LOW-COMPLEXITY ADAPTIVE FILTERING FOR ACOUSTIC ECHO CANCELLATION IN AUDIO CONFERENCING SYSTEMS

Christian Schüldt

Blekinge Institute of Technology
Licentiate Dissertation Series No. 2009:01

School of Engineering

Low-Complexity Adaptive Filtering for Acoustic Echo Cancellation in Audio Conferencing Systems

Christian Schüldt

Blekinge Institute of Technology Licentiate Dissertation Series
No 2009:01
ISSN 1650-2140
ISBN 978-91-7295-153-2

Low-Complexity Adaptive Filtering for Acoustic Echo Cancellation in Audio Conferencing Systems

Christian Schüldt

Department of Signal Processing
School of Engineering
Blekinge Institute of Technology
SWEDEN

Department of Signal Processing
School of Engineering
Publisher: Blekinge Institute of Technology
Printed by Printfabriken, Karlskrona, Sweden 2009
ISBN 978-91-7295-153-2

Preface

This licentiate thesis marks an important milestone in my applied signal processing research. The work has been carried out at the Department of Signal Processing at Blekinge Institute of Technology in Ronneby, Konftel AB in Umeå, and Limes Technology AB in Umeå and Stockholm as a collaboration between academia and industry.

Working close to the industry, focusing on problems with real life significance, has both been extremely interesting and a strong motivation.

Christian Schüldt
Stockholm, October 2008

Acknowledgments

If I were to thank only one person, it would without a doubt be Dr. Fredric Lindström, without whom I would not be where I am today. His drive and inspiration has helped me immensely to get motivation at times when everything seemed hopeless.

I thank Prof. Ingvar Claesson, who is the other reason for me being where I am today, for providing guidance as well as great scientific insights.

I would also like to thank my colleagues at Blekinge Institute of Technology, Konftel AB and Limes Technology AB for their support.

Finally, I would like to thank my girlfriend Linn Ristborg for all the love and support throughout the years.

Christian Schüldt
Umeå, October 2008

Contents

Publication list

Part I has been published as:

F. Lindstrom, C. Schüldt and I. Claesson, “Efficient Multichannel NLMS Implementation for Acoustic Echo Cancellation”, EURASIP Journal on Audio, Speech, and Music Processing, Volume 2007, Article ID 78439, 6 pages, `doi:10.1155/2007/78439`, 2007.

Part II has been published as:

C. Schüldt, F. Lindstrom and I. Claesson, “Low-Complexity Adaptive Filtering Implementation for Acoustic Echo Cancellation”, In Proceedings of IEEE TENCON, Hong Kong, November 2006.

Part III has been published as:

C. Schüldt, F. Lindstrom and I. Claesson, “A Low-Complexity Delayless Selective Subband Adaptive Filtering Algorithm”, IEEE Transactions on Signal Processing, vol. 56, no. 12, pp. 5840-5850, December 2008.

Part IV has been published as:

F. Lindstrom, C. Schüldt, M. Dahl and I. Claesson, “Improving the Performance of a Low-Complexity Doubletalk Detector by a Subband Approach”, Proceedings of IEEE International Conference on Signals, Systems and Devices, vol. III, Sousse, Tunisia, March 2005.

Other publications in conjunction with the thesis

C. Schüldt, F. Lindstrom and I. Claesson "A Distortion Reducing Subband Limiter Implementation for Conference Phones", In Proceedings of IEEE International Conference on Consumer Electronics, Las Vegas, NV, January 2008.

F. Lindstrom, C. Schüldt, M. Långström and I. Claesson, "A Method for Reduced Finite Precision Effects in Parallel Filtering Echo Cancellation", IEEE Transactions on Circuits and Systems Part I: Regular Papers, vol. 54, pp. 2011-2018, September 2007.

F. Lindstrom, C. Schüldt and I. Claesson, "An Improvement of the Two-Path Algorithm Transfer Logic for Acoustic Echo Cancellation", IEEE Transactions on Audio, Speech and Language Signal Processing, vol. 15, pp. 1320-1326, May 2007.

F. Lindstrom, C. Schüldt and I. Claesson, "A Hybrid Acoustic Echo Canceller and Suppressor", Signal Processing, vol. 87, pp. 739-749, April 2007.

C. Schüldt, F. Lindstrom and I. Claesson "A Combined Implementation of Echo Suppression, Noise Reduction and Comfort Noise in Speaker Phone Application", In Proceedings of IEEE International Conference on Consumer Electronics, Las Vegas, NV, January 2007.

F. Lindstrom, C. Schüldt and I. Claesson, "Reusing Data During Speech Pauses in an NLMS-based Acoustic Echo Canceller", Proceedings of IEEE TENCON, Hong Kong, November 2006.

Introduction

With the invention of the telephone in late 1800 by Antonio Meucci (and the commercialization by Alexander Graham Bell), the speech form of communication was no longer limited to a close distance, but could literally be achieved around the world given a telephone wire connection. In the early 1900, loudspeaker telephones primarily intended for managers in an office environment were introduced, eliminating the need of a handset and thus allowing the user to have his or her hands free while communicating [1, 2]. Initial problems were echoes and self-reverberation (so-called *howling*). The introduction of the transistor in the 1950s gave birth to a new generation of loudspeaker telephones using automatic voice control, i.e. switching between the incoming and outgoing speech to prevent echoes and howling. Advancements in adaptive filter theory [3, 4] and digital technology allowed real time echo cancellation, which in turn led to *full-duplex* conference phones where speech signals can flow in both directions simultaneously (as opposed to the *half-duplex* switching loudspeaker telephones).

The market for conference phones and audio conferencing systems has shown tremendous growth in the past few years and is expected to continue to grow in the near future. Economic savings, as well as reduced environmental impacts by avoiding travel, i.e. avoiding pollution contributing to green house effects, are the key factors explaining this development. Further, the market for cheap desktop products previously dominated by cheap half-duplex loudspeaker telephones, has during the last years seen the introduction of low-cost echo cancelling full-duplex solutions.

The importance of audio quality in a conference system

It is often said that a majority of communication is non-verbal, although this statement is possibly a misinterpretation of an early study concluding that body language accounts for approximately 55% of the *liking* for the person who puts forward the message in a face-to-face communication [5]. Obviously, body language mainly mediates emotions and feelings, while spoken communication constitutes both nonverbal cues such as tone, pitch and accent, and words (essentially the basis of communication). If forced to choose between an audio-only conference and a video-only conference, a vast majority would probably choose the former.

Hence, the speech quality is of great importance in any audio-based conference system. In this context the speech quality can be considered with respect to mainly intelligibility and listener comfort. Intelligibility is the degree to which speech can be understood, while listener comfort basically is the degree of comfort experienced by the listener. Poor intelligibility also causes poor listener comfort, since the listener will be forced to concentrate and perhaps has to strain to understand.

However, since speech is fairly resistant to many types of frequency cut-offs (and/or maskings) [6], the speech intelligibility can be maintained, or even increased, in some cases, by simple processing of the signal. For example, an average living room typically has longer reverberation time (i.e. the time it takes for the built-up echo to decay by 60 dB below the level of the direct sound) for lower frequencies than for higher frequencies [7] due to the sound absorbtion properties of the furniture, walls and the air. By removing low frequencies of the microphone signal, the speech intelligibility could be increased, but generally at the cost of reduced listener comfort, since the sound will be perceived as sharper and more tiresome.

If the speech signal is corrupted by (residual) echo, a time-varying attenuator may be used to remove or lessen the echo. By designing the attenuator to operate heavily on the signal, a large part of the echo is reduced but also a large part of the desired speech, resulting in poor intelligibility. On the other hand, with an attenuator set to only slightly damp the signal (or in a situation completely without an attenuator), the speech intelligibility will probably be higher, but the listening comfort will be reduced due to the echo.

Echos and a simple model of room acoustics

Two types of echoes are predominant in an teleconferencing environment, line echo and acoustic echo. Line echo originates from the 2-4 wire conversion in hybrid circuits in the telephone network [4], while acoustic echoes arise due to acoustic coupling between loudspeaker and microphone. The most significant differences between line echo and acoustic echo are that the amount of returning line echo is limited by regulations and recommendations [8] and that the transfer function of the line echo is sparse, while the acoustic echo typically has a non-sparse transfer function with an exponentially decaying envelope [9]. Since this thesis is mainly focused on the cancellation of acoustic echoes, a more detailed description of this phenomenon is described below.

Sound waves emitted by the loudspeaker propagate in the room and are attenuated and reflected by the air itself as well as walls, furniture and objects, see figure 1. By using these properties, it is possible to form a simple model of the room acoustics.

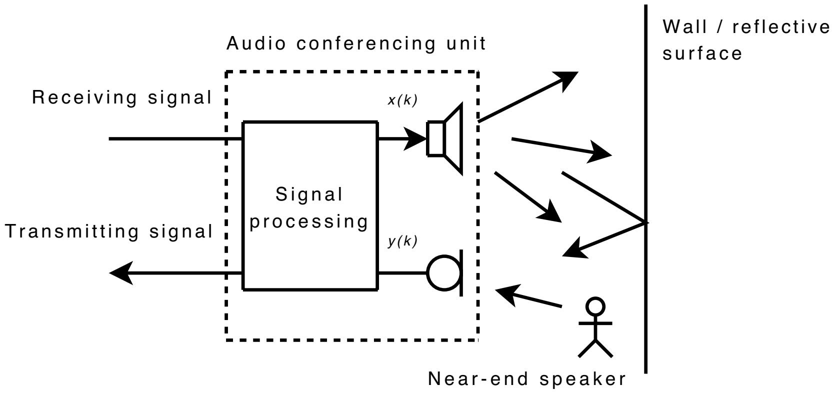

Figure 1: Scheme of a audio conferencing setup.

First, consider a digital signal $x(k)$ sampled at 8 kHz, where k is the sample index, passed through a digital-to-analog converter and then to a loudspeaker. A simple linear model of the acoustic echo present on the microphone signal after analog-to-digital conversion $y(k)$ can then be formed as a sum of more or less attenuated and time-delayed versions of the loudspeaker signal $x(k)$ as

$$\sum_{i=0}^{\infty} h_i x(k-i), \tag{1}$$

where h_i denotes the attenuation of the loudspeaker signal as received on the microphone after i samples. Plotting estimates of h_i measured in a room against the parameter i typically gives a result similar to what is shown in figure 2, also called *impulse response*. What can be seen is first the intrinsic delay in the system, resulting in the amplitude being approximately zero for the first 100 samples in figure 2. This is followed by a few samples of large magnitude representing sound traveling straight from the loudspeaker to the microphone without, or with just a few, reflections. Then as i increases the sound reaching the microphone is more and more attenuated.

Acoustic Echo Cancellation

The method of acoustic echo cancellation uses the linear model of the echo described equation (1) to create a replica of the expected echo

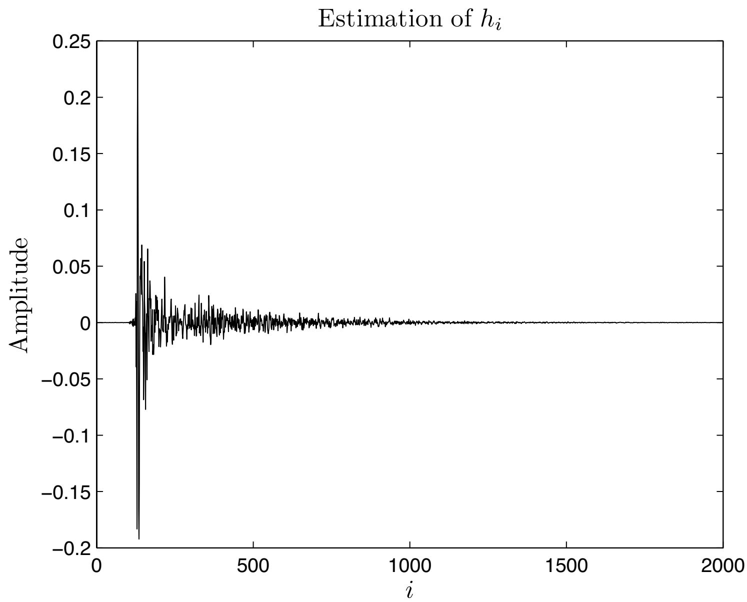

Figure 2: Estimated impulse response (transfer function) of a typical room.

$$\hat{d}(k) = \sum_{i=0}^{N-1} \hat{h}_i x(k-i), \tag{2}$$

where N is the model order and $\hat{h}_i$ is an estimate of h_i. This echo replica is then subtracted from the microphone signal to form an echo-cancelled microphone signal

$$e(k) = y(k) - \hat{d}(k). \tag{3}$$

The benefit of this, instead of just using a variable attenuator to damp the echo, is that the echo cancellation only removes the echo, leaving near-end speech unaffected in a *doubletalk* situation (i.e. when both parties of the conversation are speaking simultaneously) and thus, at least in theory, allowing full-duplex. In practice however, a level of 20-30 dB acoustic echo cancellation is achievable [9], implying a need of both echo cancellation and attenuation.

What is needed to perform the echo cancellation as described is to decide an appropriate model order (adaptive filter length) N and to estimate $\hat{h}_i \quad i \in \{0, \cdots, N\}$, i.e. the adaptive filter coefficients. If the filter length is too short the filter will obviously not be able to model the full echo path of the setup, resulting in poor cancellation performance. On the other hand, a too long filter uses an unnecessary amount of memory and computational resources, which could perhaps be better used for something else. It is also well known that a long filter converges slower than a short one [10]. In practice, N is often set as large as allowed by the given memory and computational resources, or set adaptively using a variable filter-length algorithm [11, 12].

Adaptation of the filter coefficients can be performed through a number of methods, whereof the normalized least mean square (NLMS) [13, 14, 15, 16] is one of the most popular owing to its ease of implementation, low computational complexity and robustness to fix-point implementation issues. From a geometric perspective, the updating of the adaptive filter can be seen as moving from one point to another in an N-dimensional space. In the case of the NLMS, a filter update constitutes a movement along the *regressor vector* $\mathbf{x}(k) = [x(k), x(k-1), \cdots, x(k-N+1)]^T$, where $[\cdot]^T$ denotes transpose. Moreover, in the NLMS updating case each update is independent, meaning that movement in the N-dimensional space is far from optimal, especially for highly colored input signals where the regressor vectors used for different updates are almost parallel. A more efficient adaptive filtering method is recursive least squares (RLS) [16], which minimizes a weighted sum of the square of all output errors, as opposed to the NLMS which minimizes the expected value of the current error. In a sense, the RLS depends on the signals themselves, whereas the NLMS depends on their statistics. The RLS provides a much faster convergence rate than the NLMS, but at the cost of much higher computational complexity and sensitivity to round off errors occurring in fixed-point implementations. An intermediate solution, in terms of both convergence speed and computational complexity, is the affine projection (AP) algorithm [17]. A fast implementation of the AP algorithm called fast affine projection [18] has also been presented, reducing the computational complexity almost to that of the NLMS, except for a matrix inversion.

Computational complexity reduction

A technique for further reduction of computational complexity is subband adaptive filtering [9], where the signals are passed through a filterbank employing downsampling. An adaptive filter is used in each subband. The key factor for complexity reduction in this case is the downsampling, resulting in shorter filters not updating as often as their traditional long fullband counterpart. Moreover, decimation stretches the signal in frequency, making it more flat ("white") which in turn results in faster filter convergence. The major downside of straight-forward subband adaptive filtering is the delay introduced by the analysis and synthesis filterbanks [9]. A low delay is important due to the dependence between echo delay and an acceptable level of echo attenuation [19], where a longer echo delay requires more echo attenuation for a maintained level of acceptance. Further, if the total delay of the communication setup is long, speech is not able to flow naturally as the speakers have to wait longer for the other party to respond. A solution for this issue is delayless subband adaptive filtering [20, 21], where the adaptive subband filters at regular intervals are merged together to form a fullband filter which in turn produces a delayless echo cancelled output.

Another method for computational complexity reduction is partial- or selective updating, where only a subsection of the adaptive filter is updated at each instant. The most basic approach is *periodic updating* [22], where the updating of the adaptive filter is restricted to every Mth sample. To avoid significant computational complexity variations in time, block processing can be performed, unfortunately causing a delay in the signal path. A similar approach is *partial updating*, where only a part of all N filter coefficients are updated at each instant. Several methods for choosing which coefficients to update at a specific instant have been proposed. Sequential updating [22] is perhaps the most rudimentary, where the selection of coefficients to be updated are performed in a sequential manner. The stability properties of sequential- and periodic updating are slightly different, although the convergence rate of the two methods is similar [22]. Randomizing the selection process has also been proposed [23, 24], aiding some of the stability issues of sequential updating. Other partial updating methods are the M-MAX NLMS [25, 26, 27], where only the filter coefficients associated with the M largest amplitudes are updated, and an approach for omitting updates when the step size is zero or very small for a significant number of possible consecutive updates [28]. However, due to the nature of the latter method, the reduction of computational complexity cannot be accounted for a priori. For finding the largest M magnitude values in the M-MAX NLMS, a running order algorithm called *sortline* [29] is usually used. However, the overall complexity in terms of processor cycles exceeds the one of the full update NLMS algorithms on most signal processors [9]. To overcome the complexity problem, an approach which operates on subblocks of the regressor vector instead of individual entries has been proposed [30].

In a situation where both parties in the conversation are active simultaneously (i.e. during doubletalk), the microphone signal contains both echo (originating from the *far-end* talker, whose speech is being fed to the loudspeaker) and speech from the *near-end* talker in the room. The speech from the near-end talker disturbs the convergence of the acoustic echo canceller and can even lead to divergence. To avoid this from happening, a *doubletalk detector* [9] is typically used to detect situations with near-end speech and then halt the updating process of the adaptive filter.

The most basic doubletalk detector is the Geigel detector [31], which compares the loudspeaker and microphone energies. If the energy picked up by the microphone is larger than the energy going out to the loudspeaker, the extra energy from the microphone must come from a near-end talker in the room, hence a doubletalk situation is detected. Other, more recent approaches to the doubletalk detection problem have been using e.g. power comparison using cepstral techniques [32] and coherence and cross-correlation-based approaches [33, 34].

It is of outmost importance that the doubletalk detector functions as intended in order to achieve high audio quality. If the doubletalk detector is configured to be too sensitive, halting of the filter adaptation could occur in situations where the acoustic environment changes abruptly (e.g. movement of the speaker or microphone), i.e. in situations where adaptation is most needed. However, if the doubletalk detector is not sensitive enough, near-end speech might not be detected in some situations which could lead to poor cancellation performance and possibly even to divergence of the adaptive filter.

Thesis summary

This licentiate thesis focuses on complexity reduction methods for adaptive filtering and doubletalk detection. The thesis is divided into four parts: Part I describes a method for reducing the computational complexity in a multi-microphone audio conferencing system with one adaptive filter for each microphone. The complexity is reduced by only updating one adaptive filter at each instant, and the adaptive filter chosen for an update is the one corresponding to the largest output error magnitude. This idea is extended in Part II to a single microphone system using a block-based approach, and in Part III to a subband-based approach where only one subband filter is updated at each instant. Part III also provides further theoretical analysis of the proposed method. Part IV presents a framework for subband-based low-complexity doubletalk detectors.

Part I — Effcient Multichannel NLMS Implementation for Acoustic Echo Cancellation

Acoustic Echo cancellation in a situation with a conference phone having a single loudspeaker and several microphones can be modeled as a single-input-multiple-output (SIMO) system with one adaptive filter for each microphone. To reduce the computational complexity of such a system, it is possible to only update one adaptive filter at each instant. This paper proposes a method for selecting which filter to update, and simulations show the benefits of the proposed method over other selection methods. One of the most interesting properties of the proposed method is its ability to select the most misadjusted filter for updating, leading to excellent performance in a situation where one microphone is subjected to an echo path change.

Part II — Low-Complexity Adaptive Filtering Implementation for Acoustic Echo Cancellation

This paper presents a block-based approach for reduced complexity of both NLMS- and FAP-type adaptive filter algorithms. The idea is to collect a number of output errors corresponding to a set of regressor vectors, and then to update the adaptive filter with the regressor vector corresponding to the largest output error magnitude. Simulations and comparisons with the M-MAX, a periodic and a random updating scheme show the advantages of the proposed method.

Part III — A Low-Complexity Delayless Selective Subband Adaptive Filtering Algorithm

Subband adaptive filtering is a method providing increased convergence speed, better robustness in case of narrowband signals and reduced complexity as compared to traditional fullband adaptive filtering. However, the downside of subband methods is the signal delay introduced by the filterbanks. A solution to this problem is *delayless* subband adaptive filtering, where the individual subband adaptive filters are used to construct a fullband filter providing a delayless output. The downside is that the computational cost of constructing the fullband filter is substantial. For reducing the computational cost, this paper presents a procedure where only one adaptive subband filter is updated at each instant. This by itself results in lower computational cost, but also allows modification of the fullband filter construction, resulting in further reduction of computational complexity.

Part IV — Improving the Performance of a Low-Complexity Doubletalk Detector by a Subband Approach

Detecting doubletalk is essential in speech-based communications systems employing echo cancellation. This paper presents a framework for a class of low-complexity doubletalk detectors implemented in a subband environment. The individual detector outputs are modified using weighting and threshold functions and are then combined using different norms. Simulations and comparison with the classic Geigel detector show that significant improvement can be obtained by using the subband approach.

Bibliography

[1] K. V. Tahvanainen, "A revolutionary speaker phone", *The history of Ericsson*, `http://www.ericssonhistory.com/templates/Ericsson/Article.aspx?id=2095&ArticleID=1369&CatID=360&epslanguage=EN`, Accessed 15th May 2008.

[2] W. F. Clemency, F. F. Romanow, A. F. Rose, "The Bell System Speakerphone", *AIEE Transactions*, vol. 76(I), pp. 148-153, 1957.

[3] B. Widrow, M. E. Hoff. "Adaptive switching circuits". *In IRE WESCON Convention Record*, vol 4, pp. 96-104, 1960.

[4] M. M. Sondhi, "An adaptive echo canceler", *Bell Syst. Tech. J.*, vol. 46, pp. 497-510, March 1967.

[5] A. Mehrabian, *Silent messages: Implicit communication of emotions and attitudes*, 2nd edition, Wadsworth, Belmont, California, 1981.

[6] B. C. J. Moore, *An Introduction to the Psychology of Hearing*, 5th edition, Elsevier Academic Press, 2004.

[7] G. M. Jackson, H. G. Leventhall, "The Acoustics of Domestic Rooms", *Applied Acoustics*, vol. 5, pp. 265-277, 1972.

[8] "G.168 Digital network echo cancellers", *ITU-T Recommendation*, ITU-T, 2002.

[9] E. Hänsler, G. Schmidt, *Acoustic Echo and Noise Control: A Practical Approach*, Wiley, 2004.

[10] B. Widrow, S. D. Stearns, *Adaptive Signal Processing*, Prentice-Hall, 1985.

[11] T. Usagawa, H. Matsuo, Y. Morita, M. Ebata, "A new adaptive algorithm focused on the convergence characteristics by colored input signal: Variable Tap Length LMS", *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. EA75-A, no. 11, pp. 1493-1499, 1992.

[12] Y. Gong, C. F. N. Cowan, "An LMS style variable tap-length algorithm for structure adaptation", *IEEE Transactions on Signal Processing*, vol. 53, no. 7, pp. 2400-2407, 2005.

[13] J. I. Nagumo, A. Noda, "A learning method for system identification", *IEEE Transactions on Automatic Control*, vol. AC-12, pp. 282-287, 1967.

[14] A. E. Albert, L. A. Gardner, *Stochastic Approximation and Nonlinear Regression*, MIT Press, Cambridge, MA, 1967.

[15] R. R. Bitmead, B. D. O. Anderson, "Performance of adaptive estimation algorithms in dependent random environments", *IEEE Transactions on Automatic Control*, vol. AC-25, pp. 788-794, 1980.

[16] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, 4th edition, 2002.

[17] K. Ozeki, T. Umeda, "An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties", *Electronics and Communication in Japan*, vol. 67-A, pp. 126-132, 1984.

[18] S. L. Gay, S. Tavathia, "The fast affine projection algorithm", *In Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, pp. 3023-3026, May 1995.

[19] K. Shenoi, *Digital Signal Processing in Telecommunications*, Prentice-Hall, 1995.

[20] D. R. Morgan, J. C. Thi, "A delayless subband adaptive filter architecture", *IEEE Transactions on Signal Processing*, vol. 43, no. 8, pp. 1819-1830, 1995.

[21] J. Huo, S. Nordholm and Z. Zang, "New weight transform schemes for delayless subband adaptive filtering", *In Proceedings of Global Telecommunications Conference*, vol. 1, pp. 197-201, 2001.

[22] S. S. Douglas, "Adaptive filters employing partial updates", *IEEE Transactions on Circuits and Systems - II: Analog and Digital Signal Processing*, vol. 44, no. 3, pp. 209-216, 1997.

[23] M. Godavarti, A. O. Hero III, "Stochastic partial update LMS algorithm for adaptive arrays", *In Proceedings of IEEE Sensor Array and Multichannel Signal Processing Workshop*, pp. 322-326, 2000.

[24] M. Godavarti, A. O. Hero III, "Partial Update LMS Algorithms", *IEEE Transactions on Signal Processing*, vol. 53, no. 7, pp. 2382-2399, 2005.

[25] T. Aboulnasr, K. Mayyas, "Selective coefficient update of gradient-based adaptive algorithms", *In Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 1929-1932, 1997.

[26] T. Aboulnasr, K. Mayyas, "Complexity reduction of the NLMS algorithm via selective coefficient update", *IEEE Transactions on Signal Processing*, vol. 47, no. 5, pp. 1421-1424, 1999.

[27] P. A. Naylor, W. Sherliker, "A short-sort M-MAX NLMS partial-update adaptive filter with applications to echo cancellation", *In Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, pp. 373-376, 2003.

[28] S. Gollamudi, S. Nagaraj and S. Kapoor and Y-F Huang, "Set-membership filtering and a set-membership normalized LMS algorithm with an adaptive step size", *IEEE Signal Processing Letters*, vol. 5, no. 5, pp. 111-114, 1998.

[29] D. E. Knuth, *The Art of Computer programming - Sorting and Searching*, vol. 3, 2nd edition, Addison-Wesley, 1998.

[30] T. Schertler, G. Smidt, "Implementation of a low-cost acoustic echo canceller", *In Proceedings of the International Workshop on Acoustic Echo and Noise Control*, pp. 49-52, 1997.

[31] D. L. Duttweiler, "A twelve-channel digital echo canceler", *IEEE Transactions on Communications*, vol. 26, pp. 647-653, May 1978.

[32] A. Mader, H. Puder, G. U. Schmidt, "Step-size control for acoustic echo cancellation filters - an overview", *Signal Processing*, vol. 80, pp. 1697-1719, 2000.

[33] T. Gänsler, M Hansson, C.-J. Ivarsson, G. Salomonsson, "A double-talk detector based on coherence", *IEEE Transactions on Communnications*, vol. 44, pp. 1421-1427, November 1996.

[34] J. Benesty, D. R. Morgan, J. H. Cho, "A new class of doubletalk detectors based on cross-correlation", *IEEE Transactions on Speech and Audio Process.*, vol. 8, pp. 168-172, March 2000.

Part I

Efficient Multichannel NLMS Implementation for Acoustic Echo Cancellation

Part I is reprinted, with permission, from

Fredric Lindstrom, Christian Schüldt, Ingvar Claesson, "Efficient Multichannel NLMS Implementation for Acoustic Echo Cancellation", EURASIP Journal on Audio, Speech, and Music Processing, Volume 2007, Article ID 78439, 6 pages, doi:10.1155/2007/78439, 2007.

Efficient Multichannel NLMS Implementation for Acoustic Echo Cancellation

Fredric Lindstrom, Christian Schüldt,
Ingvar Claesson

Abstract

An acoustic echo cancellation structure with a single loudspeaker and multiple microphones is, from a system identification perspective, generally modelled as a single input multiple output system. Such a system thus implies specific echo-path models (adaptive filter) for every loudspeaker to microphone path. Due to the often large dimensionality of the filters, which is required to model rooms with standard reverberation time, the adaptation process can be computationally demanding. This paper presents a selective updating normalized least mean square (NLMS)-based method which reduces complexity to nearly half in practical situations, while showing superior convergence speed performance as compared to conventional complexity reduction schemes. Moreover, the method concentrates the filter adaptation to the filter which is most misadjusted, which is a typically desired feature.

1 Introduction

Acoustic echo cancellation (AEC) [1, 2] is used in teleconferencing equipment in order to provide high quality full-duplex communication. The core of an AEC solution is an adaptive filter which estimates the impulse response of the loudspeaker-enclosure-microphone (LEM) system. Typical adaptive algorithms for the filter update procedure in the AEC are the least mean square, normalized least mean square (LMS, NLMS) [3], affine projection (AP) and recursive least squares (RLS) algorithms [4]. Of these, the NLMS based algorithms are

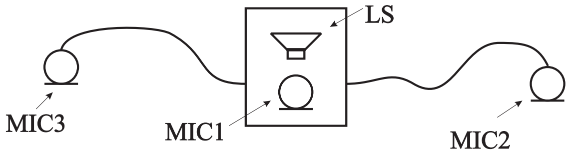

Figure 1: AEC unit with expansion microphones

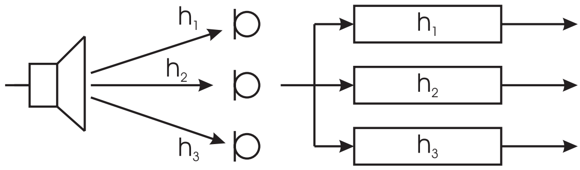

Figure 2: Schematic picture over multimicrophone system modelled as a single input multiple output system

popular in industrial implementations thanks to their low-complexity and finite precision robustness.

Multi-microphone solutions are frequent in teleconferencing equipment targeted for larger conference rooms. This paper considers a system consisting of one loudspeaker and three microphones. The base unit of the system contains the loudspeaker and one microphone and it is connected to two auxiliary expansion microphones, as shown in figure 1. Such a multi-microphone system constitutes a single input multiple output (SIMO) multichannel system with several system impulse responses to be identified, figure 2. Thus, the signal processing task can be quite computational demanding.

Several methods for computational complexity reduction of the LMS/NLMS algorithms have been proposed and analyzed, e.g. [5]–[14]. In this paper a related low complexity algorithm for use in a multi-microphone system is proposed.

2 Complexity reduction methods

The LEM system can be modelled as a time invariant linear system, $\mathbf{h}(k) = [h_0(k), \cdots, h_{N-1}(k)]^T$, where $N-1$ is the order of the finite impulse response (FIR) model [1] and k is the sample index. Thus, the desired (acoustic echo) signal $d(k)$ is given by $d(k) = \mathbf{h}(k)^T\mathbf{x}(k)$, where $\mathbf{x}(k) = [x(k), \cdots, x(k - N + 1)]^T$ and $x(k)$ is the input (loudspeaker) signal. The measured (microphone) signal $y(k)$ is obtained as $y(k) = d(k) + n(k)$, where $n(k)$ is near-end noise. Assuming an adaptive filter $\hat{\mathbf{h}}(k)$ of length N is used, i.e. $\hat{\mathbf{h}}(k) = [\hat{h}_0(k), \cdots, \hat{h}_{N-1}(k)]^T$, the NLMS algorithm is given by

$$e(k) = y(k) - \hat{d}(k) = y(k) - \mathbf{x}(k)^T\hat{\mathbf{h}}(k), \tag{1}$$

$$\beta(k) = \frac{\mu}{||\mathbf{x}(k)||^2 + \epsilon}, \tag{2}$$

$$\hat{\mathbf{h}}(k+1) = \hat{\mathbf{h}}(k) + \beta(k)e(k)\mathbf{x}(k). \tag{3}$$

where $\hat{d}(k)$ is the estimated echo, $e(k)$ the error (echo cancelled) signal, $\beta(k)$ the step-size, $||\mathbf{x}(k)||^2 = \mathbf{x}(k)^T\mathbf{x}(k)$ the squared Euclidian norm, μ the step-size control parameter and ϵ a regularization parameter [4].

Low-complexity periodical and partial updating schemes reduce the computational complexity of the LMS/NLMS by performing only a part of the filtering update, equations (2)–(3). The periodic NLMS performs the filter update only at periodical sample intervals. This updating can be distributed over the intermediate samples [5]. The sequential NLMS updates only a part of the N coefficients at every sample in a sequential manner [5]. Several methods for choosing which coefficients to update at what sample instant have been proposed, e.g. choosing a subset containing the largest coefficients in the regressor vector [6], low-complexity version of largest regressor vector coefficient selection [7], block based regressor vector methods [8, 9], and schemes based on randomization in the update procedure [10]. The updating can also be based on assumptions of the unknown plant [11, 12]. Another approach of omitting updates is possible in algorithms where the step-size is zero for a large number of updates [13, 14].

In a SIMO-modelled M microphone system, there are M adaptive filters $\hat{\mathbf{h}}_m(k)$ with $m \in \{1, \cdots, M\}$, to be updated at each sample, i.e.

$$\hat{\mathbf{h}}_m(k+1) = \hat{\mathbf{h}}_m(k) + \frac{\mu e_m(k)\mathbf{x}(k)}{||\mathbf{x}(k)||^2 + \epsilon} \qquad m = 1, \cdots, M. \tag{4}$$

see figure 2 for an example with $M = 3$. The updating scheme proposed in this paper explores the possibility of choosing between the different update equations based on comparison between the M different error signals $e_m(k)$.

3 The proposed algorithm

An adaptive linear filtering process can generally be divided in two parts; the *filtering*, equation (1), and the *adaptation*, equation (2) and (3). In an echo cancellation environment, the filtering part generally is performed at every sample instant in order to produce a constant audio stream. Although it is most often efficient (in terms of convergence) to perform filter updating at every sample instant, it is not necessary. In practice this might not even be possible due to complexity issues. This especially applies to acoustic echo cancellation environments where the dimension of the system filters is large.

One approach in a M-microphone system is to update only one adaptive filter every sample in a round-robin manner, i.e. periodic NLMS. This also ensures equal (for all filters) and predictable convergence since the update occurrences are deterministic. The disadvantage is that convergence is slow.

This paper proposes another updating method which instead updates the filter with the largest output error. To illustrate the method, assume that $M = 3$ (3 adaptive filters), that the present sample index is k and that filter 1 was updated at sample index $k-1$, filter 3 at $k-2$, and filter 2 at $k-3$, as illustrated in table 1. Thus, the available errors that can be used in the update at the present sample index k are $e_1(k)$ for filter 1, $e_2(k)$, $e_2(k-1)$ and $e_2(k-2)$ for filter 2, and $e_3(k)$ and $e_3(k-1)$ for filter 3. For example the error $e_1(k-2)$ cannot be used since it is related to the configuration of filter 1 prior to the latest update. From the available errors the algorithm chooses the error with the largest magnitude and then perform the corresponding update, compare with equations (7) and (8) below.

Sample index	Filter 1	Filter 2	Filter 3
k	$e_1(k)$	$e_2(k)$	$e_3(k)$
$k-1$	UPDATE	$e_2(k-1)$	$e_3(k-1)$
$k-2$	X	$e_2(k-2)$	UPDATE
$k-3$	X	UPDATE	X

Table 1: Example to illustrate the matrix $\mathbf{E}(k)$

An algorithm for the method is as follows. After filtering all M output channels according to equation (1), the output errors from all filters are inserted in a $L \times M$ matrix

$$\mathbf{E}(k) = \begin{pmatrix} e_1(k) & e_2(k) & e_3(k) & \cdots & e_M(k) \\ & & \underline{\mathbf{E}}(k-1) & & \end{pmatrix}, \tag{5}$$

where M is the number of adaptive filters (channels) and L determines the number of previous samples to consider. The $L-1 \times M$ matrix $\underline{\mathbf{E}}(k-1)$ consists of the $L-1$ upper rows of $\mathbf{E}(k-1)$, i.e.

$$E(l+1,m,k) = E(l,m,k-1) \quad \begin{array}{l} l = 1,\cdots,L-1 \\ m = 1,\cdots,M \end{array} \tag{6}$$

where l and m denotes row and column indexes, respectively, and $E(l,m,k)$ is the element at row l and column m in $\mathbf{E}(k)$.

The decision of which filter to update and with what output error (and corresponding input vector) is determined by the element in $\mathbf{E}(k)$ with maximum absolute value,

$$e_{\max}(k) = \max_{l,m} |E(l,m,k)| \quad \begin{array}{l} l = 1,\cdots,L \\ m = 1,\cdots,M. \end{array} \tag{7}$$

The row and column indexes of the element in $\mathbf{E}(k)$ with the maximum absolute value are denoted $l_{\max}(k)$ and $m_{\max}(k)$. For clarity of presentation the sample index is omitted, i.e. $l_{\max} = l_{\max}(k)$ and $m_{\max} = m_{\max}(k)$.

The filter corresponding to the row index $m_{\max}$, i.e. the filter $\hat{\mathbf{h}}_{m_{\max}}(k)$, is then updated with

$$\hat{\mathbf{h}}_{m_{\max}}(k+1) = \hat{\mathbf{h}}_{m_{\max}}(k) + \frac{\mu e_{\max}(k)\mathbf{x}(k-l_{\max}+1)}{||\mathbf{x}(k-l_{\max}+1)||^2+\epsilon}. \tag{8}$$

This filter update of filter $\hat{\mathbf{h}}_{m_{\max}}(k)$ will make the error elements $E(l,m_{\max},k),\ l = 1,\cdots,L$ obsolete, since these are errors generated by $\hat{\mathbf{h}}_{m_{\max}}(k)$ prior to the update. Consequently, to avoid future erroneous updates these elements should be set to 0, i.e. set

$$E(l,m_{\max},k) = 0 \qquad \text{for } l = 1,\cdots,L. \tag{9}$$

An advantage over periodic NLMS is that the proposed structure does not limit the update to be based on the current input vector $\mathbf{x}(k)$, but allows updating based on previous input vectors as well, since the errors not yet used for an update are stored in $\mathbf{E}(k)$. Further, largest output error update will concentrate the updates to the corresponding filter. This is normally a desired feature in an acoustic echo cancellation environment with multiple microphones. For example, consider the setup in figure 1 with all adaptive filters fairly converged. If then one of the microphones is dislocated, this results in an echo-path change for the corresponding adaptive filter. Naturally, it is desired to concentrate all updates to this filter.

4 Analysis

In the previously described scenario, where several input vectors are available but only one of them can be used for adaptive filter updating (due to complexity issues), it might seem intuitive to update with the input vector corresponding to the largest output error magnitude. In this section it is shown analytically that, under certain assumptions, choosing the largest error maximizes the reduction.

The error deviation vector for the m:th filter $\mathbf{v}_m(k)$ is defined as $\mathbf{v}_m(k) = \mathbf{h}_m(k) - \hat{\mathbf{h}}_m(k)$, and the mean squared deviation as $\mathcal{D}(k) = \mathbb{E}\{||\mathbf{v}_m(k)||^2\}$, where $\mathbb{E}\{\cdot\}$ denotes expectation [4]. Assume that no near-end sound is present, $n(k) = 0$, and no regularization is used, $\epsilon = 0$, and that the errors available for updating filter m are $e_m(k - l_m)$ with $l_m = 0, \cdots, L_m$ and $L_m < L$, i.e. the available errors in matrix $\mathbf{E}(k)$ that corresponds to filter m. Updating filter m using error $e_m(k - l_m)$ gives

$$\begin{aligned} ||\mathbf{v}_m(k+1)||^2 &= \\ &= ||\mathbf{v}_m(k) - \beta(k) e_m(k-l_m)\mathbf{x}(k-l_m)||^2 \end{aligned} \tag{10}$$

and by using

$$e_m(k-l_m) = \mathbf{x}(k-l_m)^T\mathbf{v}_m(k) = \mathbf{v}_m(k)^T\mathbf{x}(k-l_m) \tag{11}$$

in (10), the following is obtained

$$\begin{aligned} ||\mathbf{v}_m(k+1)||^2 &= \\ &= \mathbf{v}_m(k)^T\mathbf{v}_m(k) \\ &- \frac{(2\mu-\mu^2)}{||\mathbf{x}(k-l_m)||^2} e_m^2(k-l_m). \end{aligned} \tag{12}$$

Thus, the difference in mean square deviation from one sample to the next is given by,

$$\mathcal{D}_m(k+1) - \mathcal{D}_m(k) = -(2\mu-\mu^2)\mathbb{E}\left\{\frac{e_m^2(k-l_m)}{||\mathbf{x}(k-l_m)||^2}\right\}, \tag{13}$$

which corresponds to a reduction under the assumption that $0 < \mu < 2$.

Further, assuming small fluctuations in the input energy $||x(k)||^2$ from one iteration to the next, i.e. assuming

$$||\mathbf{x}(k)||^2 = ||\mathbf{x}(k-1)||^2 =, \cdots, = ||\mathbf{x}(k-L_m+1)||^2, \tag{14}$$

gives [4]

$$\mathcal{D}_m(k+1) - \mathcal{D}_m(k) = -(2\mu - \mu^2)\frac{\mathbb{E}\{e_m^2(k-l_m)\}}{\mathbb{E}\{||\mathbf{x}(k)||^2\}}. \tag{15}$$

The total reduction $r(k)$ in deviation, considering all M filters is thus

$$r(k) = \sum_{m=1}^{M} \mathcal{D}_m(k+1) - \mathcal{D}_m(k). \tag{16}$$

Only one filter is updated each time instant. Assume error $E(l, m, k)$ is chosen for the update. Then $r(k)$ is given by

$$r(k) = -(2\mu - \mu^2)\frac{\mathbb{E}\{E^2(l,m,k)\}}{\mathbb{E}\{||\mathbf{x}(k)||^2\}}. \tag{17}$$

From equation (17) it can be seen that the reduction is maximized if $e_{\max}(k)$, (see equation (7)), is chosen for the update, i.e. as done in the proposed algorithm.

The proposed algorithm can be seen as a version of the periodic NLMS. Analysis of convergence, stability and robustness for this branch of (N)LMS algorithms are provided in e.g. [5, 15].

5 Complexity and implementation

The algorithm proposed in this paper is aimed for implementation in a general Digital Signal Processor (DSP), typically allowing multiply-add-and-accumulate arithmetic operations to be performed in parallel with memory reads and/or writes (e.g. [16]). In such a processor the filtering operation can be achieved in N instructions and the NLMS update will require $2N$ instructions. Both the filtering and the update requires two memory reads, one addition and one multiplication per coefficient, which can be performed by the DSP in one instruction. However, the result from the filter update is not accumulated but it needs be written back to memory. Therefore, the need for two instructions per coefficient for the update operation.

Suppose an M-channel system with the same number of adaptive filters, all with the length of N. The standard NLMS-updating thus requires $3MN$ DSP-instructions.

Updating the matrix $\mathbf{E}(k)$, equation (5), can be implemented using circular buffering and thus requires only M store-instructions (possible pointer modifications disregarded), while clearing of $\mathbf{E}(k)$, equation (9), takes a maximum of

L instructions (also disregarding possible pointer modifications). Searching for the maximum absolute valued element in $\mathbf{E}(k)$, equations (7), requires a maximum of $2LM$ instructions (LM abs-instructions and LM max-instructions). The parameter $||\mathbf{x}(k)||^2$ can be calculated very efficient through recursion, i.e.

$$||\mathbf{x}(k)||^2 = ||\mathbf{x}(k-1)||^2 + x^2(k) - x^2(k-N), \tag{18}$$

and its computational complexity can be disregarded in this case.

All together, this means that the number of DSP-instructions required for the proposed solution can be approximated with

$$MN + M + L + 2ML + 2N. \tag{19}$$

For acoustic echo cancellation, N is generally quite large (> 1000) due to room reverberation time. In this case we typically have $N \gg L$ and $N \gg M$, which means that equation (19) is approximately $N(M+2)$. The complexity reduction in comparison with standard NLMS-updating is then

$$\frac{M+2}{3M}, \tag{20}$$

which for $M = 3$ gives a complexity reduction of nearly a half (5/9). For higher values of M the reduction is even larger. Further reduction in complexity can also be achieved if updates are performed say every other or every third sample.

6 Simulations

The performance of the proposed method was evaluated through simulations with speech as input signal. Three impulse responses ($\mathbf{h}_1$, $\mathbf{h}_2$ and $\mathbf{h}_3$), shown in figure 3, all of length $N = 1800$ were measured with three microphones, according to the constellation in figure 1, in a normal office. The acoustic coupling between the loudspeaker and the closest microphone, AC1, was manually normalized to 0dB and the coupling between the loudspeaker and the second and third microphone, AC2 and AC3, were then estimated to -6dB and -7dB respectively. Thus, $10\log_{10}\left(||\mathbf{h}_2||^2/||\mathbf{h}_1||^2\right) = -6$dB and $10\log_{10}\left(||\mathbf{h}_3||^2/||\mathbf{h}_1||^2\right) = -7$dB.

Output signals $y_1(k)$, $y_2(k)$ and $y_3(k)$ were obtained by filtering the input

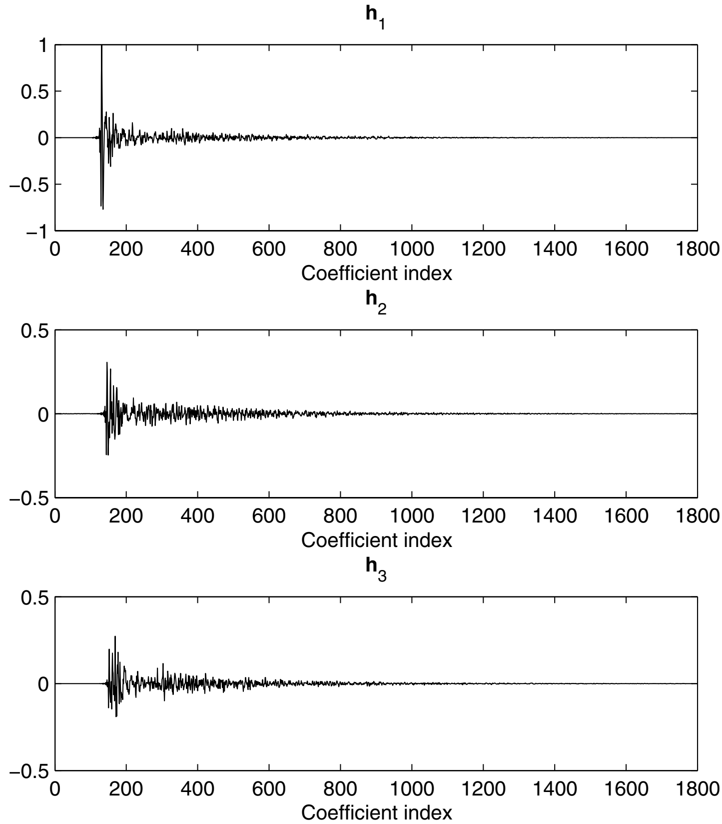

Figure 3: Impulse responses used in the simulations.

signal $x(k)$ with the three obtained impulse responses and adding noise,

$$\begin{aligned} y_1(k) &= \mathbf{x}(k)^T\mathbf{h}_1 + n_1(k) \\ y_2(k) &= \mathbf{x}(k)^T\mathbf{h}_2 + n_2(k) \\ y_3(k) &= \mathbf{x}(k)^T\mathbf{h}_3 + n_3(k). \end{aligned}$$

The noise sources $n_1(k)$, $n_2(k)$ and $n_3(k)$ were independent, but had the same characteristics (bandlimited flat spectrum). Echo-to-noise ratio was approximately 40dB for microphone 1 and 34dB and 33dB for microphone 2 and 3, respectively.

In the simulations four low-complexity methods of similar complexity were compared; the periodic (N)LMS [5], random NLMS (similar to SPU-LMS [10]) selecting which filter to be updated in a stochastic manner (with all filters having equal probability of an update), M-Max NLMS [6] and the proposed NLMS. The performance of the full update NLMS is also shown for comparison. The periodic NLMS, random NLMS and the proposed method limits the updates to one whole filter at each time interval, while M-Max NLMS instead updates all filters but only does this for a subset (1/3 in this case) of all coefficients. However, since M-Max NLMS requires sorting of the input vectors, the complexity for this method is somewhat larger ($2\log_2 N + 2$ comparisons and $(N-1)/2$ memory transfers [9]). Zero initial coefficients were used for all filters and methods. The result is presented in figure 4, where the normalized filter mismatch, calculated as

$$10\log_{10}\left(\frac{||\mathbf{h}_m - \hat{\mathbf{h}}_m(k)||^2}{||\mathbf{h}_m||^2}\right) \qquad m = 1, 2, 3, \tag{21}$$

for the three individual filters and solutions are presented. Of the four variants with similar complexity, the proposed method is clearly superior to the conventional periodic NLMS and also to the random NLMS. The performance of the M-Max NLMS and the proposed solution is comparable, although the proposed solution performs better or equal for all filters.

The algorithm automatically concentrates computational resources to filters with large error signals. This is demonstrated in figure 5, where filter 2 undergoes an echo path change, i.e. a dislocation of the microphone. In figure 5 it can be seen that the proposed algorithm basically follows the curve of the full update NLMS immediately after the echo path change.

If one specific microphone is subject to an extreme acoustic situation, e.g. it is placed in another room or placed immediately next to a strong noise source, there is a risk of "getting stuck", i.e. the corresponding filter has large output error for all input vectors and thus is updated all the time. This problem can be reduced by setting a limit on the lowest rate of updates for a filter, i.e. if filter m has not been updated for the last U samples it is forced to update the next iteration. However, this does not resolve the issue optimally. A more sophisticated method is to monitor the echo reduction of the filters and bypass or reduce the resources allocated to filters not providing significant error reduction. Implementing these extra functions will of course add complexity.

7 Conclusions

In an acoustic multichannel solution with multiple adaptive filters, the computation power required to update all filters every sample can be vast. This paper has presented a solution which updates only one filter every sample and thus significantly reduces the complexity, while still performing well in terms of convergence speed. The solution also handles echo-path changes well, since the most misadjusted filter gets the most computation power, which often is a desirable feature in practice.

References

[1] E. Hänsler and G. Schmidt, *Acoustic Echo and Noise Control: A Practical Approach*, Wiley, 2004.

[2] M.M. Sondhi, "An adaptive echo canceler," *Bell Syst. Tech. J.*, vol. 646, pp. 497–510, January 1967.

[3] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*, Prentice-Hall, 1985.

[4] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, 4th edition, 2002.

[5] S. C. Douglas, "Adaptive filters employing partial updates," *IEEE Transactions on Circuits and Systems - II: Analog and Digital Signal Processing*, vol. 44, no. 3, pp. 209–216, 1997.

[6] T. Aboulnasr and K. Mayyas, "Complexity reduction of the nlms algorithm via selective coefficient update," *IEEE Transactions on Signal Processing*, vol. 47, no. 5, pp. 1421–1424, 1999.

[7] P. A. Naylor and W. Sherliker, "A short-sort m-max nlms partial-update adaptive filter with applications to echo cancellation," *Proc. of IEEE ICASSP'03*, vol. 5, pp. 373–376, 2003.

[8] K. Dogancay and O. Tanrikulu, "Adaptive filtering with selective partial updates," *IEEE Transactions on Circuits and Systems - II: Analog and Digital Signal Processing*, vol. 48, no. 8, pp. 762–769, 2001.

[9] T. Schertler, "Selective block update of nlms type algorithms," *Proc. of IEEE ICASSP'98*, vol. 3, 1998.

[10] M. Godavarti and A. O. Hero III, "Partial update lms algorithms," *IEEE Transactions on Signal Processing*, vol. 53, no. 7, pp. 2382–2397, 2005.

[11] E. Hänsler, G. Schmidt (Eds: J. Benesty, and Y. Huang), *Adaptive Signal Processing*, Springer, 2003.

[12] M. Kuo and J. Chen, "Multiple-microphone acoustic echo cancellation system with the partial adaptive process," *Digital Signal Processing*, vol. 3, no. 1, pp. 54–63, 1993.

[13] S. Gollamudi, S. Kapoor, S. Nagaraj, and Y-F Huang, "Set-membership adaptive equalization and updator-shared implementation for multiple channel communications systems," *IEEE Transactions on Signal Processing*, vol. 46, no. 9, pp. 2372–2385, 1998.

[14] S. Werner, M. L. R. de Campos, and S. R. Diniz, "Low-complexity constrianed affine projection algorithms," *IEEE Transactions on Signal Processing*, vol. 53, no. 12, pp. 4545–4555, 2005.

[15] W. A. Gardner, "Learning characteristics of stochastic-gradient-descent algorithms: A general study, analysis, and critique," *Signal Processing*, vol. 6, no. 2, pp. 113–133, 1984.

[16] *ADSP-BF533 Blackfin processor hardware reference*, Analog Devices, 2005.

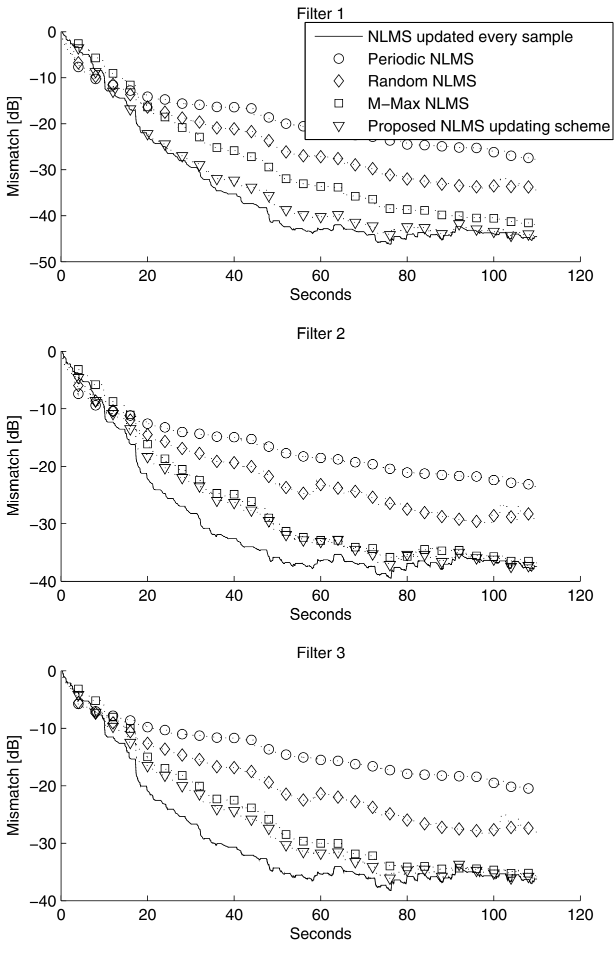

Figure 4: Mismatch for the the evaluated methods.

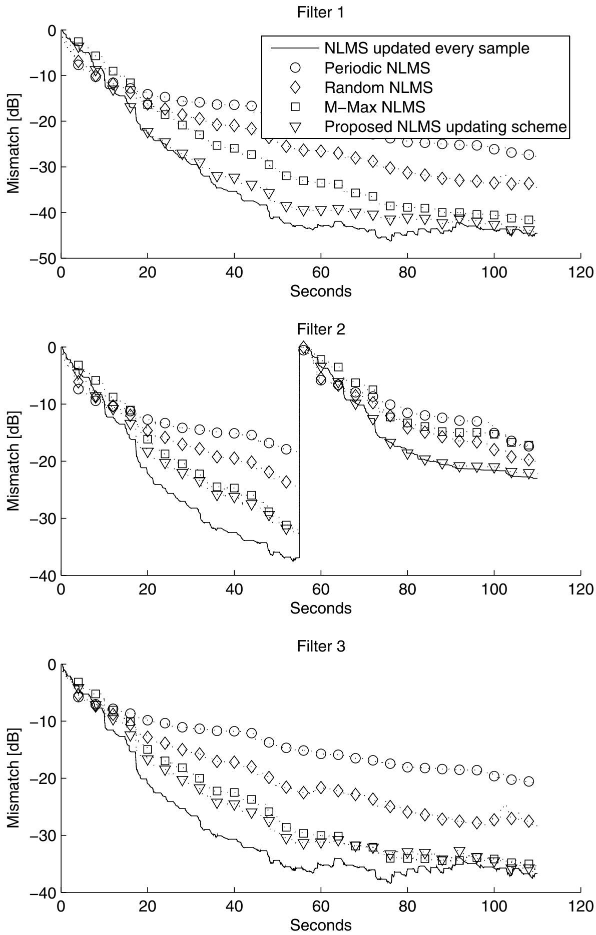

Figure 5: Mismatch for the the evaluated methods, where an echo-path change occurs for filter 2 after 55 seconds.

Part II

Low-Complexity Adaptive Filtering Implementation for Acoustic Echo Cancellation

Part II is reprinted, with permission, from

Christian Schüldt, Fredric Lindstrom, Ingvar Claesson, "Low-Complexity Adaptive Filtering Implementation for Acoustic Echo Cancellation", Proceedings of IEEE TENCON, Hong Kong, November 2006.

Low-Complexity Adaptive Filtering Implementation for Acoustic Echo Cancellation

Christian Schüldt, Fredric Lindstrom, Ingvar Claesson

Abstract

Acoustic echo cancellation is generally achieved with adaptive FIR filters. Due to the often large dimensionality of the adaptive filters, required to model rooms with standard reverberation time, the adaptation process can be computationally demanding. This paper presents a block based selective updating method which reduces the complexity with nearly a half in practical situations, while showing superior convergence speed performance as compared to conventional partial update complexity reduction schemes.

1 Introduction

Acoustic echo cancellation (AEC) [1] is used in teleconferencing equipment in order to provide high quality full-duplex communication. The core of an AEC solution is generally an adaptive filter which estimates the impulse response of the loudspeaker-enclosure-microphone (LEM) system. Typical adaptive algorithms for the AEC filter update procedure are the least mean square, normalized least mean square (LMS, NLMS), affine projection (AP) and recursive least squares (RLS) algorithms [2]. Of these, the NLMS based algorithms are particulary popular in industrial implementations thanks to their low complexity and finite precision robustness.

The echo cancellation environment can vary significantly, and in order to maintain acceptable echo reduction in rooms with long reverberation time, large dimensionality of the adaptive filters is required. Thus, the signal processing task can be computationally demanding.

Several partial update methods for computational complexity reduction of various adaptive filtering algorithms have been proposed and analyzed, e.g. [3, 4, 5, 6, 7, 8, 9, 10, 11] for the LMS/NLMS algorithms and [12, 6, 11] for AP. In this paper, a low complexity scheme applicable to both NLMS and AP, is presented and compared to several other complexity reduction methods.

2 Complexity reduction methods

Commonly, the LEM system is modelled as a finite impulse response (FIR) model [1], $\mathbf{h} = [h_0, \cdots, h_{N_\mathrm{L}-1}]^T$, where N_L is the filter order. Filtering of the input signal $x(k)$ then produces the desired (acoustic echo) signal $d(k)$, given by $d(k) = \mathbf{h}^T \mathbf{x}_\mathrm{L}(k)$, where $\mathbf{x}_\mathrm{L}(k) = [x(k), \cdots, x(k - N_\mathrm{L} + 1)]^T$ and k is the sample index. By adding near-end noise $n(k)$, the measured (microphone) signal $y(k)$ is obtained, $y(k) = d(k) + n(k)$. The NLMS algorithm is then

$$e(k) = y(k) - \hat{d}(k) = y(k) - \mathbf{x}(k)^T \mathbf{h}(k) \tag{1}$$

$$\hat{\mathbf{h}}(k+1) = \hat{\mathbf{h}}(k) + \mu \frac{e(k)\mathbf{x}(k)}{||\mathbf{x}(k)||^2 + \epsilon}, \tag{2}$$

where $\hat{\mathbf{h}}(k) = [\hat{h}_0(k), \cdots, \hat{h}_{N-1}(k)]^T$ is the adaptive FIR filter of length N, $\mathbf{x}(k) = [x(k), \cdots, x(k - N + 1)]^T$ the N length regressor vector, $\hat{d}(k)$ the estimated echo, $e(k)$ the error (echo cancelled) signal, μ is the step-size, and ϵ is a positive constant to avoid division by zero or near-zero [2].

Several low-complexity methods which only perform a part of the filtering update, equation (2), have been proposed. Partial NLMS is performed by dividing the N filter coefficients into B blocks and only updating M of these blocks each sample. Which blocks to update can be selected in either a sequential manner [3], randomly [8] or by updating the parts which correspond to large energy of the regressor vector [6, 7]. For $B = N$, i.e. the block-size set to 1 sample, the later method becomes the M-Max NLMS [4, 5]. In contrast, the periodic NLMS, updates all N filter coefficients but only at periodical sample intervals. The update can also be partitioned over all samples [3]. Other low-complexity updating schemes, possible in NLMS based algorithms where the step-size is zero for a large number of updates [10, 11], have been proposed as well as those based on assumptions of the unknown plant [9].

Complexity-reductions methods based on AP are for example the Fast Affine Projection (FAP) algorithm [12], which reduces the complexity almost to a NLMS level, except for a matrix inversion. Other methods are selective partial AP [6] and M-Max AP [13] (both analog to their NLMS counterpart).

3 The proposed algorithm

The method proposed in this paper is similar to the periodic NLMS, but updates with as large error as possible, instead of updating the filter with input vectors obtained at fixed time instances. This is achieved by a buffering technique, where blocks of L samples are collected and processed. As a result, a delay of L samples is introduced in the signal path, but since L is relatively small (< 10), this normally has insignificant impact on the whole system. Similarity to the block LMS algorithm [2] is also apparent, with the difference being the number of updates per block (L for the block LMS and only one for the proposed method).

The L samples of one block are filtered and the corresponding output errors are calculated according to

$$e(lL+i) = y(lL+i) - \mathbf{x}(lL+i)^T \mathbf{h}(l) \quad i = 0, \cdots, L-1, \tag{3}$$

where the block index l is related to the original sample index k and block length L as

$$\begin{aligned} k = lL + i, \quad & i = 0, \cdots, L-1, \\ & l = 1, 2, \cdots \end{aligned} \tag{4}$$

The decision of what output error (and corresponding input vector) should be used for the update is determined by

$$i_l = \underset{i \in \{0, \cdots, L-1\}}{\arg\max} \; |e(lL+i)|, \tag{5}$$

and an update of the filter is then performed with

$$\hat{\mathbf{h}}(l+1) = \hat{\mathbf{h}}(l) + \frac{\mu e(lL+i_l)\mathbf{x}(lL+i_l)}{||\mathbf{x}(lL+i_l)||^2 + \epsilon}. \tag{6}$$

Since this structure updates the filter with the input vector resulting in the largest largest output error, small errors will generally be ignored. However, the resulting impact on the filter convergence is likely to be minor, since a small error occurs due to a well converged filter or due to orthogonality between the input vector and the filter mismatch vector. In both cases, a filter update will not result in any significant convergence.

4 Complexity and implementation

The whole purpose of the proposed solutions is to reduce the complexity of an NLMS-implementation without sacrificing too much convergence speed. In a real application, the solution is generally realized through a Digital Signal Processor (DSP) which is capable of performing multiply-add-and-accumulate arithmetic instructions in parallel with memory reads and/or writes.

FIR filtering with a filter length N typically requires N DSP-instructions. Searching for the maximum absolute valued element in equation (5) all together requires $2L$ instructions. An NLMS-update, equation (6), requires $2N$ instructions. However, since equations (5) and (6) are only calculated once for ever block of L samples, the number of instructions required per sample is $2(L+N)/L$. The included scalar product $\mathbf{x}(lL+i_{\max})^T\mathbf{x}(lL+i_{\max})$ can be calculated very efficient through recursion [2] and its computational complexity can be disregarded in this case.

All together, this means that the number of DSP-instructions required for the proposed solution is $N+2(L+N)/L$ in comparison with the complexity of a standard NLMS-update which is $3N$. For acoustic echo cancellation, N is generally quite large (> 1000) due to room reverberation time. In this case we typically have $N \gg L$, which means that the complexity can be approximated as $N+2N/L$. The complexity reduction in comparison with standard NLMS-updating is then

$$\frac{1}{3}+\frac{2}{3L}, \tag{7}$$

which for $L=4$ gives one half of the standard NLMS complexity.

5 Simulations

Comparisons between the proposed method denoted Max-E NLMS, M-Max NLMS, partial NLMS, random NLMS and standard NLMS were performed through simulations. The parameters for Max-E NLMS are shown in table 1 and the parameters for the other methods were chosen so that the complexity of the adaptation for each method was similar. However, for the M-Max NLMS, there is also additional complexity in the sorting of the input vectors ($2\log_2 N+2$ comparisons and $(N-1)/2$ memory transfers). Furthermore, N additional memory locations are also required [7].

The LEM system was modelled with a FIR filter $\mathbf{h}$ of length $N_L = 1024$, obtained through impulse response measurements of a normal office. The acoustic coupling between the loudspeaker and the microphone was normalized to 0dB.

To obtain the microphone signal $y(k)$, the loudspeaker signal $x(k)$ was filtered with the measured impulse response $\mathbf{h}$ and band limited flat spectrum

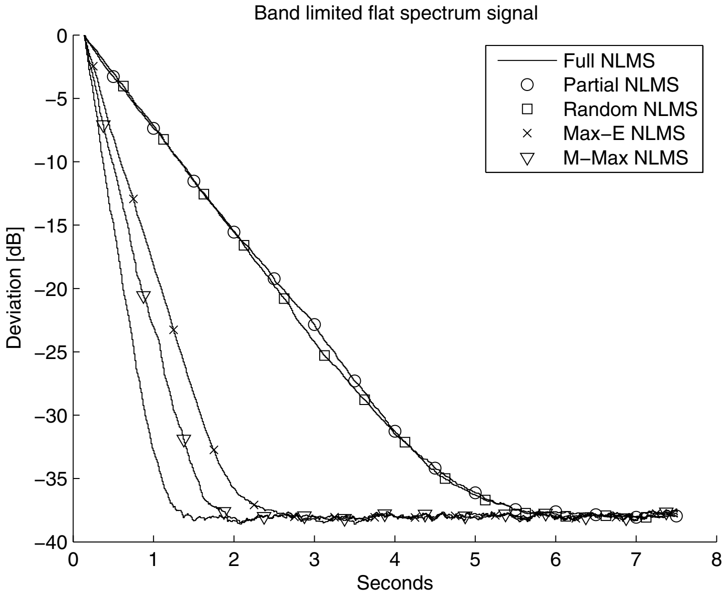

Figure 1: Filter deviation with a band limited flat spectrum signal as input.

noise was added so that the resulting echo-to-noise ratio was approximately 28dB. The sampling rate was 8kHz.

The results of the simulations for a band limited flat spectrum input signal is presented in figure 1, where the filter deviation, calculated as $\sum_{j=0}^{N-1}(h_j - \hat{h}_j(k))$, where h_j is the j:th element of $\mathbf{h}$ and $\hat{h}_j(k)$ the j:th element of the considered adaptive filter $\hat{\mathbf{h}}(k)$, respectively, is presented. Not surprisingly, the full NLMS has the fastest convergence (but also twice the complexity, compared to the other). Of the four variants with reduced complexity, the M-Max NLMS has slightly reduced convergence speed compared to full NLMS, followed by the proposed Max-E solution, random and partial NLMS.

Simulations were also performed with speech as input signal. However, an issue arises with the M-Max NLMS, since the step-size stability condition is tighter for this type of signal [14], and with the current setting M-Max NLMS diverges. Due to this, μ for M-Max NLMS is set as high as possible, but still allowing convergence, which in this case is $\mu = 0.5$. Step-size settings for the

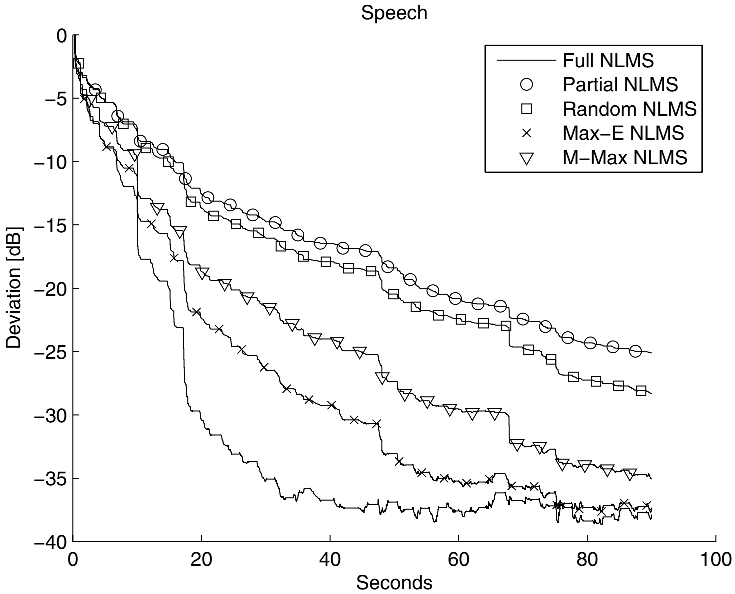

Figure 2: Filter deviation with a speech signal as input.

other methods were unchanged. Figure 2 shows the results of the simulation. Here, the proposed solution has the fastest convergence speed of the solutions with similar complexity, while full NLMS has the fastest convergence of all.

It can also be noted that the random NLMS performs slightly better than partial NLMS for speech signals (while for band limited flat spectrum signals there is no difference), which agrees with the results in [8].

6 Fast affine projection-version

The previously described scheme, where L samples are collected and the sample which produces the largest error is used for the filter update, can be applied to the fast affine projection (FAP) algorithm [12] as well. The procedure is highly similar, but since FAP is recursive and the consecutive updates are dependant, the implementation is somewhat less straight-forward.

For the sake of simplicity (both implementation- and notation-wise), only

FAP with the projection dimension 2 (FAP-2) is considered in this paper. Higher dimensions are possible, but perhaps not directly tractable due to the necessary $D \times D$ matrix inversion, where D is the projection dimension. For $D = 2$, the matrix inversion can be performed directly. Moreover, it has also been shown that considerable improvement (over NLMS) is gained for just $D = 2$ and that further significant convergence improvement for speech signals is not reached until D is increased up to 10 [12, 1].

The original FAP [12] is modified analogous to the NLMS version presented earlier, where filtering is performed in blocks of L samples and the output and updating is

$$\begin{aligned} e(lL+i) = y(lL+i) \quad - \\ (\mathbf{x}(lL+i)^T\hat{\mathbf{h}}(l) + \mu z(i, i_{l-1})\underline{\phi}(l-1)), \end{aligned} \tag{8}$$

$$z(i, i_{l-1}) = \mathbf{x}(lL+i)^T\mathbf{x}((l-1)L+i_{l-1}), \tag{9}$$

$$\mathbf{e}(l) = \begin{pmatrix} e(lL+i_l) \\ (1-\mu)e((l-1)L+i_{l-1}) \end{pmatrix}, \tag{10}$$

$$\boldsymbol{X}(l) = \begin{pmatrix} ||\mathbf{x}(lL+i_l)||^2 & z(i_l, i_{l-1}) \\ z(i_l, i_{l-1}) & ||\mathbf{x}((l-1)L+i_{l-1})||^2 \end{pmatrix} \tag{11}$$

$$\boldsymbol{\xi}(l) = \boldsymbol{X}(l)^{-1}\mathbf{e}(l), \tag{12}$$

$$\boldsymbol{\phi}(l) = \begin{pmatrix} 0 \\ \underline{\phi}(l-1) \end{pmatrix} + \boldsymbol{\xi}(l) \tag{13}$$

and

$$\hat{\mathbf{h}}(l+1) = \hat{\mathbf{h}}(l) + \mu\phi_1(l)\mathbf{x}((l-1)L+i_{l-1}), \tag{14}$$

where $\underline{\phi}(l-1)$ is the upper-most element of $\boldsymbol{\phi}(l-1)$, $\phi_1(l)$ is the lower-most element of $\boldsymbol{\phi}(l)$ and l,L,i and i_l are as defined in equations (4) and (5).

The complexity of the proposed FAP-2 solution is similar to that of the NLMS-version, with the difference being the correlation matrix $\boldsymbol{X}(l)$, which

Table 1: Max-E NLMS parameter settings.

Parameter	Value
N	1024
μ	0.95
L	4
ϵ	4

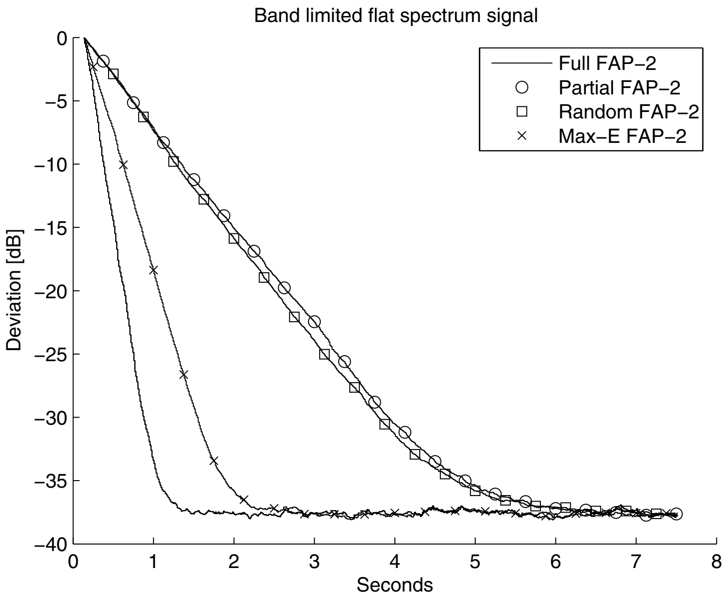

Figure 3: Filter deviation with a band limited flat spectrum signal as input.

can be calculated recursively, and its inverse, which then can be calculated directly.

Simulations were also performed for the FAP-2 version of the proposed algorithm. Figure 3 shows the results for a band limited flat spectrum input signal, while figure 4 shows the results for a speech signal. Similar to the NLMS results, the superior convergence performance of the proposed algorithm can be noted, while random FAP-2 performs better than the periodic FAP-2, which is also similar to the results of the NLMS simulations.

7 Conclusions

In acoustic echo cancellation environments, the computational resources required to update all adaptive filter coefficients every sample can be too costly. This paper has presented a block based solution which updates with the input vectors which produce the largest output error. The proposed solution sig-

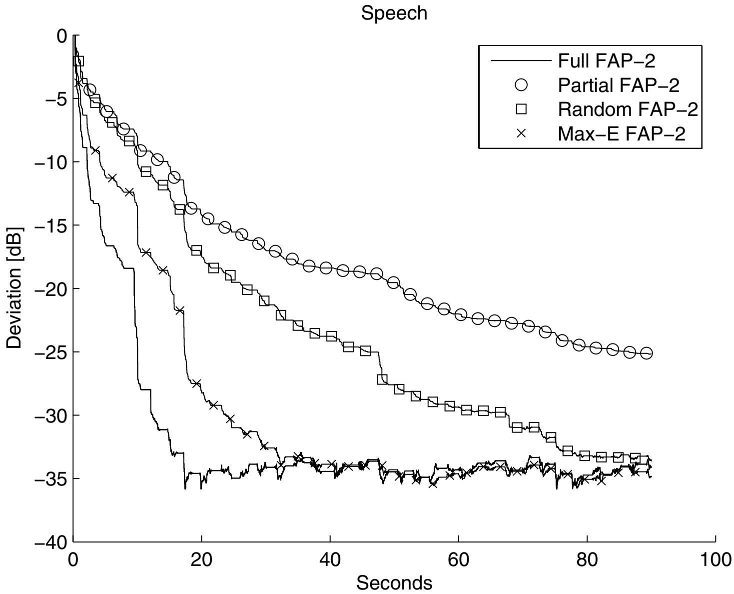

Figure 4: Filter deviation with a speech signal as input.

nificantly reduces the complexity of the adaptive filter, while still performing well in terms of convergence, compared to other partial update methods. As shown, this is specially true for speech signals.

Acknowledgment

The authors would like to thank the Swedish Knowledge Foundation (KKS) for funding.

References

[1] E. Hänsler and G. Schmidt, *Acoustic Echo and Noise Control: A Practical Approach.* Wiley, 2004.

[2] S. Haykin, *Adaptive Filter Theory*, 4th ed. Prentice-Hall, 2002.

[3] S. C. Douglas, "Adaptive filters employing partial updates," *IEEE Trans. on Circuits and Systems - II: Analog and Digital Signal Processing*, vol. 44, no. 3, pp. 209–216, 1997.

[4] T. Aboulnasr and K. Mayyas, "Complexity reduction of the NLMS algorithm via selective coefficient update," *IEEE Trans. on Signal Processing*, vol. 47, no. 5, pp. 1421–1424, 1999.

[5] P. A. Naylor and W. Sherliker, "A short-sort M-Max NLMS partial-update adaptive filter with applications to echo cancellation," *Proc. of IEEE ICASSP*, vol. 5, pp. 373–376, 2003.

[6] K. Dogancay and O. Tanrikulu, "Adaptive filtering with selective partial updates," *IEEE Trans. on Circuits and Systems - II: Analog and Digital Signal Processing*, vol. 48, no. 8, pp. 762–769, 2001.

[7] T. Shertler, "Selective block update of NLMS type algorithms," *Proc. of IEEE ICASSP*, vol. 3, 1998.

[8] M. Godavarti and A. O. Hero III, "Partial update LMS algorithms," *IEEE Trans. on Signal Processing*, vol. 53, no. 7, pp. 2382–2397, 2005.

[9] M. Kuo and J. Chen, "Multiple-microphone acoustic echo cancellation system with the partial adaptive process," *Digital Signal Processing*, vol. 3, no. 1, pp. 54–63, 1993.

[10] S. Gollamudi, S. Kapoor, S. Nagaraj, and Y.-F. Huang, "Set-membership adaptive equalization and updator-shared implementation for multiple channel communications systems," *IEEE Trans. on Signal Processing*, vol. 46, no. 9, pp. 2372–2385, 1998.

[11] S. Werner, M. L. R. de Campos, and S. R. Diniz, "Low-complexity constrianed affine projection algorithms," *IEEE Trans. on Signal Processing*, vol. 53, no. 12, pp. 4545–4555, 2005.

[12] S. L. Gay and S. Tavathia, "The fast affine projection algorithm," *Proc. of IEEE ICASSP*, vol. 5, pp. 3023–3026, 1995.

[13] P. A. Naylor and A. W. H. Khong, "Affine projection and recursive least squares adaptive filters employing partial updates," *Conference Record of the 38th Asilomar Conference on Signals, Systems and Computers*, vol. 1, pp. 950–954, 2004.

[14] I. Kammoun and M. Jaidane, “Exact performances analysis of a selective coefficient adaptive algorithm in acoustic echo cancellation,” *Proc. of IEEE ICASSP*, vol. 5, pp. 3245–3248, 2001.

Part III

A Low-Complexity Delayless Selective Subband Adaptive Filtering Algorithm

Part III is reprinted, with permission, from

Christian Schüldt, Fredric Lindstrom, Ingvar Claesson, “A Low-Complexity Delayless Selective Subband Adaptive Filtering Algorithm”, IEEE Transactions on Signal Processing, vol. 56, no. 12, pp. 5840-5850, December 2008.

A Low-Complexity Delayless Selective Subband Adaptive Filtering Algorithm

Christian Schüldt, Fredric Lindstrom, Ingvar Claesson

Abstract

Adaptive filters of significant order, requiring high computational complexity, are necessary in many applications such as acoustic echo cancellation and wideband active noise control. Successful approaches to lessen the computational complexity of such filters are subband methods, and partial updating schemes where only a part of the filter is updated at each instant. To avoid the time delay introduced by the subband-splitting, delayless structures which reconstructs a fullband filter, producing delayless output, from the adaptive subband filters have been proposed.

This paper proposes a delayless subband adaptive filter partial updating scheme, where the general idea is to only update the most misadjusted subband filter(s). Analysis in terms of mean square deviation is presented and shows that the fullband filter convergence speed is significantly increased, even for flat spectrum signals, as compared to traditional periodic subband filter update with the same computational complexity. Echo cancellation simulations with an artificial system to verify the analysis, using both flat spectrum signals and speech, is also presented, as well as off-line calculations using signals from a real system.

1 Introduction

Adaptive finite impulse response (FIR) filters is a vital component in many echo cancellation- and system estimation arrangements. The general idea is to feed the same input signal to both the system to be estimated and the adaptive filter, and using the difference of the respective outputs produced, i.e. the output error, as a measure of estimation performance. The output error is used for updating the adaptive filter. Perhaps the most frequently used adaptive filter updating algorithm is the (normalized) least mean square ((N)LMS) [1], owing to its ease of implementation, low complexity and robustness to fix-point arithmetic implications. One drawback of the (N)LMS is however slow convergence speed, especially in the case of colored input signals. To speed up the filter convergence, at the cost of increased computational complexity, algorithms such as the recursive least squares (RLS), affine projection (AP) [1], and its computationally more efficient approximation fast affine projection (FAP) [2] have been proposed.

Another approach for both increased convergence speed, mainly in the case of colored input signals, and reduced computational complexity is subband adaptive filtering [3]. This can be performed either in a transform domain [4, 5], or in the time domain [6, 7]. Other means for reduced complexity include partial updating algorithms, where the idea is to avoid updating of all filter coefficients at each time instant. Periodic NLMS performs the filter update only at periodical sample intervals, while the sequential NLMS updates only a part of all coefficients at every sample in a sequential manner. In essence, although having different stability properties, the convergence performance of these two methods are similar [8]. Other suggested methods for better convergence performance have been e.g. choosing a subset of the regressor vector containing the largest coefficients [9] and block based regressor vector methods [10, 11]. Several combinations of subband structures and partial updating algorithms have also been proposed. These have either been based on sequential updating [12] or used the magnitude of the regressor vectors in the respective subbands as selection criterion [4, 13].

A disadvantage of conventional subband structures is the delay introduced in the signal path by the filterbanks. To avoid this issue, *delayless* subband adaptive filter architectures have been proposed [6], where the general idea is to reconstruct a fullband filter from the adaptive subband filters. The reconstructed fullband filter is then used to produce the fullband output. Thus, the signal filtering is performed using the fullband filter, avoiding the delay introduced by the filterbanks, while adaptive filters are adapted in the subbands.

This paper proposes a delayless subband adaptive filter partial updating scheme, based on the idea to update only the subband filters which are most misadjusted. The outline of the paper is as follows: In section 2, the proposed subband filtering method and filterbank structure are described and in section 3, the conventional delayless subband NLMS is described. Section 4 shows the relation between the fullband filter mean square deviation and the mean square deviation of the individual subband filters. Results from this section is then used in section 5, where the proposed selective subband updating scheme is derived. Theoretical analysis of the proposed algorithm, periodic NLMS and full updating scheme is presented in section 6. The computational complexity of the proposed algorithm is presented in section 7. Section 8 verifies the analytical results through simulations using flat spectrum signals, colored stationary signals and speech with an artificial system in an echo cancellation application and in section 9, the proposed algorithm is subjected to speech signals recorded in a real setup in an office.

2 Polyphase filterbank structure

The delayless subband structure used in this paper is essentially the same as in [6], where the subband signals are obtained by convolution with a frequency shifted prototype lowpass filter [3]. The prototype filter used here is designed using the fast converging iterative least squares method provided by [14]. Thus, in the case of dividing the input signal $x(k)$ (see figure 1) into M subbands, the signal in subband $m \in \{0, \cdots, M-1\}$ will be

$$x_m(n) = \sum_{i=0}^{K-1} x(k-i) g_i e^{j\frac{2\pi m}{M} i}, \tag{1}$$

where n is the decimated subband sample index, R is the decimation ratio, $k = Rn$ is the fullband sample index, g_i is the i:th prototype filter coefficient, and K is the number of prototype filter coefficients. Rearranging the summation index in equation (1) according to

$$\begin{aligned} i = sM + q \qquad & q \in \{0, \cdots, M-1\} \\ & s \in \{0, \cdots, S-1\}, \end{aligned} \tag{2}$$

where $K = SM$ gives

$$x_m(n) = \sum_{q=0}^{M-1} e^{j\frac{2\pi m}{M} q} \sum_{s=0}^{S-1} x(Rn - sM - q) g_{sM+q}. \tag{3}$$

Thus, the subband filtering can be implemented very efficiently through M convolutions (each of length S) and one inverse FFT (fast fourier transform) every R:th fullband sample [3].

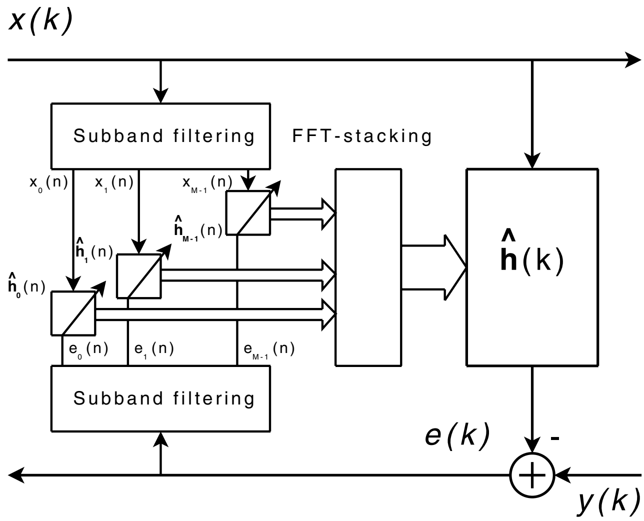

Figure 1: Closed loop delayless subband adaptive filtering configuration.

3 Subband normalized least mean square adaptive filtering implementation

For the closed loop case, which is considered in this paper, the non-delayed fullband output error is calculated directly as (see figure 1)

$$e(k) = y(k) - \hat{\mathbf{h}}(k)^T \mathbf{x}(k), \tag{4}$$

where $y(k) = d(k) + w(k)$, and $d(k)$ is the signal to be estimated, $w(k)$ is the local noise, $\hat{\mathbf{h}}(k) = [\hat{h}_0(k), \cdots, \hat{h}_{N-1}(k)]^T$ is the fullband adaptive filter, and $\mathbf{x}(k) = [x(k), \cdots, x(k-N+1)]^T$ is the regressor vector of length N. $[\cdot]^T$ denotes transpose. The fullband error $e(k)$ is partitioned into subbands $e_m(n)$, just as the input signal $x(k)$, see equation (1).

Then, NLMS updating of subband filter $\hat{\mathbf{h}}_m(n) = [\hat{h}_{m,0}(n), \cdots, \hat{h}_{m,N_M-1}(n)]^T$ of length N_M is performed as

$$\hat{\mathbf{h}}_m(n+1) = \hat{\mathbf{h}}_m(n) + \beta_m e_m^*(n)\mathbf{x}_m(n), \tag{5}$$

where

$$\beta_m = \frac{\mu_m}{||\mathbf{x}_m(n)||^2 + \epsilon}, \tag{6}$$

and μ_m is a step-size control parameter, ϵ is a regularization parameter [1], and * denotes complex conjugate. However, in the case of real fullband signals, which is what is considered in this paper, it is only necessary to update the $M/2+1$ first subband filters owing to Hermitian symmetry.

Transformation of the $M/2+1$ subband filters, in the case of $R = M/2$, is then performed through a technique called FFT-2 stacking [15], which is a refinement of the FFT-stacking technique suggested by [6]. The FFT-2 stacking is performed by taking a $2N_M$-point FFT of each subband filter and then stacking the DFT (discrete Fourier transform)-coefficients as

$$\hat{H}(l) = \begin{cases} \hat{H}_{[lM/2N]}(l \bmod 4N/M) & l \in [0, N) \\ 0 & l = N \\ \hat{H}^*(2N - l) & l \in (N, 2N) \end{cases} \tag{7}$$

where $\hat{H}(l)$ denotes DFT-coefficient l of the fullband filter and $\hat{H}_m(l \bmod 4N/M)$ denotes DFT-coefficient l modulo $4N/M$ of subband filter m, respectively. In this case $[\cdot]$ denotes rounding towards nearest integer.

Finally, the fullband filter is the N first samples of the $2N$-point inverse FFT of $\hat{H}(l)$.

4 Fullband- and subband filter deviation

This section describes the relation between the fullband filter mean square deviation and the mean square deviation of the individual subband filters in the FFT-2 stacking case. This relation is then used in the following section, where the proposed algorithm is derived.

The fullband filter deviation vector is defined as $\mathbf{v}(k) = \mathbf{h}_{\text{opt}} - \hat{\mathbf{h}}(k) = [v_0(k), \cdots, v_{N-1}(k)]^T$, where $\mathbf{h}_{\text{opt}}$ describes the unknown system to be estimated. It is assumed that $\hat{\mathbf{h}}(k)$ and $\mathbf{h}_{\text{opt}}$ are of equal length N. By taking the $2N$-point FFT of the adaptive filter and the optimal filter, respectively, the corresponding expressions can be obtained in the DFT-domain as $\text{DFT}_{2N}\{\mathbf{v}(k)\} = V(l) = H_{\text{opt}}(l) - \hat{H}(l)$. Defining the $2N$-point DFT of the N coefficient vector $\mathbf{v}(k)$ as

$$V(l) = \sum_{q=0}^{N-1} v_q(k) e^{-j\frac{2\pi l}{2N}q} \tag{8}$$

where $l \in \{0, \cdots, 2N\}$, and thus the inverse formula as

$$v_q(k) = \frac{1}{2N} \sum_{l=0}^{2N-1} V(l) e^{j \frac{2\pi q}{2N} l} \tag{9}$$

gives the mean square deviation (MSD) as

$$\mathcal{D}(k) = \mathbb{E}[||\mathbf{v}(k)||^2] = \mathbb{E}[\frac{1}{2N} \sum_{l=0}^{2N-1} |V(l)|^2], \tag{10}$$

according to Parseval's relation. $\mathbb{E}[\cdot]$ denotes expectation. Similarly, the MSD of subband filter m is defined as

$$\mathcal{D}_m(k) = \mathbb{E}[||\mathbf{v}_m(k)||^2] = \mathbb{E}[\frac{1}{2N_M} \sum_{l=0}^{2N_M-1} |V_m(l)|^2], \tag{11}$$

where $V_m(l)$ is the $2N_M$-point FFT of $\mathbf{v}_m(k)$, which in turn is the subband deviation vector for band m.

Now, examining the effect of the FFT-2 stacking procedure, equation (7), on the MSD, it is clear that $V(l)$ can be seen as being built up by stacked versions of $V_m(l)$. However, not all frequency coefficients of $V_m(l)$ are used to build up $V(l)$. In fact, it can be seen by examining equation (7) that only half of the coefficients of $V_m(l)$ are used. Moreover, due to the 2-times oversampling, there is a frequency overlap between the adaptive subband filters. Considering ideal subband filtering, it can be assumed that the overlapping frequency bins of two neighboring adaptive filters are approximately equal. This means that

$$\frac{1}{2} \sum_{m=0}^{M-1} \sum_{l=0}^{2N_M-1} \mathbb{E}[|V_m(l)|^2] \approx \sum_{l=0}^{2N-1} \mathbb{E}[|V(l)|^2], \tag{12}$$

and inserting equation (11) gives

$$\begin{aligned} \mathcal{D}(k) = \mathbb{E}[||\mathbf{v}(k)||^2] = \\ \frac{1}{2N} \sum_{l=0}^{2N-1} \mathbb{E}[|V(l)|^2] \approx \frac{N_M}{2N} \sum_{m=0}^{M-1} \mathcal{D}_m(k). \end{aligned} \tag{13}$$

From equation (13) it can be seen that the mean square deviation of the fullband filter is proportional to the sum of the mean square deviation of the subband filters. The proportionality constant will depend on the FFT-scaling and the amount of oversampling.

5 Proposed selective subband updating

By allowing only a subset of the adaptive subband filters to update at each instant, reduction of the computational complexity can be achieved. In this particular approach, the updating of only *one* subband filter each sample instant will be considered. The proposed scheme is shown in figure 2.

Besides the reduced computational complexity achieved through absent filter updates, the FFT-2 stacking in this case can be modified for further reduced complexity. Since only one subband filter has changed since the last subband sample, it is only necessary to compute the $2N_M$-point FFT of the corresponding filter for the stacking. Further, when constructing the fullband filter, instead of performing the $2N$-point FFT of the whole filter, it is possible to consider the difference between the fullband filter from the previous update and the currently updated fullband filter, i.e.

$$\mathbf{c}(k) = \hat{\mathbf{h}}(k) - \hat{\mathbf{h}}(k-R), \tag{14}$$

and thus in the DFT-domain

$$C(l) = \hat{H}(l) - \hat{H}_p(l) \qquad l \in \{0, \cdots, 2N-1\}, \tag{15}$$

where $\text{DFT}_{2N}\{\hat{\mathbf{h}}(k-R)\} = \hat{H}_p(l) \quad l \in \{0, \cdots, 2N-1\}$. Obviously, $C(l)$ will be 0 for all l which corresponds to an unchanged subband filter (see equation (7)). Hence, $C(l)$ will only contain N_M non-zero components, which means that the inverse FFT of $C(l)$, i.e. $\mathbf{c}(k)$, can be computed very efficiently. Finally, the updated fullband filter is obtained through $\hat{\mathbf{h}}(k) = \hat{\mathbf{h}}(k-R) + \mathbf{c}(k)$. This technique is denoted *FFT-difference stacking*, see figure 2.

For selecting which subband filter to update at each instant, a periodic selection scheme where the filters are sequentially selected in a round-robin manner, or a random selection scheme where the filters are selected randomly could be used. However, as obviously a low fullband filter deviation is desired in as few updates as possible, m should be chosen as the subband corresponding to the largest current deviation reduction. This is since the fullband mean square filter deviation is proportional to the sum of the mean square deviation of the subband filters, as given by equation (13). The general idea is similar to the multichannel reasoning in [16] and the buffering technique described in [17].

The square deviation change in one subband m, using equation (5), is given as

$$||\mathbf{v}_m(n+1)||^2 = ||\mathbf{v}_m(n) - \beta_m(n) e_m^*(n) \mathbf{x}_m(n)||^2. \tag{16}$$

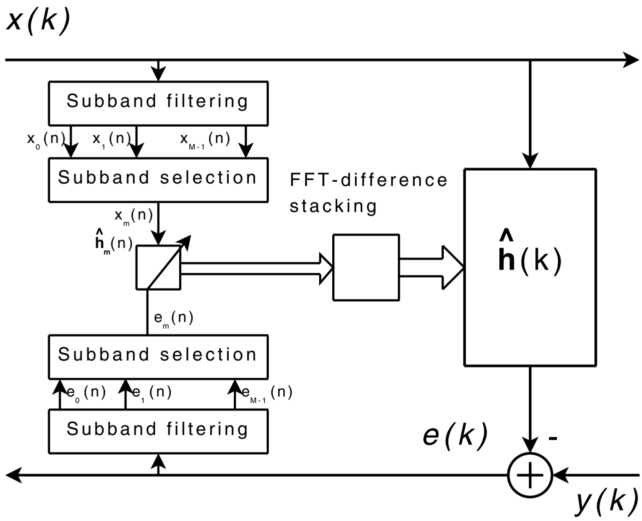

Figure 2: Proposed adaptive filtering configuration.

By inserting equation (6), assuming no regularization is used, i.e. $\epsilon = 0$, and the definition

$$t_m(n) = \mathbf{v}_m^H(n)\mathbf{x}_m(n) \tag{17}$$

into equation (16), the following expression is obtained [1]

$$\begin{aligned} ||\mathbf{v}_m(n+1)||^2 = ||\mathbf{v}_m(n)||^2 \quad + \\ \mu_m^2 \frac{|e_m(n)|^2}{||\mathbf{x}_m(n)||^2} - 2\mu_m \frac{\mathrm{Re}\{e_m^*(n)t_m(n)\}}{||\mathbf{x}_m(n)||^2}. \end{aligned} \tag{18}$$

Assuming that that $x_m(n)$ and $w_m(n)$, i.e. the driving subband signal and the local subband noise, are independent and zero mean, and using the relation

$$e_m(n) = t_m(n) + w_m(n), \tag{19}$$

the difference in mean square deviation from one update to the next is given by

$$\begin{aligned} \mathcal{D}_m(n+1) - \mathcal{D}_m(n) = \\ \mu_m(\mu_m - 2)\mathbb{E}\Big[\frac{|t_m(n)|^2}{||\mathbf{x}_m(n)||^2}\Big] + \mu_m^2 \mathbb{E}\Big[\frac{|w_m(n)|^2}{||\mathbf{x}_m(n)||^2}\Big]. \end{aligned} \tag{20}$$

Further, assuming small fluctuations in the input energy $||\mathbf{x}_m(n)||^2$ from one iteration to the next gives [1]

$$\mathcal{D}_m(n+1) - \mathcal{D}_m(n) = \mu_m(\mu_m - 2)\frac{\mathbb{E}[|t_m(n)|^2]}{\mathbb{E}[||\mathbf{x}_m(n)||^2]} + \mu_m^2 \frac{\mathbb{E}[|w_m(n)|^2]}{\mathbb{E}[||\mathbf{x}_m(n)||^2]}. \quad (21)$$

From this expression it can be seen that the first term contributes to a deviation reduction under the assumption that $0 < \mu_m < 2$. Moreover, a large $\mathbb{E}[|t_m(n)|^2]$ gives a significant deviation decrease. On the other hand, a large noise level $\mathbb{E}[|w_m(n)|^2]$ counteracts the deviation decrease and could even cause an increase.

Based on these observations and with equation (13) in mind, it is clear that updating the subband filter corresponding to the largest output error magnitude, disregarding the noise, will cause the largest possible fullband filter deviation reduction. To minimize the impact of the noise, stationary noise could be estimated in each subband and then subtracted from the output error prior to deciding which subband to update. Estimation of this noise could be done in a number of ways, e.g. using minimum statistics, or schemes with fast and slow estimators [3]. Non-stationary noise could e.g. in an echo cancellation scenario be detected by a *doubletalk detector* (DTD)[3]. The DTD would then indicate when it is safe to update the filter. Details on noise estimation is, however, out of scope for this paper.

Once the estimated subband noise level, denoted $\hat{\sigma}^2_{w_m}(n)$, is obtained, this variable can be subtracted from the squared subband error to obtain an estimate of $t^2_m(n)$. However, as a margin for minor noise estimation errors, a constant T_n is multiplied with $\hat{\sigma}^2_{w_m}(n)$ prior to the subtraction. Thus,

$$\hat{t}^2_m(n) = e^2_m(n) - T_n \hat{\sigma}^2_{w_m}(n). \quad (22)$$

From equation (22), it can be seen that the setting of T_n controls how much influence the noise is allowed to have on the selection of which subband filter to update.

Finally, which subband filter to be updated is decided through

$$i = \arg\max_m \frac{\hat{t}^2_m(n)}{||\mathbf{x}_m(n)||^2} \quad m \in \{0, \cdots, M/2\}. \quad (23)$$

Thus, the proposed algorithm could be summarized as follows

1: Estimate the noise level $\hat{\sigma}^2_{w_m}(n)$ in all subbands.

2: Before every filter update, calculate an estimate of $t_m^2(n)$ as in equation (22) for each subband, where T_n is a pre-defined constant.
3: Calculate equation (23) and update subband filter $\hat{\mathbf{h}}_i(k)$.
4: Perform FFT-difference stacking (as described in the beginning of this section) to construct an updated fullband filter.

6 Mean Square Deviation Analysis

To analytically show the benefits of the proposed algorithm, mean square deviation expressions for uncorrelated input samples for the standard NLMS, periodic NLMS and the proposed updating method are presented and compared.

6.1 Subband mean square deviation for NLMS

Considering only the deviation of a single constantly updating subband m and inserting $e_m^* = \mathbf{x}_m^H(n)\mathbf{v}_m(n) + w_m^*(n)$ and equation (6) into equation (16) and taking expectation gives

$$\mathbb{E}[\mathbf{v}_m^H(n+1)\mathbf{v}_m(n+1)] = \mathbb{E}\Big[||\big(\mathbf{I} - \mu_m \frac{\mathbf{x}_m(n)\mathbf{x}_m^H(n)}{\mathbf{x}_m^H(n)\mathbf{x}_m(n)}\big)\mathbf{v}_m(n) - \mu_m \frac{w_m^*(n)\mathbf{x}_m(n)}{\mathbf{x}_m^H(n)\mathbf{x}_m(n)}||^2\Big], \tag{24}$$

where $\mathbf{I}$ is the identity matrix with dimensions $N_M \times N_M$. Again using the assumption that $x_m(n)$ and $w_m(n)$ are uncorrelated and zero mean allows reduction to

$$\mathbb{E}[\mathbf{v}_m^H(n+1)\mathbf{v}_m(n+1)] = \mathbb{E}\big[\mathbf{v}_m^H(n)\big(\mathbf{I} - \mu_m(2-\mu_m)\frac{\mathbf{x}_m(n)\mathbf{x}_m^H(n)}{\mathbf{x}_m^H(n)\mathbf{x}_m(n)}\big)\mathbf{v}_m(n)\big] + \mu_m^2\mathbb{E}\big[\frac{|w_m(n)|^2}{\mathbf{x}_m^H(n)\mathbf{x}_m(n)}\big]. \tag{25}$$

Using the independence assumption [18], i.e. that $\mathbf{x}_m(n)$ and $\mathbf{v}_m(n)$ are independent, and assuming that the individual entries of $\mathbf{x}_m(n)$ are uncorrelated allows separate evaluation of $\mathbb{E}[\frac{\mathbf{x}_m(n)\mathbf{x}_m^H(n)}{\mathbf{x}_m^H(n)\mathbf{x}_m(n)}]$ as $\mathbb{E}[\frac{|x_m(n)|^2}{\mathbf{x}_m^H(n)\mathbf{x}_m(n)}]\mathbf{I}$ and equation (25) can be rewritten as

$$\begin{aligned}\mathcal{D}_m(n+1) = \\ &\left(1-\mu_m(2-\mu_m)\mathbb{E}\left[\frac{|x_m(n)|^2}{\mathbf{x}_m^H(n)\mathbf{x}_m(n)}\right]\right)\mathcal{D}_m(n) + \\ &\mu_m^2\mathbb{E}\left[\frac{|w_m(n)|^2}{\mathbf{x}_m^H(n)\mathbf{x}_m(n)}\right]. \end{aligned} \quad (26)$$

The assumption of small input signal energy fluctuations from one iteration to the next allows the approximation $\mathbb{E}[\frac{|x_m(n)|^2}{\mathbf{x}_m^H(n)\mathbf{x}_m(n)}] \approx \frac{\mathbb{E}[|x_m(n)|^2]}{\mathbb{E}[\mathbf{x}_m^H(n)\mathbf{x}_m(n)]}$ [1] which leads to

$$\mathcal{D}_m(n+1) = \left(1-\frac{\mu_m(2-\mu_m)}{N_M}\right)\mathcal{D}_m(n) + \frac{\mu_m^2}{N_M}\frac{\sigma_{w_m}^2}{\sigma_{x_m}^2}, \quad (27)$$

where $\sigma_{x_m}^2 = \mathbb{E}[|x_m(n)|^2]$ and $\sigma_{w_m}^2 = \mathbb{E}[|w_m(n)|^2]$. It is then obvious that by letting $n \to \infty$, the steady state deviation becomes

$$\mathcal{D}_m(\infty) = \frac{\mu_m}{2-\mu_m}\frac{\sigma_{w_m}^2}{\sigma_{x_m}^2}. \quad (28)$$

6.2 Subband mean square deviation for periodic NLMS

Now, in the case of periodic updating every $P = M/2+1$:th subband sample, the expression for the periodic NLMS becomes

$$\mathcal{D}_m(n+P) = \left(1-\frac{\mu_m(2-\mu_m)}{N_M}\right)\mathcal{D}_m(n) + \frac{\mu_m^2}{N_M}\frac{\sigma_{w_m}^2}{\sigma_{x_m}^2}. \quad (29)$$

From equation (29) it can be seen that the steady state in equation (28) also holds for the periodic NLMS, and that stability is ensured for $0 < \mu_m < 2$ just as for equation (27). It is clear that the convergence speed will be reduced with a factor P, still under the assumption of uncorrelated input samples, compared to the full updating scheme.

6.3 Subband mean square deviation for the proposed algorithm

For the proposed algorithm, the updating scheme of subband m will depend on the input signal as well as the state of the adaptive filters in the other subbands. Assuming that the subband filters will update approximately equally often, the considered filter $\hat{\mathbf{h}}_m$ will on average update every P:th subband sample, just as for the periodic approach. However, the statistical distribution of the input will be different; in the periodic case the original distribution of the input samples is maintained, but not for the proposed approach. In the event of an update of subband m, and disregarding the noise, $|e_i(n)|^2 \quad i \in \{0, \cdots, M/2\}$ is largest for $i = m$.

If it is assumed that the error contribution from the filter mismatch is significantly larger than the local noise, and the spectral content of the input signal $x_m(n)$ is essentially flat over a frequency band larger than that occupied by each element of the deviation vector $\mathbf{v}_m(n)$, the mean square error can be approximated as [1]

$$\begin{aligned} \mathbb{E}[|e_m(n)|^2] &= \mathbb{E}[|\mathbf{v}_m^H(n)\mathbf{x}_m(n)|^2] + \mathbb{E}[|w_m(n)|^2] \\ &\approx \mathbb{E}[||\mathbf{v}_m(n)||^2]\mathbb{E}[|x_m(n)|^2] \\ &= \mathcal{D}_m(n)\mathbb{E}[|x_m(n)|^2]. \end{aligned} \tag{30}$$

Further, the assumption of all subband filters on average updating equally often gives that the filter deviation of all subband filters are approximately equal, i.e. $\mathcal{D}_i(n) \approx \mathcal{D}_j(n) \quad i, j \in \{0, \cdots, M/2\}$. Thus, defining $e_{\mathrm{cur}}(n)$ as the error of the filter which is updated at the current subband sample index n, the mean square error of the filter to be updated is

$$\mathbb{E}[|e_{\mathrm{cur}}(n)|^2] = \mathcal{D}_{\mathrm{cur}}(n)\mathbb{E}[\max_i |x_i(n)|^2] \quad i \in \{0, \cdots, M/2\}. \tag{31}$$

Now, again considering the subband m, it is obvious that $\mathbb{E}[|e_{\mathrm{cur}}(n)|^2] = \mathbb{E}[|e_m(n)|^2]$ when updating subband filter m. Also, under the assumption given above, each subband filter is updated approximately every P:th subband sample. Using this, and inserting equation (31) into equation (25) yields

$$\begin{aligned} &\mathcal{D}_m(n+P) = \\ &\qquad \left(1 - \frac{\mu_m(2-\mu_m)}{N_M}\frac{\sigma_{x_\infty^2}}{\sigma_{x_m}^2}\right)\mathcal{D}_m(n) + \frac{\mu_m^2}{N_M}\frac{\sigma_{w_m}^2}{\sigma_{x_m}^2}, \end{aligned} \tag{32}$$

where $\sigma_{x_\infty}^2 = \mathbb{E}[\max_i |x_i(n)|^2] \quad i \in \{0, \cdots, M/2\}$. From equation (32) it can be seen that since obviously $\frac{\sigma_{x_\infty}^2}{\sigma_{x_m}^2} \geqslant 1$, the convergence speed of the proposed updating scheme is in general higher than for the periodic NLMS (equation (29)). It can also be seen that if $\frac{\sigma_{x_\infty}^2}{\sigma_{x_m}^2} < N_M$, the stability is ensured for $0 < \mu_m < 2$. The first condition holds under essentially all practical circumstances since typically N_M is fairly large (especially for acoustic echo cancellation) while $\sigma^2_{x_\infty}$ and $\sigma^2_{x_m}$ generally are in the same order of magnitude.

However, equation (32) only describes the the deviation of the proposed algorithm during the initial converging phase when the filter mismatch component of the squared error $|e_m(n)|^2$ (see equation (30)) is larger than the noise component, i.e. when $\mathcal{D}_m(n) > \frac{\sigma^2_{w_m}}{\sigma^2_{x_m}}$. In this case, the selection of which subband filter to update will be optimal in the sense that the subband filter which will contribute most to reducing the total deviation will be updated. After convergence, however, the selection of which subband filter to update will also depend on the noise components in the different bands. Assuming that the noise influence will disturb the subband selection so that the selection will be in a totally random manner gives that equation (29) better describes the proposed algorithm after convergence. Incorporating this property into equation (32) yields

$$\mathcal{D}_m(n+P) = \left(1 - \frac{\mu_m(2-\mu_m)}{N_M} q_m(n)\right)\mathcal{D}_m(n) + \frac{\mu_m^2}{N_M}\frac{\sigma^2_{w_m}}{\sigma^2_{x_m}}, \tag{33}$$

where

$$q_m(n) = \begin{cases} \frac{\sigma_{x^2_\infty}}{\sigma^2_{x_m}} & \text{if} \quad \mathcal{D}_m(n) > \frac{\sigma^2_{w_m}}{\sigma^2_{x_m}} \\ 1 & \text{otherwise.} \end{cases} \tag{34}$$

6.4 Reconstructed fullband filter mean square deviation

Updating one adaptive subband filter will, since the individual bands in the oversampled filterbank (shown in figure 3) are overlapping, also affect the future input error to its neighbors due to the closed-loop structure (see figure 2). This means that the convergence speed of the neighboring adaptive filters will be affected. Experiments have shown that for a high order ($K = 8192$) near-ideal 2-times oversampled filterbank with 100% frequency overlap (i.e. two filter bands overlap at every frequency), equation (13) holds fairly well. In this case, the spectrum of the signal will be fairly flat within each subband if the fullband signal spectrum is flat. However, this is not true if the filter order is changed so that the amount of overlap is decreased (e.g. as in figure 3). As the filter bands become narrower, although remaining flat in the frequency region corresponding to the components used in the FFT-2-stacking, the convergence speed of the fullband filter is increased. It seems like this amount

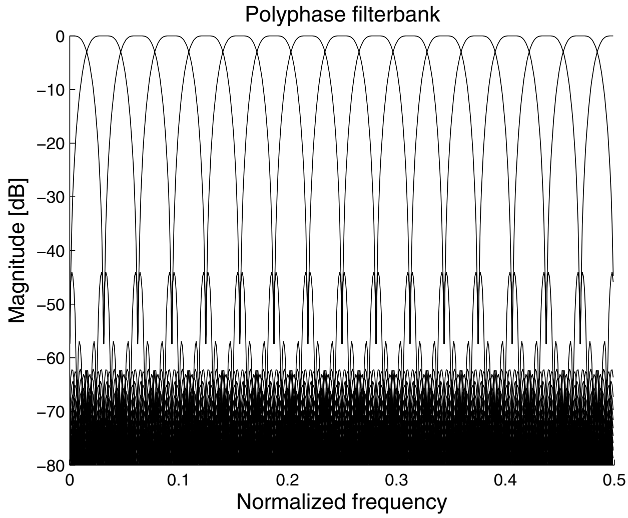

Figure 3: Frequency response of polyphase FFT filterbank.

of increase can be up to a factor of about 3/2 in practice, which should be taken account for by multiplying the right hand side of equation (13) by 2/3. Mathematical analysis of this effect is a subject of further studies.

7 Computational complexity

The delayless adaptive filtering can be divided into four steps; subband filtering, subband filter updating, FFT-2 stacking and the signal path fullband filtering. For the sake of simplicity, the computational complexity considered here is measured in multiplications and divisions per sample.

7.1 Subband filtering

For M subbands, the 2-times oversampled polyphase FFT requires one K-coefficient prototype filter convolution and one M point FFT for each set of $M/2$ input samples. An M-point real FFT requires about $M \log_2 M$ multiplications, giving a total of [6]

$$2(K/M + 2\log_2 M) \tag{35}$$

real multiplications per input sample.

7.2 Subband filter updating

Since the input signal is real, only $M/2+1$ subbands need to be considered. Updating one subband adaptive filter requires $4N_M$ multiplications and one division. In the full updating case, the number of multiplications per sample then is

$$(M/2+1)8N_M/M \tag{36}$$

and for the case where only one subband is updated, i.e. the periodic scheme and the proposed algorithm,

$$8N_M/M. \tag{37}$$

In the full updating case, $M/2+1$ real divisions are also required by the subband filter updating (see equation (5)). The proposed algorithm which updates only one subband requires only one division for updating. However, an additional $M/2+1$ real divisions are required for the subband selection in equation (23), resulting in a total of $M/2+2$ real divisions. Thus, the number of divisions per input sample is $1+2/M$ in the full updating case and $1+4/M$ for the proposed algorithm.

Furthermore, for the proposed algorithm, there are $M/2+1$ comparisons/max-operations (for finding the largest value) and $M/2+1$ multiplications (for calculating the squared errors), yielding

$$1+2/M \tag{38}$$

additional multiplications per input sample. The imposed need of noise estimation in each subband by the proposed algorithm has less importance to the complexity, since the estimation is performed when there is no input signal present (i.e. only local noise), hence not at the same time as the adaptive filter updating.

7.3 FFT-2 stacking

The $2N_M$-point complex FFT for each subband, which is originally of length N_M and zeropadded up to $2N_M$, requires $4N_M \log_2(2N_M)$ real multiplications. Since the echo path is assumed to be real, only $M/2+1$ subbands need to be considered. Thus, in the full updating case, a total of $(M/2+1)4N_M \log_2(2N_M)$ real multiplications is required for calculating all the DFT coefficients, which are stacked to form the fullband filter. The fullband filter is then calculated using an complex-to-real inverse FFT, requiring $N \log_2(2N)$ real multiplications (half of a full complex-to-real inverse FFT, since only the N first time domain filter coefficients are calculated). This gives the total number of real multiplications for the FFT-2 stacking as

$$(M/2+1)4N_M \log_2(2N_M) + N \log_2(2N). \tag{39}$$

For the periodic updating scheme and the proposed algorithm, only one subband filter is updated each subband sample. This means that only one subband filter has changed since the last sample, and it is possible to use the FFT-2-difference stacking previously described. Hence, it is only necessary to compute one $2N_M$-point real FFT for the updated subband filter, i.e. $4N_M \log_2(2N_M)$ real multiplications are required. Further, when constructing the fullband filter, the number of real multiplications needed is $N \log_2 N_M$, since only N_M components are non-zero, and only the N first coefficients are considered. The total number of real multiplications for the FFT-2-difference stacking is then

$$4N_M \log_2(2N_M) + N \log_2 N_M. \tag{40}$$

Since the FFT-stacking is relatively computationally demanding, [6] suggested to only perform this stacking every N/J input samples. This is motivated through the fact that the fullband filter cannot change much faster than the length of its impulse response. The computational complexity of the FFT-2 stacking is then

$$J\left((M/2+1)4N_M \log_2(2N_M) + N \log_2(2N)\right)/N, \tag{41}$$

which can be rewritten as

$$J\left(\log_2(2N_M) + \log_2(2N) + \frac{4N_M \log_2 2N_M}{N}\right) \tag{42}$$

real multiplications per input sample.

The FFT-2-difference stacking can of course also be performed only every N/J input samples, yielding the total number of real multiplications per input sample as

$$J\left(4\frac{N_M}{N} \log_2(2N_M) + \log_2 N_M\right). \tag{43}$$

7.4 Signal path fullband filtering

The signal path fullband filtering can be performed through *fast convolution* [19]. The fullband filter coefficients are divided into L blocks, where the first block is processed with direct convolution, resulting in N/L real multiplications per input sample and the remaining blocks are processed in the DFT-domain. The DFT-domain processing requires a $2N/L$-point real FFT of each of the remaining $L-1$ blocks, as well as one $2N/L$-point real FFT of the block of N/L input samples, $(L-1)N/L$ complex multiplications in the frequency domain and one $2N/L$-point complex-to-real inverse FFT. This gives [6]

$$N/L + 2(L-1)\log_2(2N/L) + 4(L-1) + 4\log_2(2N/L) \tag{44}$$

real multiplications per input sample, which reduces to

$$N/L + 2(L+1)\log_2(2N/L) + 4(L-1) \tag{45}$$

real multiplications per input sample.

7.5 Examples

Considering a fullband filter of length $N = 512$, $M = 32$ subbands and a prototype filter of length $K = 128$ and a decimation ratio of $R = 16$ and $J = 4$, the computational complexity in the full updating case will be $18 + 136 + 142 + 218 = 514$ (equations (35), (36), (42) and (45)) real multiplications per input sample, whereas for the case where only one subband is updated every subband sample will be $18 + 8 + \frac{33}{32} + 26 + 218 \approx 271$ (equations (35), (37), (38), (43) and (45)) real multiplications per input sample. The number of divisions per input sample for the full updating case is $\frac{33}{32}$ and $\frac{9}{8}$ (see section 7.2) for the proposed algorithm. Additionally, $\frac{33}{32}$ max-operations per sample are required for the proposed algorithm.

As can be seen, the significant difference in computational complexity between the considered methods lies in the number of multiplications per input sample. Hence, the proposed algorithm almost halves the computational complexity in this case. For $J = 16$, i.e. updating of the fullband filter after each subband update, the full updating scheme requires $18 + 136 + 568 + 218 = 940$ real multiplications per input sample, while the proposed algorithm requires $18 + 8 + \frac{33}{32} + 104 + 218 \approx 349$ real multiplications per input sample.

It is clear that in these cases, the proposed algorithm requires significantly less complexity as compared to the full updating delayless subband approach.

For further comparison, the conventional fullband (N)LMS requires $2N = 1024$ multiplications per sample. Thus, in this case the proposed algorithm requires only about one fourth of the fullband (N)LMS complexity for $J = 4$ and one third of the fullband (N)LMS complexity for $J = 16$.

8 Simulations

To verify the results obtained in the previous section, various simulations were performed. In all simulations, the sampling frequency was 8 kHz. As a first setup, an "ideal" single-input-multiple output (SIMO) setup with four different FIR-filters, each of length 512, were studied. The coefficients of the four filters were realizations of zero mean gaussian random variables with variance 1. A bandlimited flat spectrum signal was used as input (hence, independent input samples as assumed in the analysis in section 6), with zero mean and variance 1. Four independent zero mean bandlimited flat spectrum signals with variance $\sigma^2_{w_m} = 2.5 \times 10^{-3}$ were used as local noise signals. To estimate the SIMO-setup, four adaptive filters of length $N = 512$ were used. Two different updating methods were then compared, one employing the periodic updating schedule and one with the proposed updating scheme. Considering this setup in terms of four different "subbands", i.e. $m = \{0, \cdots, 3\}$, the parameters were $\sigma^2_{x_m} = 1$ and $\sigma^2_{x_\infty} \approx 2.47$ (estimated through Monte Carlo simulation). The squared deviation of one filter (or "subband" $m = 0$), calculated as $||\mathbf{v}_m(n)||^2 = ||\mathbf{h}_{m,\text{opt}} - \hat{\mathbf{h}}_m(n)||^2$ is shown in figure 4. As can be seen, the estimated deviations, obtained through equation (29) and equation (33), follow the simulated deviations fairly well. Moreover, the proposed algorithm converges faster than the periodic updating scheme, and reaches the same steady state, as expected. Behavior of filters $m = \{1, 2, 3\}$ is identical.

Next, the performance of the two updating schemes, the periodic and the proposed, were compared in an actual subband setup. In this case the number of subband were chosen as $M = 32$ with a decimation ratio of $M/2 = 16$. A band limited flat spectrum signal with zero mean and variance 1 were used as input signal, resulting in $\frac{\sigma^2_{x_\infty}}{\sigma^2_{x_m}} \approx 3.6$ (again, estimated through Monte Carlo simulation). In this case, the squared deviation of the fullband filters are compared, since the optimal subband filters are not directly available due to the delayless structure. Estimation of the fullband MSD for the periodic and the proposed updating scheme, respectively, is as described in section 6.4.

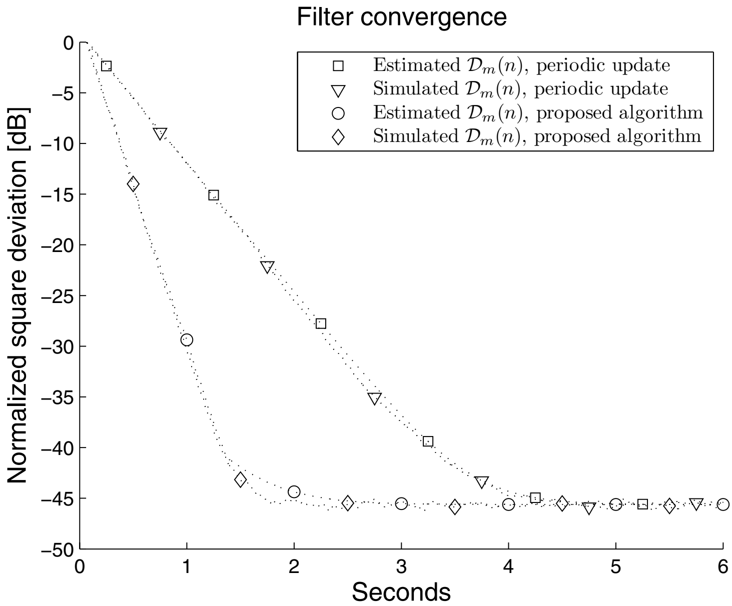

Figure 4: Comparison between the estimated and simulated squared deviation of the proposed methods and the periodic NLMS, respectively. Simulations were performed with an ideal setup, i.e. independent input samples and orthogonal filter coefficient vectors. Estimated deviations were obtained through equation (29) for the periodic updating scheme and equation (33) for the proposed algorithm.

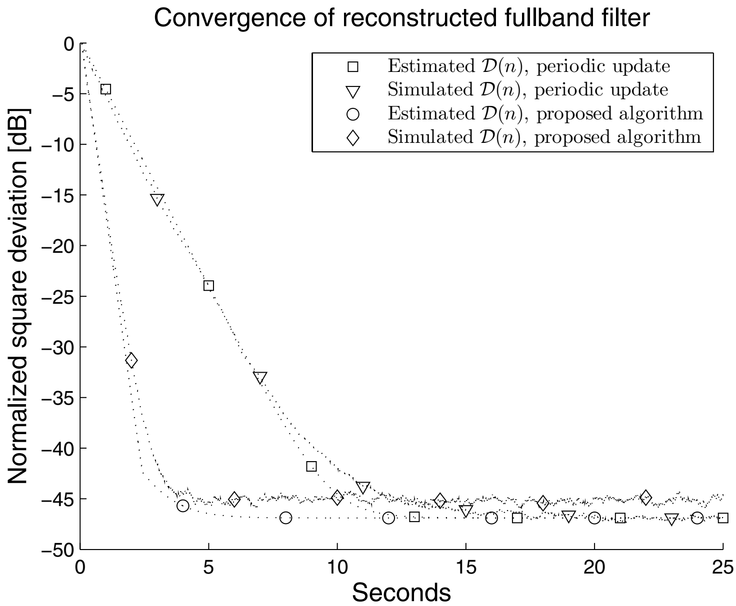

Figure 5: Comparison between the estimated and simulated squared fullband filter deviation of the proposed methods and the periodic NLMS, respectively, in a situation with a flat spectrum signal as input. 32 subbands were used, out of which only one subband filter is updated at each time instant.

Results are shown in figure 5, again showing the faster convergence of the proposed method compared to the periodic updating. However, the simulated deviation curves show a deviant behavior compared to the estimated curves. This happens below approximately −30 dB, at which point the influence of the subband filter band edges on the adaptive filter deviation start to become significant. The reduction in convergence speed is believed to be caused by the small eigenvalues associated with the band edges of the subband filters [6]. Moreover, the steady state divergence is slightly higher (about 2 dB) for the proposed method. The reason for this is not clear at the moment.

Simulations with real speech signals and real recorded stationary ambient noise with different characteristics were used in further evaluation of the proposed algorithm. Important to note is that the spectrum of the noise in this case is highly non-flat, i.e. the amount of noise in each subband is very different. The noise was recorded in a standard office and originates mainly from computer fans and air-conditioners. Parameters were as shown in table 1.

A simple voice activity detector (VAD) [3] was set to operate in each subband, assuring update only when sufficient echo-to-noise ratio in the subband. In practice, this means that $\hat{t}_m^2(n)$ for all subbands was calculated for both algorithms. Thus, if $\hat{t}_m^2(n)$ is negative, the next subband for which $\hat{t}_m^2(n)$ is not negative is updated instead for the periodic NLMS. If $\hat{t}_m^2(n)$ is negative for all subbands, no subband is updated. For the proposed algorithm, a corresponding approach implies not updating any subband if the largest $\hat{t}_m^2(n)$ is negative.

The results, in the form of squared averaged output error, calculated as $\tilde{e}(n) = \frac{1}{N_i}\sum_{i=1}^{N_i} |e(n-i)|^2$, with $N_i = 4000$ (and $\tilde{y}(n) = \frac{1}{N_i}\sum_{i=1}^{N_i} |y(n-i)|^2$) from both updating methods, are shown in the upper plot of figure 6. Clearly, it can be seen that the proposed algorithm converges faster than the periodic updating scheme also in this case. In this simulation, comparison with a full updating scheme is also presented. It could also be seen that the convergence of the proposed algorithm is comparable (only slightly slower) than the full updating scheme, albeit much lower computational complexity. In the lower plot of figure 6, the update distribution among the different subband filters for the proposed algorithm is presented. One observation that can be made from this plot is that at the very beginning (0-2 seconds), the updates are primarily concentrated to lower bands. This is since the largest output error magnitudes are originating from those subbands in this case. Then, as the lower bands converge, producing lower output error magnitudes, the higher bands have a chance to update.

This property is also shown in figures 7 and 8. The upper plot of figure 7 shows a simulation with a flat spectrum signal filtered through a bandpass filter, and the lower plot shows the update distribution among the different subband filters for the proposed algorithm in this simulation. It can be seen that the proposed algorithm initially concentrates the updates to the middle bands (bands 7 to 9), containing the largest error magnitudes at this point. The upper plot of figure 8 shows a simulation with a flat spectrum signal filtered through a bandstop filter, and the lower plot shows the update distribution among the different subband filters for the proposed algorithm in this simulation. In this case, the updates are concentrated to the uppermost and lowermost subbands, not allowing subband 8, corresponding to the notch of the stopband filter, to update. Subband 7 and 9, corresponding to the band edges, are allowed to update after about 1 second, i.e. after the other bands have reached a sufficient level of convergence.

9 Off-line calculations

Evaluation using real signals, recorded in a normal office with a loudspeaker and a microphone, were also performed. In this case, the adaptive filter length was changed to $N = 1024$ and thus $N_M = 64$ to be able to model the long impulse response of the room. Moreover, the parameters $\mu = 0.75$ and $\epsilon = 10^{-5}$ were used. The result is presented in figure 9, where the averaged square output error from the full updating, periodic and the proposed scheme is shown. Like in the previous simulations, the full updating scheme displays only slightly faster convergence compared to the proposed scheme, while the convergence of the periodic updating scheme is significantly slower.

10 Conclusions

In this paper, a method for reduced computational complexity of delayless subband adaptive filters has been presented. An analytical expression for the mean square deviation for the proposed algorithm in a situation with uncorrelated input samples has also been presented, and verified through simulations. Comparison between the proposed algorithm and a periodic updating scheme, in both artificial situations with flat spectrum signals and speech, as well as in real situations with speech signals in an acoustic echo cancellation setup, shows the advantage of the proposed algorithm in terms of convergence speed. Moreover, comparison with a full updating delayless scheme shows that the proposed solution exhibits only slightly slower convergence, while requiring only about half of the computational complexity.

Implementation of the proposed algorithm in a fixed point environment would be straight-forward, except possible implications in the balance between keeping sufficient precision and avoid saturation of the FFTs and the inverse FFTs. However, it is believed that this can be handled through appropriate (perhaps dynamic) scaling, and is a subject of further study.

References

[1] S. Haykin, *Adaptive Filter Theory*, 4th ed. Prentice-Hall, 2002.

[2] S. L. Gay and S. Tavathia, "The fast affine projection algorithm," *In proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, pp. 3023–3026, 1995.

[3] E. Hänsler and G. Schmidt, *Acoustic Echo and Noise Control: A Practical Approach*. Wiley, 2004.

[4] K. Dogancay and O. Tanrikulu, "Generalized subband decomposition LMS algorithm employing selective partial updates," *In proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, pp. 1377–1380, 2002.

[5] S. Hosur and A. Tewfik, "Wavelet transform domain adaptive FIR filtering," *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 617–630, 1997.

[6] D. Morgan and J. Thi, "A delayless subband adaptive filter architecture," *IEEE Transactions on Signal Processing*, vol. 43, no. 8, pp. 1819–1830, 1995.

[7] H. Huang and C. Kyriakakis, "Real-valued delayless subband affine projection algorithm for acoustic echo cancellation," *Conference Record of the Thirty-Eighth Asilomar Conference on Signals, Systems and Computers*, vol. 1, pp. 259–262, 2004.

[8] S. C. Douglas, "Adaptive filters employing partial updates," *IEEE Transactions on Circuits and Systems - II: Analog and Digital Signal Processing*, vol. 44, no. 3, pp. 209–216, 1997.

[9] T. Aboulnasr and K. Mayyas, "Complexity reduction of the NLMS algorithm via selective coefficient update," *IEEE Transactions on Signal Processing*, vol. 47, no. 5, pp. 1421–1424, 1999.

[10] K. Dogancay and O. Tanrikulu, "Adaptive filtering with selective partial updates," *IEEE Trans. on Circuits and Systems - II: Analog and Digital Signal Processing*, vol. 48, no. 8, pp. 762–769, 2001.

[11] T. Shertler, "Selective block update of NLMS type algorithms," *In proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, 1998.

[12] R. Brennan and H. Sheikhzadeh, "Advances in subband adaptive filtering using a low-resource oversampled filterbank implementation," *Conference Record of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers*, vol. 2, pp. 1473–1477, 2003.

[13] S. Attallah, "The wavelet transform-domain LMS adaptive filter with partial subband-coefficient updating," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 53, no. 1, pp. 8–12, 2006.

[14] S. Weiss, M. Harteneck, and R. Stewart, "On implementation and design of filter banks for subband adaptive systems," *IEEE Workshop on Signal Processing Systems*, pp. 172–181, 1998.

[15] J. Huo, S. Nordholm, and Z. Zang, "New weight transform schemes for delayless subband adaptive filtering," *Global Telecommunications Conference, GLOBECOM*, vol. 1, pp. 197–201, 2001.

[16] F. Lindstrom, C. Schuldt, and I. Claesson, "Efficient multichannel NLMS implementation for acoustic echo cancellation," *EURASIP Journal on Audio, Speech, and Music Processing*, 2007, article ID 78439, 6 pages, doi:10.1155/2007/78439.

[17] C. Schuldt, F. Lindstrom, and I. Claesson, "Low-complexity adaptive filtering implementation for acoustic echo cancellation," *In proceedings of IEEE TENCON*, November 2006.

[18] W. Gardner, "Learning characteristics of stochastic-gradient-descent algorithms: A general study, analysis, and critique," *Signal Processing*, vol. 6, no. 2, pp. 113–133, 1984.

[19] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing.* Prentice-Hall, 1989.

Parameter	Value
N	512
R	16
M	32
N_M	32
S	128
J	16
μ	0.5
ϵ	10^{-6}
T_n	0.1

Table 1: Simulation parameter settings.

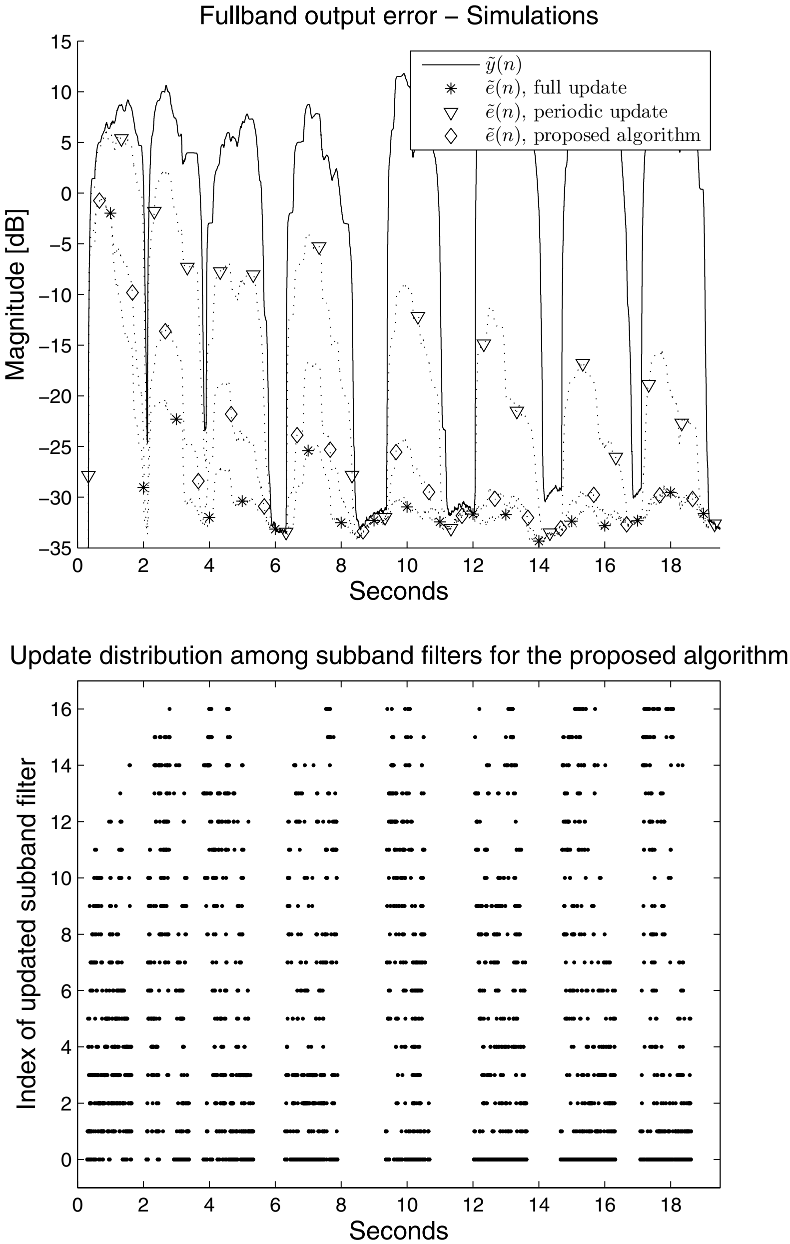

Figure 6: Upper plot shows the output error from the proposed algorithm, the periodic updating scheme and full updating, respectively, with a speech signal as input and simulated impulse response. Lower plot shows the update distribution among subband filters for the proposed algorithm. A dot denotes and update of the corresponding filter.

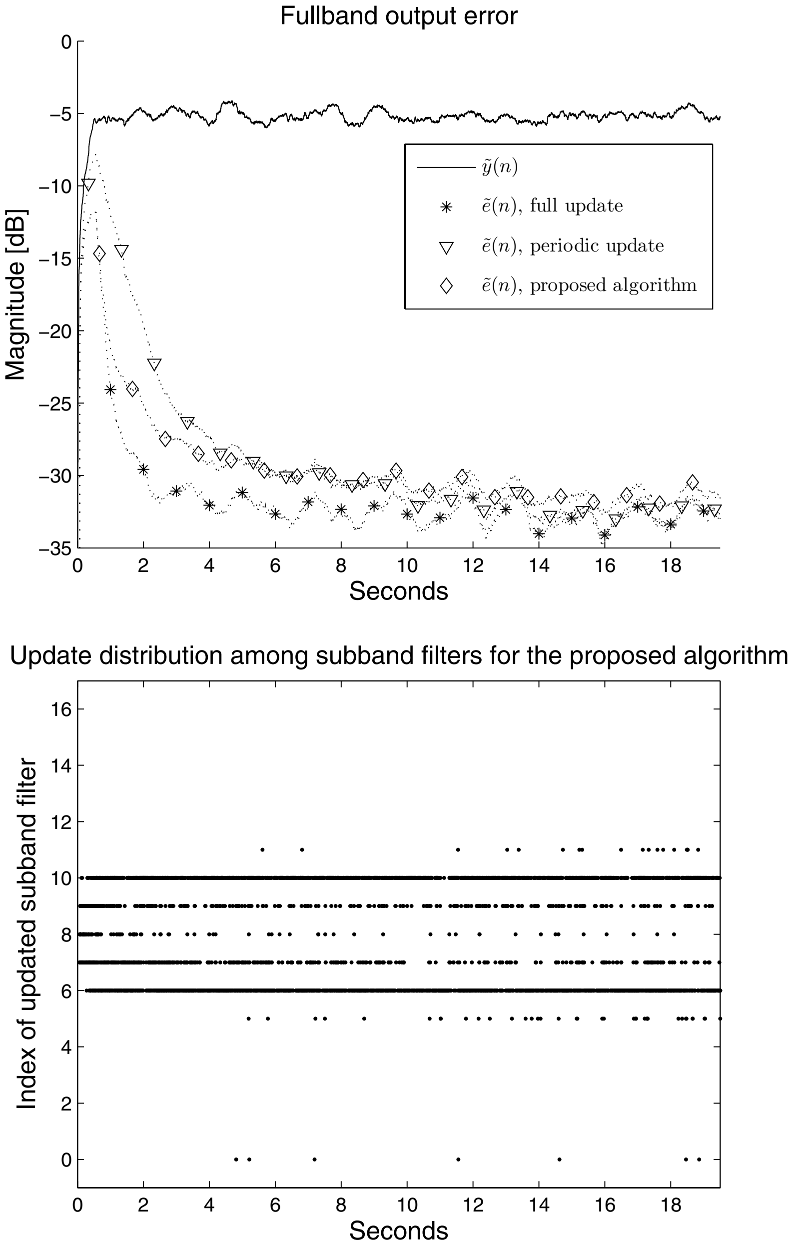

Figure 7: Upper plot shows the output error from the proposed algorithm, the periodic updating scheme and full updating, respectively, with a flat spectrum signal filtered through a bandpass filter as input and simulated impulse response. Lower plot shows the update distribution among subband filters for the proposed algorithm. A dot denotes and update of the corresponding filter.

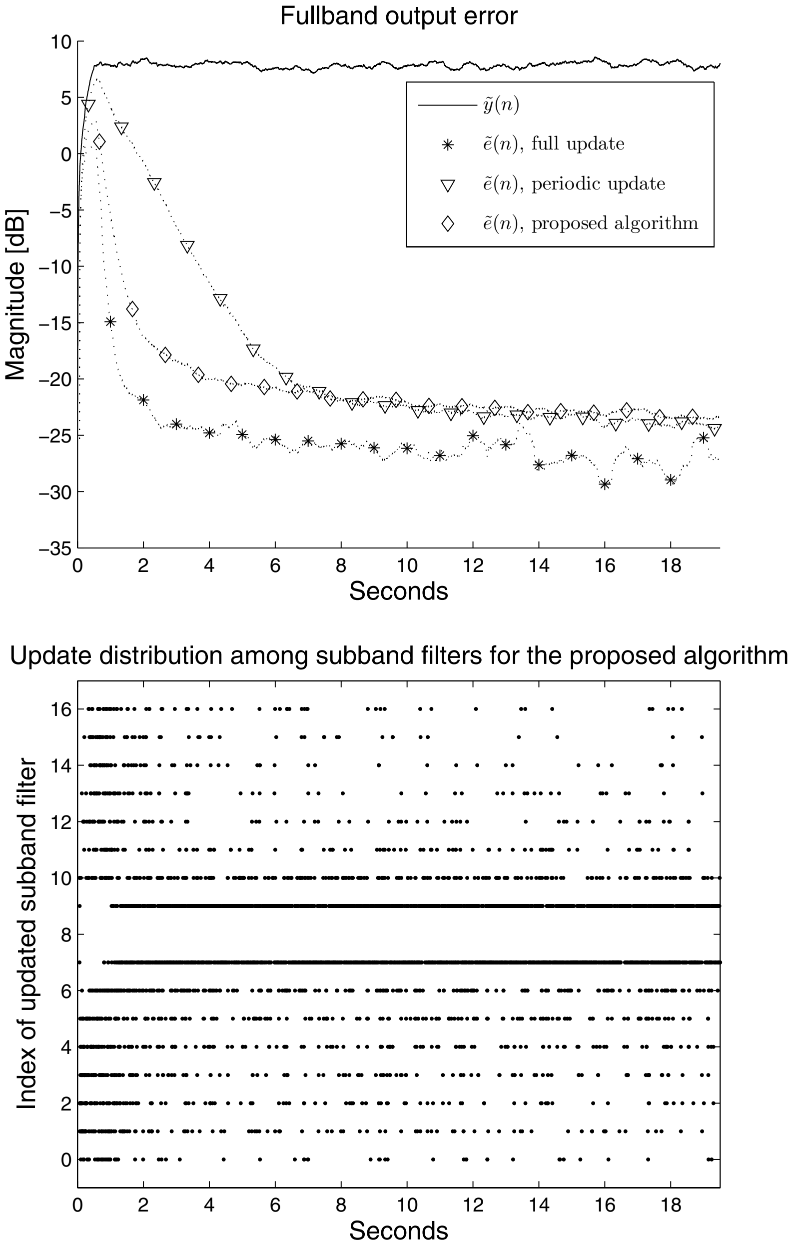

Figure 8: Upper plot shows the output error from the proposed algorithm, the periodic updating scheme and full updating, respectively, with a flat spectrum signal filtered through a bandstop filter as input and simulated impulse response. Lower plot shows the update distribution among subband filters for the proposed algorithm. A dot denotes and update of the corresponding filter.

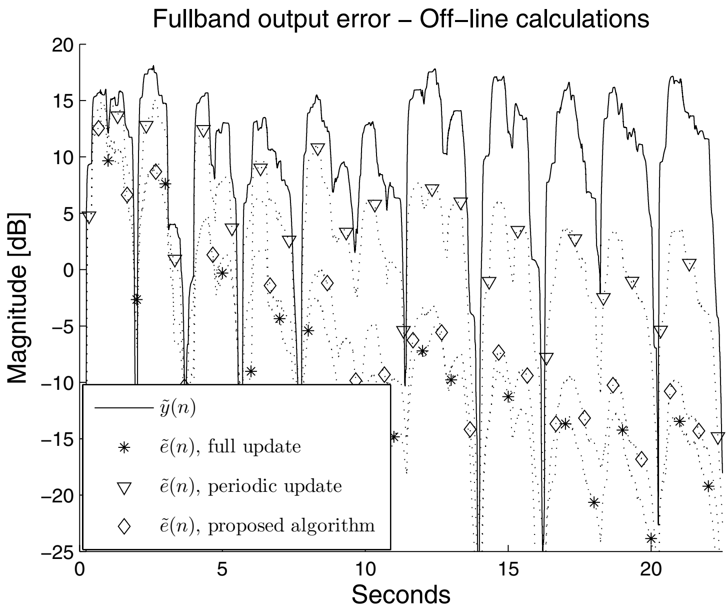

Figure 9: Output error from the proposed algorithm and the periodic updating scheme, respectively, with a speech signal as input and "real" impulse response, i.e. the signals were recorded from an acoustic setup in a normal office.

Part IV

Improving the Performance of a Low-Complexity Doubletalk Detector by a Subband Approach

Part IV is reprinted, with permission, from

Fredric Lindstrom, Christian Schüldt, Mattias Dahl, Ingvar Claesson, "Improving the Performance of a Low-Complexity Doubletalk Detector by a Subband Approach", Proceedings of IEEE ICSSD, Sousse, Tunisia, March 2005.

Improving the Performance of a Low-complexity Doubletalk Detector by a Subband Approach

Fredric Lindstrom, Christian Schüldt,
Mattias Dahl, Ingvar Claesson

Abstract

This paper presents a common framework for subband doubletalk detectors. Within this framework a number of low-complexity subband doubletalk detectors are evaluated in comparison with a corresponding fullband detector. The evaluation is performed by using real-data off-line calculations. The evaluation indicates that the subband approach significantly improves the performance.

1 Introduction

Hands-free operation is desirable in many different situations and in relation to many products, e.g. car phones, videoconference systems, conference phones, etc. In hands-free systems acoustic echoes inevitably arise. Acoustic echoes arise when the far-end speech signal produced by the loudspeaker is picked up by the microphone and transmitted back to the far-end talker [1]. Acoustic echoes are, in general, considered quite annoying. The effect of acoustic echoes can be reduced by the use of an Acoustic Echo Canceler (AEC) [1]-[3]. The performance of an AEC is linked to the estimation of certain parameters, such as speech activity, acoustic coupling between the loudspeaker and the microphone, etc [4]. The detection of speech activity, in particular doubletalk detection, constitutes a crucial task for most AEC systems. Several doubletalk detectors have been proposed, e.g. the Giegel detector [5], cross-correlation and coherence based detectors [6]-[8], and detectors using power comparison or cepstral techniques [4].

The use of subband or frequency domain based DTDs have been proposed earlier [8]-[10]. This paper proposes a general approach to subband DTDs. The approach is used to evaluate a low-complexity fullband detector in comparison with subband versions.

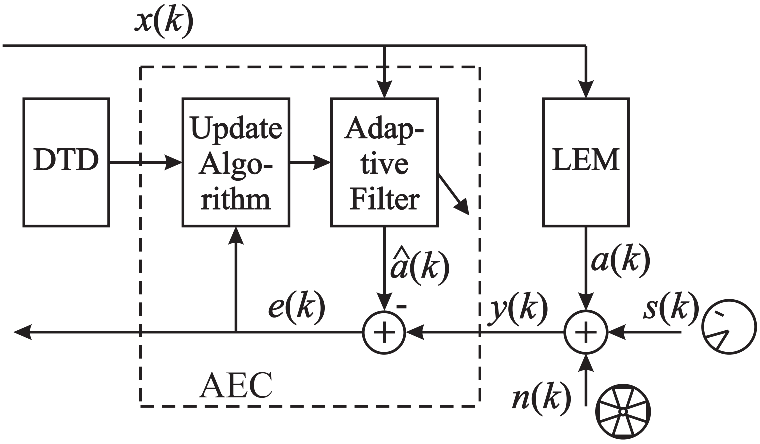

Figure 1: The AEC and its environment.

2 The Doubletalk Detection Problem

An AEC consists of an adaptive filter and an adaptive filter update algorithm, see figure reffig:aec. Commonly used update algorithms are: the Normalized Least Mean Squares (NLMS), the Recursive Least Squares (RLS), and the Affine Projection Algorithm (APA) [2]. The far-end signal $x(k)$ and the microphone signal $y(k)$ are input signals to the AEC, (k is the sample index). The microphone signal $y(k)$ consists of the acoustic echo $a(k)$, the near-end speech signal $s(k)$, and the near-end background noise $n(k)$, see figure 1. The acoustic echo $a(k)$ results from a filtering of the far-end signal $x(k)$ by the Loudspeaker-Enclosure-Microphone (LEM) system [1].

Internally calculated signals are: the estimated echo signal, $\hat{a}(k)$, and the error signal, $e(k)$, i.e. the near-end line-out signal. The purpose of the AEC is to adapt the adaptive filter in such a manner that $\hat{a}(k) = a(k)$, yielding an echo free signal $e(k)$. In the AEC, the signal $e(k)$ is used as a feed-back input to the update algorithm. If a near-end speech signal $s(k)$ exists, the adaptive filter encounter convergence difficulties, and thus an increased portion of the acoustic echo will be transferred back to the far-end talkers. If the far-end signal $x(k)$ is not present, there is no acoustic echo $a(k)$, and thus adaptation should not be done. Detecting the presence of the far-end signal $x(k)$ is quite easy since this signal is directly accessible. Therefore, it is the detection of doubletalk that is crucial, i.e. the detection of simultaneous activity in the $x(k)$ and $s(k)$ signals. The purpose of the DTD is to halt the update of the adaptive filter in situations of doubletalk.

3 Doubletalk Detection

Many proposed DTDs are single parameter detection DTDs. These detectors produce a detection parameter $\xi(k)$, which is a function of the input signals $x(k)$ and $y(k)$. The detection parameter $\xi(k)$ is compared with a threshold T; doubletalk is declared if $\xi(k) > T$. Commonly a hold feature is used, i.e. if doubletalk is declared for a sample, the detector continues to declare doubletalk for the next N_{hold} samples, no matter the value of $\xi(k)$.

Examples of single parameter detectors are: the short-term normalized correlation algorithm [4], the Geigel detector [5], the cross-correlation detector [6], and the normalized correlation algorithm [7].

4 Subband Doubletalk Detection

Subband doubletalk detection can be performed by dividing the input signals $x(k)$ and $y(k)$ into several subband signals, $\mathbf{x}_{\text{sub}}(k) = [x_0(k), \cdots, x_{N-1}(k)]$ and $\mathbf{y}_{\text{sub}}(k) = [y_0(k), \cdots, y_{N-1}(k)]$, where N is the number of subbands. For every subband a detection parameter is calculated, resulting in N parameters $\boldsymbol{\xi}_{\text{sub}}(k) = [\xi_0(k), \cdots, \xi_{N-1}(k)]$. These subband parameters can be individually modified by a function $\mathbf{g}(\cdot)$, such as a limiter, operating on each subband

$$\mathbf{g}(\boldsymbol{\xi}_{\text{sub}}(k)) = [g(\xi_0(k)), \cdots, g(\xi_{N-1}(k))]. \tag{1}$$

The modified subbands are combined into one single detection parameter $\xi(k)$ by a combination function $f(\cdot)$, i.e.

$$\xi(k) = f(\mathbf{g}(\boldsymbol{\xi}_{\text{sub}}(k))) \tag{2}$$

This combined parameter $\xi(k)$ is then compared to a threshold.

5 Combination Functions

In this section, three combination functions are proposed. These functions can be seen as generalizations of earlier proposed combination functions, e.g. [9], [10]. The proposed functions in this paper are based on the L_1, L_2, and L_∞ norms, yielding three detection parameters $\xi_{L_1}(k)$, $\xi_{L_2}(k)$ and $\xi_{L_\infty}(k)$ defined as

$$\xi_{L_1}(k) = \sum_{i=0}^{N-1} g(\xi_i(k)) \quad (3)$$

$$\xi_{L_2}(k) = \sum_{i=0}^{N-1} g(\xi_i^2(k)) \quad (4)$$

$$\xi_{L_\infty}(k) = \max_i(g(\xi_i(k))) \quad (5)$$

where i denotes the subband index.

6 Implemented DTDs

Three different subband DTD:s were implemented, denoted DTD_{L_1}, DTD_{L_2}, DTD_{L_∞} corresponding to the three combination functions presented in section 5. Further, a fullband version, $\mathrm{DTD}_{\mathrm{full}}$, was implemented in order to serve as a reference. The detection parameter used in all three DTDs is calculated by using a low-complexity method given by

$$\xi_i(k) = \frac{\overline{y}_i(k)}{\max\{\overline{x}_i(k), \cdots, \overline{x}_i(k-N_x)\}}, \quad (6)$$

where N_x is a positive integer constant, and $\overline{x}_i(x)$ and $\overline{y}_i(k)$ are smoothed magnitudes given by

$$\overline{x}_i(k) = (1-\gamma)\overline{x}_i(k) + \gamma|x_i(k)| \quad (7)$$

$$\overline{y}_i(k) = (1-\gamma)\overline{y}_i(k) + \gamma|y_i(k)| \quad (8)$$

where γ is a forgetting factor constant. The low-complexity is achieved by implementing the max function in equation (6) as a ”running” max. (The fullband detection parameter is calculated in a corresponding manner).

The performance of a fullband version of the type of DTD presented in equations (6)-(8) are generally considered inadequate [11]. This paper investigates the extent to which a low-complexity detector, such as the one defined in equations (6)-(8), can be improved by a subband approach.

The presence of a far-end speech signal is detected using the smoothed magnitude of the full-band far-end signal $\overline{x}(k)$

$$\overline{x}(k+1) = (1-\gamma_2)\overline{x}(k) + \gamma_2|x(k)|. \tag{9}$$

where γ_2 is another forgetting factor. Far-end speech is considered present when $\overline{x}(k) > T_{\overline{x}}$, where $T_{\overline{x}}$ is a threshold.

The subband filtering is performed by a uniform finite impulse response (FIR) filter bank consisting of N subbands, and all subband signals are down-sampled with a factor N_{down} using polyphase filtering [12]. Each filter has a filter order of N_{FIR}. The filter coefficients were obtained by using the Remez algorithm [13]. This implementation of the filterbank might not be computationally optimal, but was chosen since it is a well know filter design procedure.

Due to the large number of calculations performed in the evaluation, the DTDs were implemented on a digital signal processor [14].

7 Modification Functions

In this paper, three different modification functions are evaluated denoted, $\mathbf{g}_1(\cdot)$, $\mathbf{g}_2(\cdot)$, $\mathbf{g}_3(\cdot)$, defined by

$$g_1(\xi_i(k)) = \xi_i(k) \tag{10}$$

$$g_2(\xi_i(k)) = \frac{\overline{y}_i(k)\xi_i(k)}{\sum_{i=0}^{N}\overline{y}_i(k)} \tag{11}$$

$$g_3(\xi_i(k)) = \begin{cases} \xi_i(k) & \text{if } \overline{y}_i(k) > T_{\overline{y}} \\ 0 & \text{otherwise,} \end{cases} \tag{12}$$

where $T_{\overline{y}}$ is a constant threshold.

The function $\mathbf{g}_1(\cdot)$ implies that no modification of the subband detection parameters is performed. A low level of $\overline{y}_i(k)$ implies that the subband i mainly contains background noise, i.e. neither acoustic echo nor near-end speech are present in band i. The functions $\mathbf{g}_2(\cdot)$ and $\mathbf{g}_3(\cdot)$ are used to reduce the influence of such noisy subbands.

The function $\mathbf{g}_2(\cdot)$ implies that each band i is weighted with the smoothed magnitude of the near-end signal $\overline{y}_i(k)$. This function, together with the combination function in equation (3), is practically the same combination function as proposed in [10]. The function $\mathbf{g}_3(\cdot)$, implies that if a band i contains only low energy noise, i.e. if $\overline{y}_i(k) < T_{\overline{y}}$, then that band is discarded, otherwise the band is used.

8 Evaluation Method

The objective evaluation proposed in [11] is used. This method does not sufficiently evaluate the performance of the DTD in echo path change situations, i.e. in situations where the transfer characteristics of the LEM change [15]. However, for the purpose of this paper, i.e. to evaluate the improvement of doubletalk detection capability, the method is suitable.

The evaluation method is inspired by Receiver Operating Characteristics (ROC). The characteristics used are the probability of a false alarm, P_f, i.e. declaring doubletalk when doubletalk is not present, and the probability of a miss, P_m, i.e. not declaring doubletalk when doubletalk in fact is present. The procedure is as follows: for a specific preset P_f value we compute the value of P_m for a number of different levels of the Near-end speech to Acoustic echo power Ratio (NAR). This measure is defined as

$$\mathrm{NAR} = \frac{\sigma_s^2}{\sigma_a^2}, \tag{13}$$

where σ_s and σ_a are the variance of the near-end speech signal, $s(k)$, and the acoustic echo, $a(k)$, respectively. Thus, a plot of P_m vs. NAR is obtained for a specified value of P_f. From these plots visual inspection is used to judge the DTD performance. In this paper, P_f is set to $P_f = 0.1$, for details see [11].

The method proposed in [11] simulates the LEM using a FIR model of a real system. When evaluating the DTDs in this paper, off-line calculations using a real LEM system are used.

9 Results

In this section, the results of the evaluations are shown. All results shown are obtained by off-line calculations using real data. The distance between the microphone and the loudspeaker was 10 cm and the background noise was estimated to 26 dB below the acoustic echo. All settings of different parameters are given in Table 1. Care must be taken in parameter setting. A fair basic default setting is given in Table 1. Since the algorithms are implemented on a fix-point processor [14], all input signals are scaled to be in the range [-1, 1]. Further, all signals are in 8kHz sampling rate. The parameter settings in Table 1 should thus be considered in relation to this range and the sampling rate.

N_{hold}	500	$T_{\overline{x}}$	0.015	γ	0.0625
N	16	N_{down}	8	$T_{\overline{y}}$	0.005
N_x	600	N_{FIR}	64	γ_2	0.001

Table 1: Parameter values of the implemented DTDs

The result of the evaluation using modification function $\mathbf{g}_1(\cdot)$, i.e. no modification, is shown in figure 2, upper plot. It can be seen that the subband approach yields a better performance for low values of NAR, while for high values the fullband DTD has the best performance. For low values of NAR, the near-end speech signal $s(k)$ is at such a low level, as compared to the acoustic echo $a(k)$, that it is in practice undetectable by the fullband detector. However, for certain subbands the near-end speech signal can be detected. Hence, the better performance of the subband DTDs for low NARs. Subbands that contain only noise, i.e. neither near-end speech nor acoustic echoes, contribute negatively to the performance. Since the estimate parameter is obtained through a division, see equation (6), the impact of noisy subbands can be significant. When the NAR increases, the fullband DTD performance is improved. However, the negative impact on the subband DTDs from subbands containing only background noise remains, thereof the better performance of the fullband DTDs for high NARs.

In figure 2, the middle and lower plots, the result when using modification functions $\mathbf{g}_2(\cdot)$, $\mathbf{g}_3(\cdot)$ are shown. The function $\mathbf{g}_3(\cdot)$ seems to be better. The best performing subband DTD, i.e. DTD_{L_∞} in the lower plot, is for NARs from -10dB to 5dB about twice as good as the fullband DTD. This increase in performance indicates that a subband approach can make low-complexity DTDs sufficiently efficient to be used in AEC applications. These observations confirm the results indicated earlier in [9].

10 Conclusion

In this paper, a general DTD framework was presented for a class of subband DTDs. The subband DTDs were implemented on a fix-point processor and evaluated through off-line calculations. The importance of reducing the impact of noise from subbands containing neither acoustic echo nor near-end speech was demonstrated. The evaluation of the subband DTDs, in comparison with their corresponding fullband version, demonstrated that a subband approach can increase the performance of low-complexity DTDs, in order to make them interesting candidates for AEC systems.

References

[1] C. Breining P. Dreiseitel E. Hansler et. al, "Acoustic echo control," *IEEE Signal Processing Magazine*, vol. 16, no. 4, pp. 42–69, July 1999.

[2] S. Haykin, *Adaptive filter theory*, Prentice-Hall, 4th edition, 2002.

[3] J. Benesty Y. Huang, *Adaptive signal processing*, Springer, 2003.

[4] A. Mader H. Puder G. U. Schmidt, "Step-size control for acoustic cancellation filters - an overview," *Signal Processing*, vol. 80, pp. 1697–1719, 2000.

[5] D. L. Duttweiler, "A twelve-channel digital echo canceler," *IEEE Transactions on Communications*, vol. COM-26, pp. 647–653, May 1978.

[6] H. Ye B. X. Wu, "A new double talk detection based on the orthogonality theorem," *IEEE Transactions on Communication*, vol. 39, pp. 1542–1545, November 1991.

[7] J. Benesty D. R. Morgan J. H. Cho, "A new class of doubletalk detectors based on cross-correlation," *IEEE Transactions on Speech and Audio Processing*, vol. 8, pp. 168–172, March 2000.

[8] T. Gansler M. Hansson C.-J. Ivarsson G. Salomonsson, "A double-talk detector based on coherence," *IEEE Transactions on Communication*, vol. 44, pp. 1421–1427, November 1996.

[9] P. L. Chu, "Weaver ssb subband acoustic echo canceller," *IEEE Workshop on applications of signal processing to audio and acoustics*, pp. 8–11, 1993.

[10] T. Jia Y. Jia J. Ji Y. Hu, "Subband doubletalk detector for acoustic echo cancellation systems," *Proceedings of IEEE ICASSP*, pp. 604–607, 2003.

[11] J. H. Cho D. R. Morgan J. Benesty, "An objective technique for evaluating doubletalk detectors in acoustic echo cancelers," *IEEE Transactions on Speech and Audio Processing*, vol. 7, pp. 718–724, November 1999.

[12] P. P. Vaidyanathan, *Multirate systems and filter banks*, Prentice-Hall, 1993.

[13] A. V. Oppenheim R. W. Schafer, *Discrete-time signal processing*, Prentice-Hall, 1989.

[14] *ADSP-BF533 Blackfin processor hardware reference*, Analog Devices, 2003.

[15] Per Ahgren, *On system identification and acoustic echo cancellation*, Ph.D. thesis, Uppsala University, 2004.

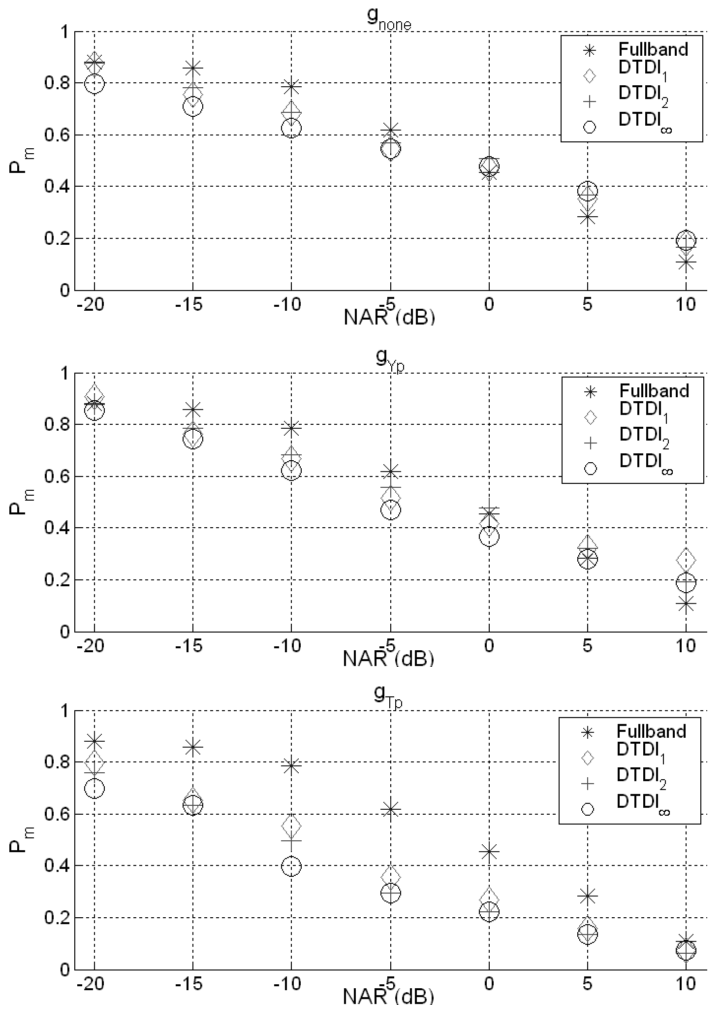

Figure 2: Results of evaluations in form of P_m vs. NAR, i.e. the probability of a miss vs. the near-end speech to acoustic echo ratio.

ABSTRACT

With the globalization of the world's economy, the demand for effortless, quick and efficient communication is increasing. Modern audio conferencing allows people at different locations to have a conversation as if they were sitting in the same room, without having to travel. This obviously saves time and money, and also lessens the environmental strain caused by travel.

Most audio conferencing systems and hands-free systems in particular, suffer from electric and/or acoustic echoes. Electric echoes typically originate from 2-4 wire conversion in hybrid circuits in the telephone network, while acoustic echoes arise due to acoustic coupling between loudspeaker and microphone. In digital audio communication equipment, the echoes are usually removed through digital signal processing methods such as adaptive filtering.

Since audio conferencing systems are consumer electronic products, the manufacturing cost is a key issue. In order to accomplish low manufacturing costs, the choice of a low cost digital signal processor (DSP) to perform the signal processing tasks is central. Further, due to the limited resources of low cost DSPs, there is an intrinsic demand for low complexity signal processing algorithms.

This thesis presents low complexity algorithms for adaptive filtering in acoustic echo cancellation applications. Both the actual update of the adaptive filter and the update control to prevent divergence and so called howling, are considered. Computer simulations, as well as real time implementations in actual acoustic systems are used to verify the performance of the proposed algorithms.

2009:01

ISSN 1650-2140
ISBN 978-91-7295-153-2