white noise process with variance

\[
S_{h_i}(\omega) = E \left[ H_i(\omega, n)H_i^*(\omega, n) \right].
\]

In the time domain, the feedback path variation vector is given by

\[
h_i(n) = h_i(n) - h_i(n - 1).
\]

The covariance matrix of the i’th and j’th feedback path variations is defined as

\[
\tilde{H}_{ij} = E \left[ \tilde{h}_i(n)\tilde{h}_j^*(n) \right],
\]

and we assume, for simplicity, that \( \tilde{H}_{ij} = 0_{L \times L} \) for \( i \neq j \).

The estimated FIR feedback path of order \( L - 1 \) is given by

\[
\hat{h}_i(n) = [\hat{h}_i(0, n), \ldots, \hat{h}_i(L - 1, n)]^T,
\]

and the corresponding estimation error vector is defined as

\[
\tilde{h}_i(n) = h_i(n) - h_i(n),
\]

with a frequency response of

\[
\tilde{H}_i(\omega, n) = \sum_{k=0}^{L-1} \tilde{h}_i(k, n)e^{-j\omega k}.
\]

The time invariant FIR filter \( g_i \) of order \( N - 1 \) is expressed by

\[
g_i = [g_i(0), \ldots, g_i(N - 1)]^T,
\]

and its frequency response is expressed by

\[
G_i(\omega) = \sum_{k=0}^{N-1} g_i(k)e^{-j\omega k}.
\]

The loudspeaker signal vector is defined as

\[
u(n) = [u(n), \ldots, u(n - L + 1)]^T.
\]

We follow the approach in [6] and consider the loudspeaker signal \( u(n) \) as a deterministic signal, because it simplifies the analysis. However, our results remain valid, even if the loudspeaker signal \( u(n) \) is considered as a realization of a stochastic process, which is statistically independent of the incoming signals \( x_i(n) \); this important point will be demonstrated by simulations in Sec. 6. Therefore, we define the matrix

\[
R_u(k) = \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} u(n)u^T(n - k).
\]

3. POWER TRANSFER FUNCTION

Considering Fig. 2 in the frequency domain, the power transfer function (PTF) denoted by \( \xi(\omega, n) \) is defined as

\[
\xi(\omega, n) = E \left[ \sum_{i=1}^{P} G_i(\omega)\tilde{H}_i(\omega, n) \right]^2
\]

\[
= \sum_{i=1}^{P} \sum_{j=1}^{P} G_i(\omega)G_j^*(\omega)\xi_{ij}(\omega, n),
\]

where \( \xi_{ij}(\omega, n) = E[\tilde{H}_i(\omega, n)\tilde{H}_j^*(\omega, n)] \). The PTF \( \xi(\omega, n) \) is a frequency domain measure which describes the expected magnitude square of the TF between points A and B in Fig. 2. Hence, \( \xi(\omega, n) \) is an MMSE system equivalent to the FDMSE \( E[|\tilde{H}(\omega, n)|^2] \) for the SDSL system shown in Fig. 1(a), and as we will show, it is directly linked to the system stability.

It is possible to express Eq. (18) without knowing \( \tilde{H}_i(\omega, n) \), but this expression is very complicated. In this work, we aim to derive a simpler, but accurate approximation \( \hat{\xi}(\omega, n) \) of \( \xi(\omega, n) \).

We introduce the notation

\[
\hat{\xi}_{ij}(\omega, n) \approx E \left[ \tilde{H}_i(\omega, n)\tilde{H}_j^*(\omega, n) \right].
\]

An approximation of the PTF \( \hat{\xi}(\omega, n) \) in Eq. (18) is then given by

\[
\hat{\xi}(\omega, n) = \sum_{i=1}^{P} \sum_{j=1}^{P} G_i(\omega)G_j^*(\omega)\hat{\xi}_{ij}(\omega, n).
\]
4. SYSTEM ANALYSIS

We derive asymptotic expressions for \( \hat{\xi}(\omega, n) \) as the FIR order \( L \to \infty \) and the step size \( \mu(n) \to 0 \). It turns out that the resulting expressions for \( \hat{\xi}(\omega, n) \) are accurate for practically usable values of \( L \) and \( \mu(n) \) as demonstrated in Sec. 6.

The LMS update of the \( i \)th microphone channel is, see e.g. [3],

\[
\hat{h}_i(n) = \hat{h}_i(n-1) + \mu(n)u(n)e_i(n).
\]  

(21)

Using Eqs. (21), (16), (15) and (5), the estimation error vector defined in Eq. (8) can be expressed by

\[
\hat{e}_i(n) = \left( I - \mu(n)u(n)u^T(n) \right) \hat{h}_i(n-1) + \mu(n)u(n)x_i(n) - \hat{h}_i(n).
\]  

(22)

where \( I \) is the identity matrix.

Now, it can be shown using Eq. (22), under the assumption of \( \mu(n) \to 0 \) and \( r_{x_i}(k) = 0 \) for \( |k| \geq k_0 \in \mathbb{N} \), that the estimation error covariance matrix \( \hat{\Sigma}_{ij}(n) \approx \mathbb{E}[\hat{h}_i(n)\hat{h}_j^T(n)] \) is expressed by

\[
\hat{\Sigma}_{ij}(n) \approx \hat{\Sigma}_{ij}(n-1) - \mu(n)\hat{\Sigma}_{ij}(n-1)R_u(0) + \hat{\Sigma}_{ij} - \mu^2(n) \sum_{k=k_0}^{k_0} R_u(k)r_{x_i}(k).
\]  

(23)

We wish to study the entries of \( \hat{\Sigma}_{ij}(n) \) in the frequency domain. Therefore, let \( \mathbf{F} \in \mathbb{C}^{L \times L} \) be a DFT matrix and define the matrix \( \hat{\Xi}_{ij}(n) = \mathbf{F}^H\hat{\Sigma}_{ij}(n)\mathbf{F} \) which can be expressed by

\[
\hat{\Xi}_{ij}(n) \approx \mathbf{F}\hat{\Sigma}_{ij}(n-1)\mathbf{F}^H
\]

\[- \mu(n)\frac{1}{L}\mathbf{F}\hat{R}_u(0)\mathbf{F}^H\hat{\Sigma}_{ij}(n-1)\mathbf{F}^H
\]

\[- \mu(n)\frac{1}{L}\mathbf{F}\hat{\Sigma}_{ij}(n-1)\mathbf{F}^H\hat{R}_u(0)\mathbf{F}^H
\]

\[+ \mathbf{F}\hat{\Sigma}_{ij}\mathbf{F}^H + \mu^2(n) \sum_{k=k_0}^{k_0} \mathbf{F}R_u(k)\mathbf{F}^Hr_{x_i}(k).
\]  

(24)

Asymptotically, as \( L \to \infty \), each term in Eq. (24) approaches a diagonal matrix due to the (asymptotical) Toeplitz structures of \( \hat{\Sigma}_{ij}(n-1) \) and \( \hat{\Sigma}_{ij} \) [8]. By looking at each diagonal element in \( \hat{\Xi}_{ij}(n) \), we get the frequency domain measure \( \hat{\xi}_{ij}(\omega, n) \) defined in Eq. (19) as

\[
\hat{\xi}_{ij}(\omega, n) \approx (1 - 2 \mu(n)S_u(\omega))\hat{\xi}_{ij}(\omega, n-1) + L\mu^2(n)S_u(\omega)S_{x_{ij}}(\omega) + S_{h_{ij}}(\omega),
\]  

(25)

where \( S_u(\omega) \) denotes the power spectrum density (PSD) of the loudspeaker signal \( u(n) \), and \( S_{x_{ij}}(\omega) \) denotes the cross(auto) PSDs of the incoming signals \( x_i(n) \) and \( x_j(n) \).

Finally, inserting Eq. (25) in Eq. (20), we obtain

\[
\hat{\xi}(\omega, n) = (1 - 2 \mu(n)S_u(\omega))\hat{\xi}(\omega, n-1) + L\mu^2(n)S_u(\omega)\sum_{i=1}^{P} G_i(\omega)G^*_j(\omega)S_{x_{ij}}(\omega)
\]

\[+ \sum_{i=1}^{P} |G_i(\omega)|^2 S_{h_{ij}}(\omega).
\]  

(26)

This result can be easily adapted to be valid for the Normalized Least Mean Square (NLMS) algorithm, simply by substituting the step size \( \mu(n) \) with \( \mu_0(n) = L\mu^2(n) \), where \( \sigma^2_u \) is the variance of loudspeaker signal \( u(n) \). Furthermore, the methodology described in this section can be used to derive similar expressions for the Recursive Least Squares (RLS) algorithm.

5. INTERPRETATION

In this section, we use Eq. (26) to derive the system stability bound, convergence rate describing the decay of \( \hat{\xi}(\omega, n) \) and the steady-state value describing \( \hat{\xi}(\omega, n) \) when the adaptation has converged.

Eq. (26) is a first order difference equation in \( \hat{\xi}(\omega, n) \) expressed by the TF \( H(z) = \frac{\beta}{1 - \alpha z^{-1}} \); the coefficient \( \alpha \) determines the pole location in \( H(z) \) and thereby its convergence rate and stability. For Eq. (26), \( \alpha \) is expressed by

\[
\alpha = 1 - 2 \mu(n)S_u(\omega).
\]  

(27)

Using Eq. (27), the maximum step size \( \mu(n) \) to maintain system stability (where \( |\alpha| < 1 \)) is

\[
\mu(n) < \frac{1}{\max_{\omega} S_u(\omega)}.
\]  

(28)

The convergence rate (CR[dB/iteration]) is calculated, based on the impulse response \( h(n) = \beta \cdot \alpha^n \) of \( H(z) = \frac{\beta}{1 - \alpha z^{-1}} \), as

\[
\text{CR[dB/iteration]} = 10 \log_{10} \alpha.
\]  

(29)

Thus, the convergence rate depends on the step size \( \mu(n) \) and the PSD \( S_u(\omega) \) of the loudspeaker signal \( u(n) \), but surprisingly not on the incoming signals \( x_i(n) \); similar results for an SMSS system are achieved in [6].

Furthermore, using Eq. (26), the steady-state value of \( \hat{\xi}(\omega, \infty) = \lim_{n \to \infty} \hat{\xi}(\omega, n) \) is derived as

\[
\hat{\xi}(\omega, \infty) = \lim_{n \to \infty} \frac{\mu(n)}{2} \sum_{i=1}^{P} \sum_{j=1}^{P} G_i(\omega)G^*_j(\omega)S_{x_{ij}}(\omega)
\]

\[+ \lim_{n \to \infty} \sum_{i=1}^{P} |G_i(\omega)|^2 S_{h_{ij}}(\omega).
\]  

(30)

The first term in Eq. (30) is the steady-state value \( \hat{\xi}(\omega, \infty) \) with time invariant feedback paths; the second term is the extra error contribution as result of variations in the feedback paths.

It can be seen from the first term of Eq. (30), that \( \hat{\xi}(\omega, \infty) \) in a time invariant system is a function of incoming signals \( x_i(n) \), but not the loudspeaker signal \( u(n) \). Furthermore, \( \hat{\xi}(\omega, \infty) \) is linearly proportional to \( L \) and \( \mu(n) \).

On the other hand, it is seen from the second term, that a larger step size \( \mu(n) \) and higher PSD \( S_u(\omega) \) lead to smaller additional error when the system is undergoing variations. This is explained by the fact that a larger \( \mu(n) \) and higher \( S_u(\omega) \) make the system faster in tracking a time varying feedback path.

Hence, the overall steady-state value \( \hat{\xi}(\omega, \infty) \) is a compromise between the steady-state and tracking behavior in situations with time invariant and time varying feedback paths, respectively.

Furthermore, it is observed that \( \hat{\xi}(\omega, n) \) is independent of the true feedback path responses \( H_i(\omega, n) \). Another interesting observation is that the frequency responses \( G_i(\omega) \) and \( G_j(\omega) \) of the beamformer filters act as weighting coefficients in the expression for the
steady-state value $\hat{\xi}(\omega, \infty)$ in Eq. (30), but they have no influence on the convergence rate according to Eq. (27).

Using Eq. (29), a desired convergence rate can be achieved by choosing the step size according to

$$\mu(n) = \frac{1}{2} \frac{10^{\left(\text{Convergence Rate} / 10\right)}}{S_{\xi}(\omega)} - 1,$$  

and using Eq. (30), a desired steady-state value $\hat{\xi}(\omega, \infty)$, ignoring the feedback path variations for simplicity, can be achieved by choosing the step size according to

$$\lim_{n \to \infty} \mu(n) = \frac{\sum_{i=1}^{L} G_i(\omega) G_i^*(\omega) S_{\xi}(\omega)}{2 \sum_{i=1}^{L} G_i(\omega) G_i^*(\omega) S_{\xi}(\omega)}.$$  

6. EXPERIMENTS

In this section, we conduct two simulation experiments, based on an MMSL system with $P = 3$ microphones, to verify Eqs. (29)-(30) and (31)-(32), respectively. Unfortunately, due to space limitations, most parameter values used in the simulations are omitted.

In both experiments, the beamformer filters $g_i$ are fixed first order FIR filters, and each true feedback path $h_i(n)$ is modeled by a first order FIR filter and thereby known. Thus, in the simulations, it is possible to compute the true PTF $\xi(\omega, n)$ given by Eq. (18) using Eqs. (8), (9) and (11). This calculation is used to verify the predicted results. In both experiments, the duration is $10^4$ iterations for each simulation run, and 100 simulation runs are performed to obtain an averaged $\xi(\omega, n)$. In each simulation run, new realizations of a standard Gaussian white noise sequence are drawn. They are then filtered by various fixed first order FIR shaping filters to generate the loudspeaker signal $u(n)$ and the incoming signals $x_i(n)$. The initial feedback path estimates are $\hat{h}_i(0) = 0_{L \times 1}$, where $L = 32$.

In the first experiment, the true feedback paths are fixed during the first half of the simulations, whereas random walk variations are added during the second half. A fixed step size $\mu(n) = 2^{-9}$ is used. The simulated and predicted results are shown in Fig. 3(a), at a representative example frequency $\omega = 2\pi L / L$, where $L = 7$. Clearly, the simulation results agree with the predicted convergence rate and steady-state values obtained using Eqs. (29)-(30). Furthermore, despite the underlying asymptotic assumptions of $\mu(n) \to 0$ in the analysis, we observe that the derived expressions are accurate for practical values as $L = 32$ and $\mu(n) = 2^{-9}$.

The fact that a new realization of $u(n)$ is drawn for each simulation run demonstrates, as expected, that the derived results are valid, when the loudspeaker signal $u(n)$ is a realization of a stochastic process, even though we considered $u(n)$ as a deterministic signal in the analysis.

In the second experiment, we calculated, using Eqs. (31)-(32), the step size $\mu(n)$ to achieve a desired convergence rate of $-0.005$ dB/iteration and a steady-state value of $-6$ dB, respectively. The settings are the same as in the first experiment, except that the feedback path remains fixed, and the step size $\mu(n)$ has been changed. The simulated and predicted results at the frequency bin $l = 7$ are shown in Fig. 3(b)-(c). Again, simulation results support the theory.

7. CONCLUSIONS

In this work, we derived expressions to predict the system stability bound, convergence rate and the steady-state behavior in terms of the power transfer function for each frequency in a general multiple microphone and single-loudspeaker system with a beamformer. We showed that these derived expressions can be used to control system parameters e.g. the step size parameter in adaptive cancelation algorithms in order to achieve desired properties such as convergence rate and steady-state behavior at a specific frequency.

The results are successfully verified by simulations.

8. REFERENCES


