ISSN 0005–1144
ATKAAF 47(1–2), 49–58 (2006)

*Mladen Maletić, Dubravko Miljković*

# The Acoustic Echo Cancellation for Hands-free Communicator

UDK 621.395.664
IFAC 5.8.1;3.2.5

**Original scientific paper**

The control of acoustic feedback has an important function in speech processing systems, whenever the microphone picks up the audio signal radiated by loudspeaker and its reflection from the enclosure. The users are annoyed by listening to their own speech delayed by the round-trip time of the hands free communicator. This paper compares convergence properties of two FIR filter adaptation algorithms used in echo cancellation. Beside base algorithms, advanced and improved versions that include adaptive step size are considered.

**Key words:** convergence, FIR filter, adaptation, echo cancellation, feedback, LEM system

## 1 INTRODUCTION

The problem of acoustic echo cancellation arises whenever a loudspeaker and a microphone are placed in such way that the microphone is picking up the signal radiated by the loudspeaker and its reflections at the borders of the enclosure [1]. Due to acoustic feedback, the electroacustic circuit may become unstable and produce howling. Introduction of short term delay into signal path to avoid howling is unacceptable in most hands-free applications, as tolerable delay times are specified by International Telecommunication Union (ITU). With increasing more signal processing power available in recent years, application of an adaptive filter parallel to the loudspeaker-enclosure-microphone system (LEM system) is economically feasible. Stereophonic systems require independent adaptation of filtering for each channel because both channels are highly correlated and there is no unique solution for impulse responses of the echo-cancellation filters to be applied to both channels. Moving the object and/or changing the temperature results also in time-variable impulse response and hence adaptivity is of key importance in echo-cancellation applications.

## 2 PROPERTIES AND MODELING OF LEM SYSTEMS

The central elements in acoustic echo cancellation are a loudspeaker and a microphone placed within one enclosure. This system may be modeled with sufficient accuracy as a linear system. Block diagram in Figure 1 describes basic features of LEM system with Echo-Cancellation Filter (ECF) and corresponding notation.

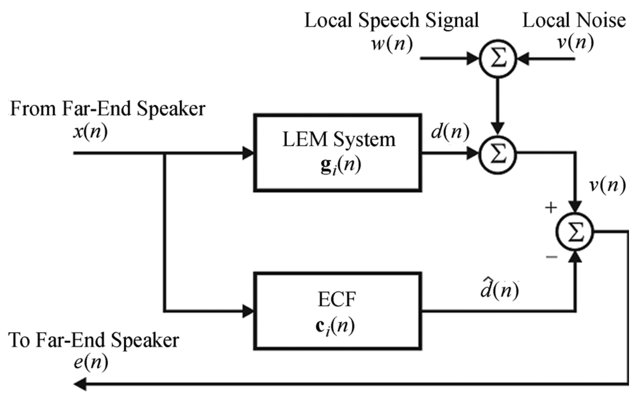


*Fig. 1 LEM System with echo-cancellation filter (ECF)*

In general, the acoustic coupling within an enclosure is formed by a direct path between the loudspeaker and the microphone, and a very large number of echo paths.

The impulse response can be described by a sequence of delta impulses delayed proportionally to the geometrical length of related path. The reverberation time of an office is typically few hundred milliseconds. These impulse responses are highly sensitive to any changes within LEM system. The distance of sound traveled between two sampling instants is of an order of few centimeters. Thus, the need for an adaptive Echo-Cancellation Filter (ECF) is evident. Long impulse response (commensurate with TR60) must be modeled that exhibits highly detailed and irregular shape and replica must offer a large number of adjustable parameters and this is the argument for preferring a FIR filter with its guaranteed stability during adaptation. FIR digital filters use only current and past input samples, and none of the previous output samples, to obtain current output sample value. A measure to express the effect of echo cancellation is the echo-return loss enhancement (ERLE) [1]:

$$\mathrm{ERLE} = 10\log\frac{E\left[d^2(n)\right]}{E\left(d(n)-\hat{d}(n)\right)^2}\,\mathrm{dB} \tag{1}$$

where $d(n)$ is the microphone output signal and $\hat{d}(n)$ describes ECF output. Denoting the impulse response of the LEM system and the filter by **g** and **c** – assuming time invariance for the moment – it follows that

$$d(n) = \sum_{i=0}^{\infty} g_i x(n-i) \tag{2}$$

and

$$\hat{d}(n) = \sum_{i=0}^{N-1} c_i x(n-i) \tag{3}$$

where $N$–1 is the degree of the FIR ECF, $g_i$ is $i$-th component of vector $\boldsymbol{c}$, and $c_i$ is $i$-th component of vector $\boldsymbol{c}$.

## 3 EXCITATION SIGNALS

The performance of adaptive echo cancellation system depends on convergence of adaptive filter and on the properties of the input and error signal. In our experiments we excite system with white noise, colored noise and speech signal. Excitation signal is always normalized (maximal amplitude touching limits ±1). White noise is produced from real random number sequence. Speech signal is characterized by nearly periodic segments, noise-like segments and pauses. Great fluctuations are common to speech signal envelope, and short term variances may differ by more than 40 dB. In our experiments we use speech signal of male talker in form of .wav file. Talker is British native speaking UK English.

Colored noise is obtained by filtering white noise with CoolEdit Pro 2 equalizer functions to adjust spectral slopes of signal in such way that PSD (Power Spectral Density) of obtained colored signal captures PSD of long term speech signal as illustrated in Figure 2.

Input signal is mixed with ambient noise that is modeled as white noise with 30 dB SNL to provide more realistic floor for error during simulations. All signals are sampled with 8 kHz and 16 bit resolution.

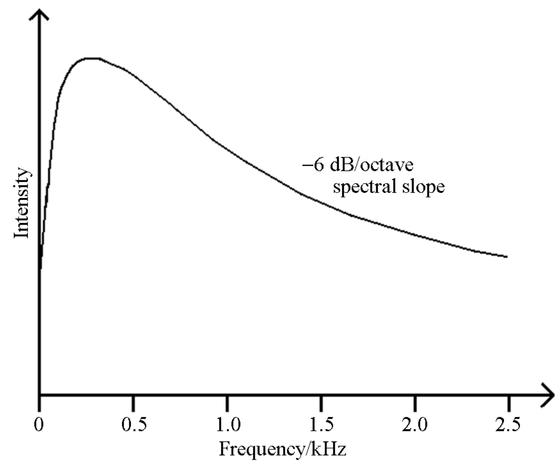


*Fig. 2 PSD of speech like signal*

Because we are mainly interested in convergence properties of algorithms and in our simplified representation of echo canceller is missing DTD (Double Talk Detector) that prevents adaptation during presence of near talker signal, we have used only ambient noise, i.e. we don't use near talker signal.

## 4 ADAPTIVE FILTER STRUCTURES

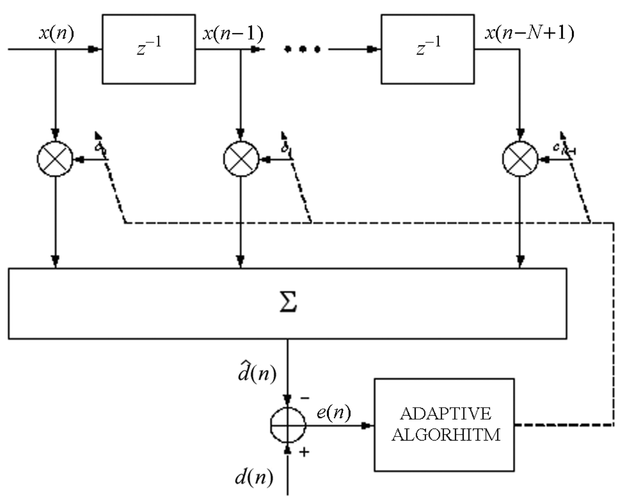


*Fig. 3 Adaptive transversal filter*

Figure 3 shows the structure of a direct-form FIR filter, also known as the tapped-delay-line or transversal filter, where $z^{-1}$ denotes the unit delay element and each $c_i(n)$ is a multiplicative gain within the system [2, 3, 4]. The echo canceller estimates the impulse response $g_i(n)$ of an echo path. An echo replica $\hat{d}(n)$ is computed by convoluting the estimated impulse response $\mathbf{c}_i(n)$ with the received input $\mathbf{x}(n)$. The echo replica is then subtracted from the real echo $d(n)$ to produce the residual echo $e(n)$.

$$e(n) = d(n) - \sum_{i=0}^{N-1} c_i(n)x(n-i). \qquad (4)$$

The error signal is fed into the procedure which alters or adapts the parameters of the filter from time $n$ to time $n+1$, as specified by the adaptive form of the adaptive LS, LMS, NLMS or AP algorithm. At each iteration $n$ the input vector includes one new input sample $x(n)$ and $N-1$ past input samples $x(n-k)$, $k=1,\ldots,N-1$ and new values of the coefficients are calculated from Equation (19). The delay element in the feedback loop, shown in Figure 4, retains the previous processed coefficient for one sampling interval.

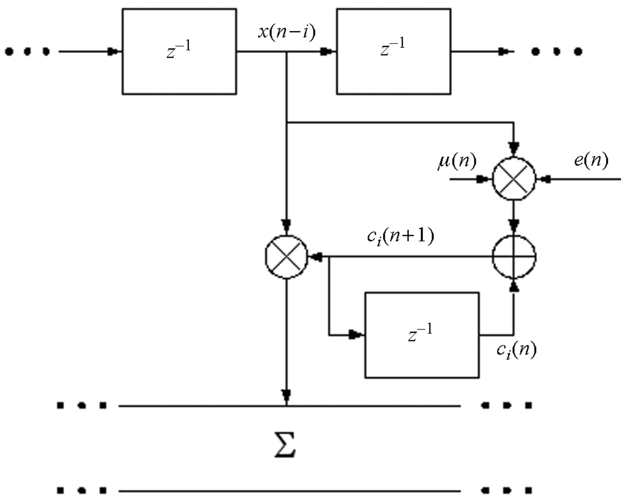


*Fig. 4 Detail of application of NLMS algorithm*

## 5 ADAPTIVE ALGORITHMS

FIR system identification setup is illustrated in Figure 5. The system model can be described by equations:

$$y(n) = \mathbf{x}_N^T(n)\mathbf{g}_N + \eta(n) \qquad (5)$$

$$\hat{d}(n) = \mathbf{x}_N^T(n)\mathbf{c}_N \qquad (6)$$

where

- $\hat{d}(n)$ is estimate for a given input data sequence $x(n)$
- $y(n)$ is the desired output sequence
- $\mathbf{x}_N$ is the regressor or data vector of dimensions $N \times 1$
- $\mathbf{g}_N$ is the filter coefficients vector of dimensions $N \times 1$
- $\mathbf{x}_N^T$ is transposed $x_N$ of dimension $1 \times N$

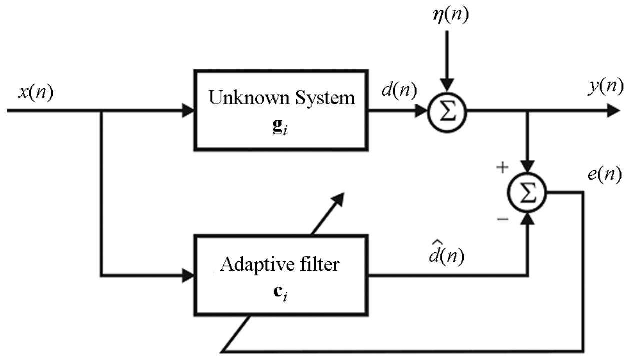


*Fig. 5 FIR systems identification setup*

- $\mathbf{c}_{Ni}$ is an estimate of the system parameters $\mathbf{g}_N$
- $\eta$ is an additional disturbance in form of near speaker talk and ambient noise:

$$\eta(n) = w(n) + v(n) \qquad (7)$$

- $w(n)$ near end speaker signal
- $v(n)$ local (ambient) noise.

The estimation error is expressed as the difference between the measured and the predicted system output.

$$e(n) = d(n) - \hat{d}(n) = d(n) - \mathbf{x}_N^T(n)\mathbf{c}_N. \qquad (8)$$

$d(n)$ is signal that propagated through LEM system (acoustic space of an office with reflections).

## 6 LEAST MEAN SQUARE (LMS) ALGORITHM

LMS (Least Mean Square) algorithm is algorithm for descending on performance surface. Detailed considerations of algorithm are given in [2] and [3]. We will concentrate mainly on algorithmic steps.

Coefficient initialization is according to Equation (9), all coefficients initialized to zero.

$$\mathbf{c}_N(-1) = 0. \qquad (9)$$

The method of steepest descent estimates the gradient $\nabla(n)$ of $E\left[e(n)^2\right]$ – and updates filter coefficients according to Equation (10):

$$\mathbf{c}_N(n+1) = \mathbf{c}_N(n) + \mu(-\nabla(n)). \qquad (10)$$

However, the method of steepest descent results in a heavy computational load and requires a large memory to calculate the expectation. When the instantaneous square error $e(n)^2$ is used instead of its expectation $E\left[e(n)^2\right]$ to estimate the gradient $\nabla(n)$ then Equaton (11) is obtained.

$$\nabla(n) = -2e(n)\mathbf{x}_N(n). \qquad (11)$$

Error term is defined as usual:

$$e(n) = d(n) - \mathbf{x}_N^T \mathbf{c}_N(n-1). \tag{12}$$

Substituting Equation (11) into Equation (10) gives Equation (13).

Equation (13) is called the LMS (Least Mean Square) algorithm

$$\mathbf{c}_N(n+1) = \mathbf{c}_N(n) + 2\mu e(n)\mathbf{x}_N(n) \tag{13}$$

where $n$ denotes discrete time index and $\mu$ is a step size.

## 7 NORMALIZED LMS (NLMS) ALGORITHM

Normalizing the adjustment vector by the norm of the input vector $\mathbf{x}(n)$ and replacing $\mu$ with $\alpha$ gives the NLMS (Normalized LMS) algorithm, as follows:

$$\mathbf{c}_N(-1) = 0 \tag{14}$$

$$e(n) = y(n) - \mathbf{x}_N^T \mathbf{c}_N(n-1) \tag{15}$$

$$\mathbf{c}_N(n) = \mathbf{c}_N(n-1) + \mu(n)\mathbf{x}_N(n)e(n) \tag{16}$$

$$\mu(n) = \frac{\alpha}{\beta + \mathbf{x}_N^T(n)\mathbf{x}(n)}, \quad \alpha \in (0,2), \ 0 \le \beta. \tag{17}$$

The algorithm is trimmed by the so-called »step size« factor $\mu(n)$. The NLMS algorithm converges in the mean to the optimum solution (16) when step-size factor $\mu$ is restricted to the interval given by equation (18) and filtering error is given by (15).

$$0 < \mu < 2. \tag{18}$$

## 8 AFFINE PROJECTION (AP) ALGORITHM

The AP algorithm is an extension of the NLMS algorithm. The optimal solution (22) is expressed in the matrix form [2, 3]. Adaptation rules are as follows:

$$\mathbf{c}_N(-1) = 0 \tag{19}$$

$$\mathbf{e}_L(n) = \mathbf{y}_L(n) - \mathbf{X}_{N,L}^T(n)\mathbf{c}_N(n-1) \tag{20}$$

$$\mathbf{R}_L(n) \equiv \mathbf{X}_{N,L}^T(n)\mathbf{X}_{N,L}(n) + \delta\mathbf{I} \tag{21}$$

$$\mathbf{w}_L(n) = \mathbf{R}_L^{-1}(n)\mathbf{e}_L(n) \tag{22}$$

$$\mathbf{v}_N(n) = \mathbf{X}_{N,L}(n)\mathbf{w}_L(n) \tag{23}$$

$$\mathbf{c}_N(n) = \mathbf{c}_N(n-1) + \mu\mathbf{v}_N(n). \tag{24}$$

Notation: $L$ is an order of AP algorithm, $\mathbf{X}_{N,L}$ input data matrix, dimensions $N \times L$; $\mathbf{y}_L(n)$ is response data matrix, dimension $L \times 1$; $\mathbf{R}_{N,L}$ is the auto-correlation matrix.

A convergence analysis can be performed similar to that of the NLMS algorithm, also resulting in condition:

$$0 < \mu(n) < 2. \tag{25}$$

AP algorithm converges much faster then NLMS and this property is extremely important when non-constant room acoustics is involved (which is always the case since people tend to move).

## 9 IMPROVEMENTS TO ALGORITHMS

Improvements to previous algorithm include two classes: time-invariant and time-variant modifications.

### 9.1 Exponentially weighted step size is introduced

Following method is an attempt to incorporate information about reverberant acoustic space to adaptation process. This improvement, called ES (exponentially weighted step size) uses a different step size for each weight of adaptive transversal filter. These step sizes are time-invariant and weighted proportional to the expected variation of a room impulse response [5]. The expected variation of a room impulse response becomes progressively smaller along the series by the same exponential ratio as the impulse response energy decay. As a result, the method adjusts coefficients with large errors in large steps, and coefficients with small errors in small steps. This method is also suitable for multiple DSP structures.

Instead of single $\alpha$ parameter, a step size matrix **A** with diagonal form is introduced:

$$A = \begin{pmatrix} a_1 & & & & 0 \\ & a_2 & & & \\ & & \cdots & & \\ & & & \cdots & \\ 0 & & & & a_N \end{pmatrix} \tag{26}$$

and substitute $\alpha$ in (17). In similar way matrix **A** replaces $\mu$ in (29).

Matrix elements are as follows:

$$a_i = a_0\gamma_i \quad a_0 = 2 \quad \text{and} \quad i = 0,\ldots, N\text{–}1 \tag{27}$$

where is

$$\gamma = \exp\left(-6.9\frac{T_S}{T_R}\right). \tag{28}$$

Exponential term uses constant described bellow:

$$-6.9 = \log_e(10^{-3}). \tag{29}$$

Two time parameters are: $T_S$ is sample period, $T_R$ is reverberation time TR60.

All above parameters combined give as exponentially decayed function that reassembles MLS Schroeder Plot derived from impulse responses as shown in Figure 6 and Figure 7. More abut reverberation and measurements one can find in [6].

Diagonal elements of **A** matrix are illustrated on Figure 8.

For this method to be effective filter must be long enough to include $T_R$ period. With shorter filter diagonal elements of matrix **A** would follow only beginning part of exponential decay, and improvement would be diminished. Also, filter that cover period shorter than $T_R$ can introduce aliasing problem. When tried with short filters ($N = 256$ and

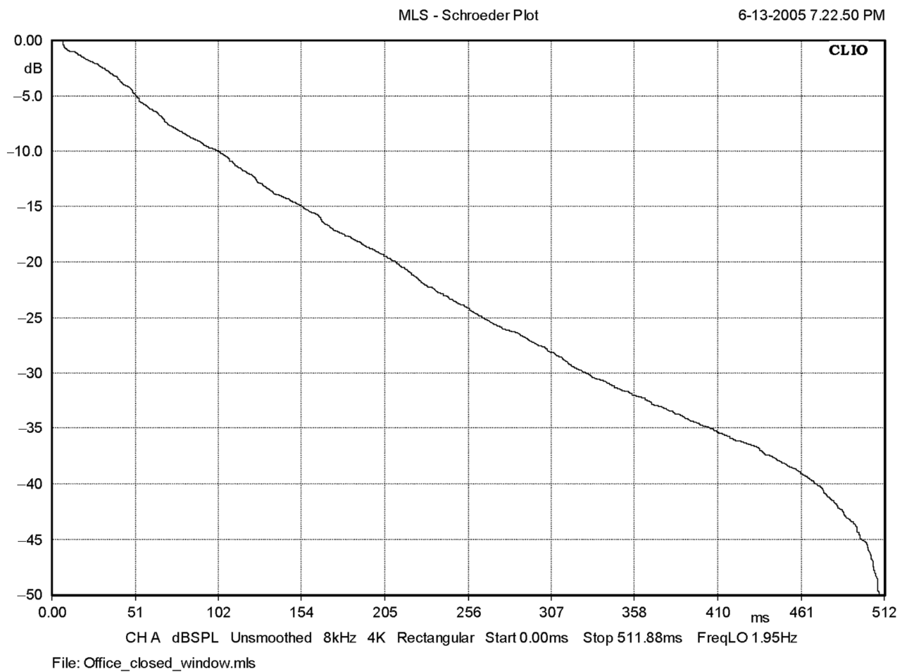


*Fig. 6 MLS Schroeder Plot – Office with closed window*

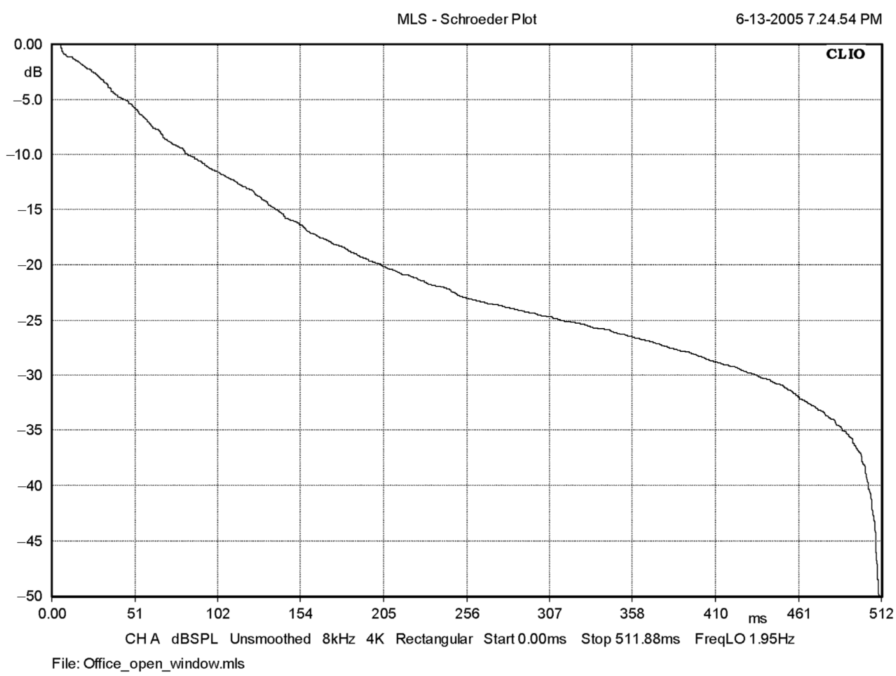


*Fig. 7 MLS Schroeder Plot – Office with open window*

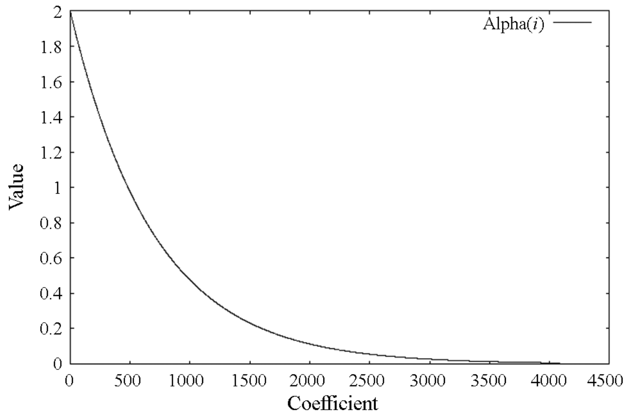


*Fig. 8 Exponentially weighted adaptation step size for appropriate weight vector coefficients*

$N=512$ we obtained only marginal improvements). Good result have been achieved with filters of 4096 taps (in accordance with sample period $T_S$ and reverberation time RT60, for our office it is about 600 ms).

### 9.2 Time varying adaptive step size

Adaptation step size is modified by correction factor $\lambda$ that depends on error rate of algorithm.

The basic idea is following: if absolute value of errors between two subsequent steps is decreasing, correction factor $\lambda$ increases adaptation step size, if absolute value of errors between two subsequent steps is increasing correction factor $\lambda$ decreases adaptation step by. This idea is further improved by specifying number of subsequent steps $N_{up}$ and $N_{down}$ when previous error condition is satisfied before correction factor is modified. To prevent correction factor to become too large or too small, limits to procedure are added: $l_{up}$ as upper limit and $l_{down}$ as upper limit.

$$\lambda = \max(l_{down}, \lambda\kappa_{down}) \tag{30}$$

if

$$|e(n)| > |e(n-1)| \tag{31}$$

for more then $N_{down}$ consequent steps.

$$\lambda = \min\left(l_{up}, \frac{\lambda}{\kappa_{up}}\right) \tag{32}$$

$$\lambda = \lambda\kappa_{up} \tag{33}$$

if

$$|e(n)| < |e(n-1)| \tag{34}$$

for more then $N_{up}$ consequent steps. Once the correction factor $\lambda$ has been updated, counters for $N_{up}$ and $N_{down}$ are reseted to zero, and new count begins. In our experiments we have achieved best results with following selection for previous parameters:

$\kappa_{down} = \kappa_{up} = 1.075$
$N_{down} = N_{up} = 12$

Following values have been chosen for limits:

$l_{up} = 2.0$
$l_{down} = 0.1$

When used together with exponentially weighted step size (EWSS), additional constraint is limiting product of $a_0\gamma^i$ and $\lambda$:

$$a_i = \min(2.0, a_0\gamma^i\lambda). \tag{35}$$

By introducing previous limit we enforce condition for algorithm convergence:

$$0 < a_i < 2. \tag{36}$$

For NLMS algorithm application of this rule is straightforward. AP algorithm has error vector $\mathbf{e}_L$ and we compare lengths of error vector $\mathbf{e}$ between two subsequent steps instead of scalar values $e$ as in NLMS. Conditions are now:

$$|\mathbf{e}_n| > |\mathbf{e}_{n-1}| \tag{37}$$

and

$$|\mathbf{e}_n| < |\mathbf{e}_{n-1}|. \tag{38}$$

The remaining steps are same as within improved NLMS algorithm. As with NLMS algorithm, when used together with EWSS same additional constraint as discussed above must be applied to ensure convergence. Convergence of algorithms also depends on choice of step size adaptation parameters.

## 10 IMPULSE RESPONSE

The two impulse responses are measured in small office (approx. $4.0 \times 3.0 \times 3.2$ m). First impulse response is taken with closed window, and second with open window. The impulse response of the LEM system was measured with MLS method using Audiomatica CLIO system, 8 kHz sample rate and MLS length $N=4096$ (same as adaptive filter length). This MLS length is comparable to RT60 of an office and alleviates possible aliasing problems [6]. The LEM system is modeled with the help of two measured impulse responses. We start by adapting filter to the first impulse response, Figure 9, and after 5 seconds switch to room model of second impulse response Figure 10. We decided to test all algorithms with filter lengths $N=4096$. All simulations are performed with sample and clock rate of 8 kHz. In order to obtain the highest convergence speed, the best choice for the step size is one.

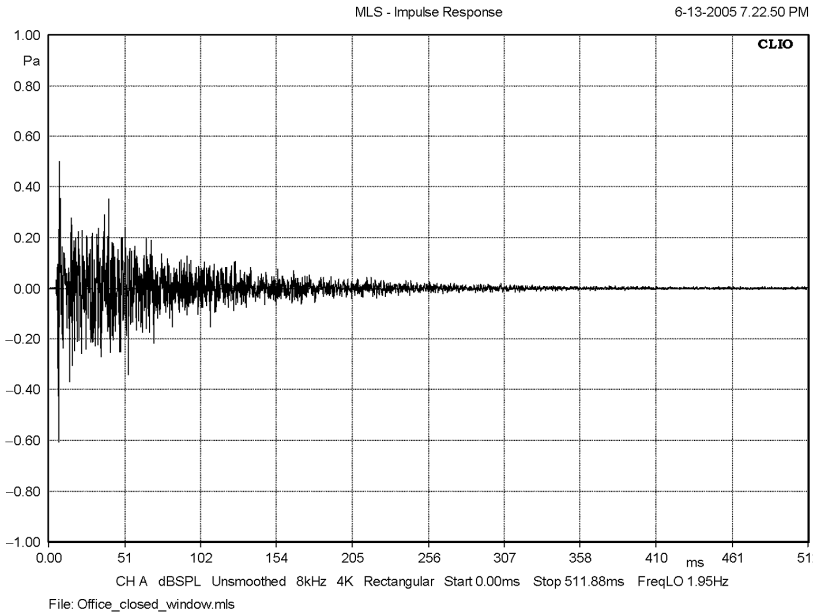


*Fig. 9 Impulse response – Office with closed window*

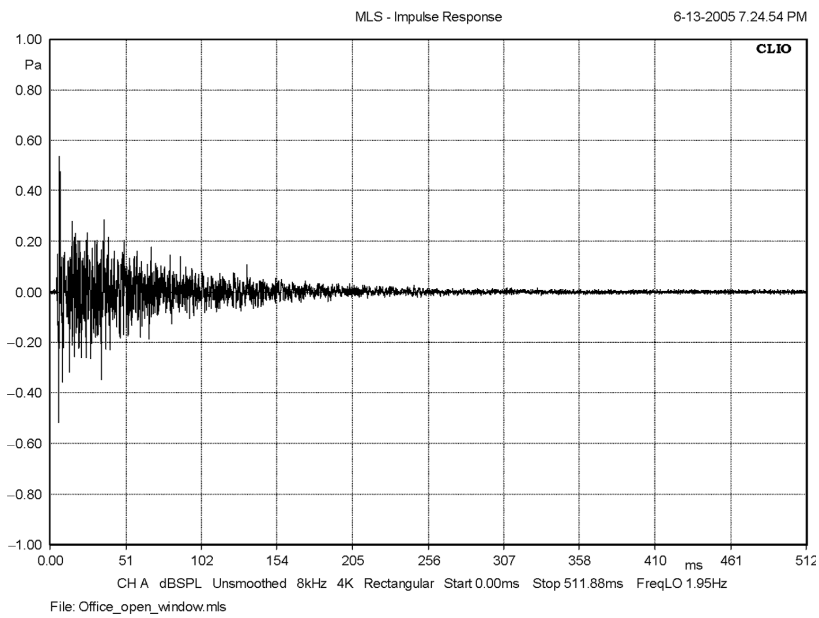


*Fig. 10 Impulse Response – Office with open window*

**11 CONVERGENCE OF THE ALGORITHMS**

Convergence of previous adaptive algorithms has been compared with the use of the following system error norm [1]:

$$10\log(\|\mathbf{c}_N(n)-\mathbf{g}_N(n)\|^2) \tag{39}$$

which is the squared norm of the difference between the LEM model $\mathbf{g}(n)$ and the adaptive filter $\mathbf{c}(n)$. Convergence of NLMS and AP algorithm was tested with white noise, colored noise and nonstationary speech excitation (male voice). AP algorithm used had an order of $L=2$.

For NLMS algorithm convergence properties are shown in Figure 11 (white noise), Figure 12 (colored noise) and Figure 13 (male speech).

Convergence properties of AP algorithm are shown in Figure 14 (white noise), Figure 15 (colored noise) and Figure 16 (male speech).

Each figure consists of three curves, one for conventional version of an algorithm (fixed $\mu$), one for exponentially weighted steps sizes (EWSS) and one for EWSS together with our improvement with time varying step size dependent on error sequence his-

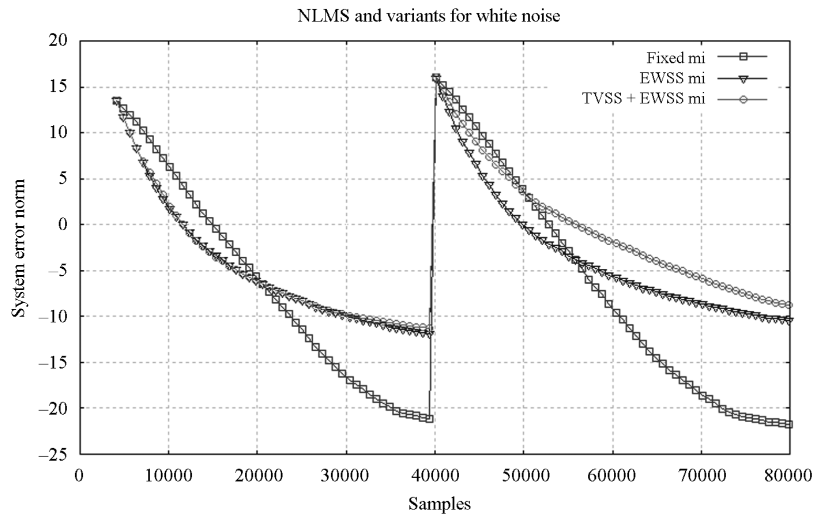


*Fig. 11 NLMS and variants – white noise*

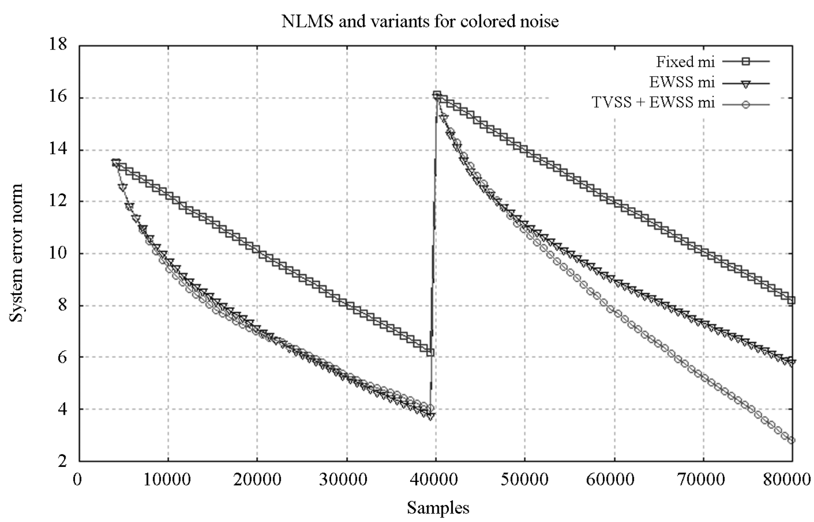


*Fig. 12 NLMS and variants – colored noise*

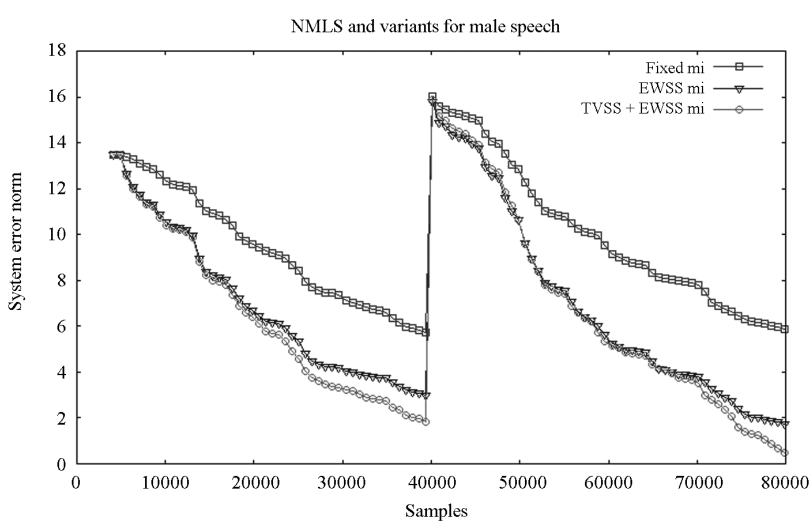


*Fig. 13 NLMS and variants – male speech*

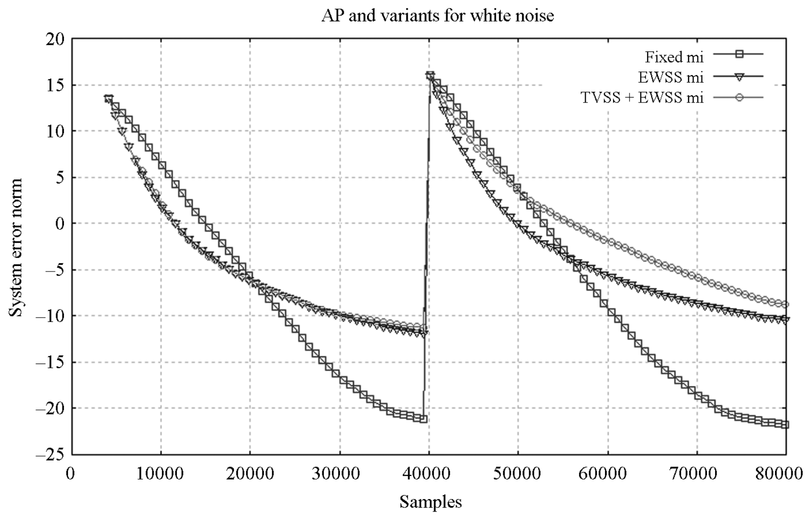


*Fig. 14 AP and variants – white noise*

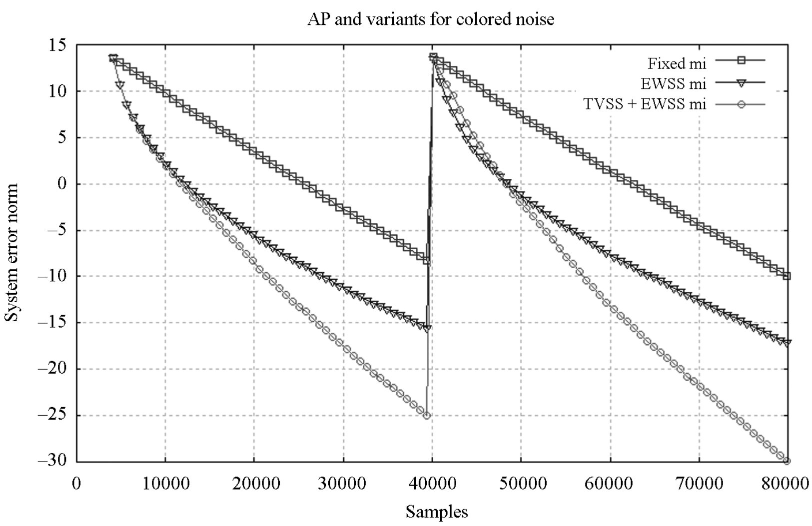


*Fig. 15 AP and variants – colored noise*

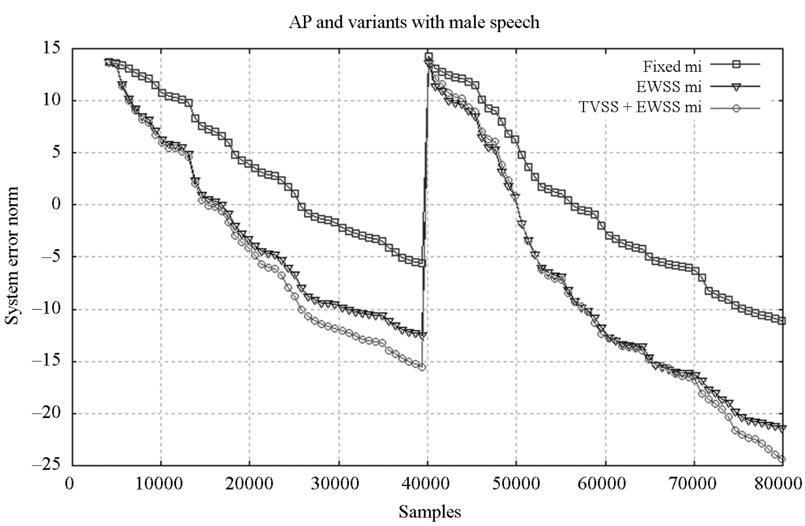


*Fig. 16 AP and variants – male speech*

tory (EWSS+TVSS). For convergence behavior, the initial adjustment of the adaptive filter to the LEM impulse response is evaluated. Sudden switch to new impulse response help us simulate tracking behavior of an algorithm. Tracking behavior indicates how well the adaptive filter can follow room-impulse response changes. Behavior of NLMS algorithm and AP algorithm is almost same for white noise excitation and no remarkable difference is visible. The advantage of the AP algorithm compared to the NLMS algorithm becomes obvious for colored excitation signal. The colored signal exhibits power spectral density comparable to the speech signal. In contrast to very slow converging NLMS algorithm, the second-order AP algorithm adapts much faster (Figure 15). For the speech excitation the results indicate that the convergence speed of the NLMS algorithm is approximately comparable to its convergence speed for colored noise excitation with PSD comparable to speech (Figure 12 and Figure 13). With white noise applied to AP algorithm adaptation is much more slower compared to its adaptation speed with colored excitation.

## 12 CONCLUSION

The control of acoustical echoes of Loudspeaker-Enclosure-Microphone (LEM) system is an important function in speech processing systems, such as hands-free mobile telephone, speakerphone, audio teleconference system, etc. Since the impulse response of the LEM system varies with time, a parallel adaptive filter is used to estimate impulse response of the LEM system. We have described an adaptive transversal filter structure because it is capable of real-time operation, guarantees stability and can use a well known adaptive algorithm for tracking the varying characteristics of the echo path, i.e. provide real-time operation, fast convergence speed.

The performance of an adaptive filter depends crucially on the properties of the input and the error, but also on adaptation algorithm and step size adaptation method. Two adaptive step size methods, both time-invariant and time-variant have been tested and proved valuable. Compared to white or colored noise excitation, the adaptation performance of the NLMS and AP algorithm in case of speech excitation is rather poor, due to strong correlation of speech signals. One way of overcoming this problem is to prewhiten or decorrelate the incoming excitation signal before passing it to the adaptive algorithm. In summary, the NLMS algorithm converges very slowly for correlated excitation signals. Therefore, the AP algorithm leads to much better results. Use of step size modifications further improves convergence.

Step size modification methods should be used in combination where each method exploits some hidden information. EWSS use characteristics of reverberant acoustic space (exponential decay of Schroeder's curve), and our TVSS exploit statistics of error sequence during adaptation.

## REFERENCES


[1] C. Breining, P. Dreiseitel, E. Hansler, A. Mader, B. Nitsch, H. Puder, T. Schertler, G. Schmidt, J. Tilp, **Acoustic Echo Control.** IEEE Signal Processing Magazine, July 1999.

[2] G. O. Glentis, K. Berberidism, S. Theodoridis, **Efficient Least Squares Adaptive Algorithms for FIR Transversal Filtering.** IEEE Signal Processing Magazine, July 1999.

[3] B. Widrow, **Adaptive Signal Processing.** Prentice Hall, 1986.

[4] L. R. Rabiner, R. W. Schafer, **Digital Processing of Speech Signals.** Prentice Hall, 1978.

[5] S. Makino, Y. Kaneda, **Exponentially Weighted Step Size NLMS Adaptive Filter Based on the Statistics of a Room Impulse Response.** IEEE Trans. on Speech and Audio Processing, Vol. 1, No. 1, January 1993.

[6] A. Everest, **Master Handbook of Acoustics.** 4th Edition, McGraw-Hill, 2001.



**Poništavanje akustičke jeke pri »hands-free« komunikaciji**. Kontrola akustičke povratne veze ima važnu ulogu u sustavima koji procesiraju govor, a pritom mikrofon prima audio signale reproducirane pomoću zvučnika kao i refleksije okolnog prostora. Korisnici su ometani pri slušanju jer čuju svoj vlastiti govor zbog vremenskog kašnjenja povratnog signala pri uporabi »hands-free« komunikatora. U članku su uspoređena konvergencijska svojstva dvaju adaptivnih algoritama za FIR filtre koji se upotrebljavaju u sustavima za potiskivanje odjeka. Osim osnovnih algoritama razmatrane su i naprednije verzije koje uključuju adaptivnu veličinu koraka.



**Ključne riječi:** konvergencija, FIR filtar, ugađanje, poništenje jeke, povratna veza, LEM sustav



**AUTHORS' ADDRESSES**

**Mladen Maletić**
**Dubravko Miljković**

**Faculty of Electrical Engineering and Computing**
**University of Zagreb**
**Unska 3, 10000 Zagreb, Croatia**


Received: 2006-03-30