ON SEMI-BLIND ESTIMATION OF ECHO PATHS DURING DOUBLE-TALK BASED ON NONSTATIONARITY

Zbyněk Koldovský[a], *Jiří Málek*[a], *Michael Müller*[a], *and Petr Tichavský*[a,b]

[a]Faculty of Mechatronics, Informatics, and Interdisciplinary Studies, Technical University of Liberec, Studentská 2, 461 17 Liberec, Czech Republic.
[b]Institute of Information Theory and Automation, Pod vodárenskou věží 4, P.O. Box 18, 182 08, Praha 8, Czech Republic.

ABSTRACT

The estimation of a filter that determines an echo path is a difficult problem when double-talk is present. We use a Blind Source Separation model based on signals' nonstationarity with a partially known mixing matrix to estimate the filter during the double talk. A second-order approximation of a log-likelihood function is used to derive a quadratic criterion. Then, we propose two methods to estimate the filter through minimizing the criterion. The first method works in a batch-offline regime, while the second method is designed to estimate the filter adaptively. The methods are verified by experiments and compared with standard least-squares-based approaches.

Index Terms— Echo cancellation, relative transfer functions, blind source separation

1. INTRODUCTION

Echo cancellation is a channel identification problem where the goal is to estimate filter $a(n)$ when observing a noise reference signal $r(n)$ and a mixture given by

$$x(n) = s(n) + \{a * r\}(n). \tag{1}$$

Here $*$ denotes the convolution, the second term in (1) is called the echo, and $s(n)$ will be referred to as the target signal. When $\widehat{a}(n)$ is an estimate of $a(n)$, $s(n)$ can be extracted from $x(n)$ as

$$\widehat{s}(n) = x(n) - \{\widehat{a} * r\}(n) = s(n) + \{(a - \widehat{a}) * r\}(n), \tag{2}$$

where the last term is the residual echo in $\widehat{s}(n)$.

This paper is focused on acoustic echo cancellation (AEC), although the conclusions could be generalized to more problems of this kind. In AEC, typically, $r(n)$ is a signal played by a loudspeaker (e.g. a radio broadcasting) while $x(n)$ is a mixed observation of a speech signal $s(n)$ and of the radio signal modified by an acoustic path $a(n)$ between the loudspeaker and microphone. The reference signal $r(n)$ can be captured through a line connection with the radio.

Since $s(n)$ is independent of $r(n)$, the mean-squared value of $\widehat{s}(n)$ evaluated on a sufficiently long interval is approximately equal to the sum of mean-squared values of $s(n)$ and of the residual echo. Therefore, $a(n)$ can be searched as the one filter that minimizes the mean-squared value of $\widehat{s}(n)$.

However, since $a(n)$ could be changing in time, adaptive estimation of the filter coefficients is desired. Adaptive LMS or RLS algorithms [1] work well unless $s(n)$ is active; see e.g. [2]. During situations where $s(n)$ is active, referred to as double-talks, the adaptive filters fail to converge. This leads either to failings to remove the echo from $\widehat{s}(n)$ or to a distortion of the target signal in $\widehat{s}(n)$ since the filters also remove a portion of its energy.

To avoid the above problem, double-talk detection is used to stop the adaptation during double-talks; see e.g. [3, 4] and references therein. However, when $a(n)$ is simultaneously changing, the echo remains present in $\widehat{s}(n)$. Recently, Gunther proposed a method based on Blind Source Separation (BSS) to estimate $a(n)$ during double-talks [5]. The method processes a batch of recording that is divided into blocks and, in one iteration, it performs a weighted least mean-square estimation of $a(n)$. The weights are chosen to maximize a log-likelihood function derived for a BSS model of signals assuming their block-wise stationarity [6].

In this paper, we follow the ideas of Gunther and derive different estimators based on the maximum likelihood principle. Using a second-order approximation of the log-likelihood function, a criterion that is a quadratic function of coefficients of $a(n)$, is derived. Then, two methods to estimated $a(n)$ using the criterion are proposed. The first one works in the batch-offline regime, similarly to the Gunther's estimator. The second method is designed to estimate $a(n)$ continuously, in an adaptive way.

In the following section, the problem of estimating $a(n)$ is formulated as a semi-blind identification problem. The quadratic criterion for the estimation is derived in Section 3

This work was supported by the Czech Science Foundation through Project No. 14-11898S.

978-1-4799-6808-4/14/$31.00 ©2014 IEEE

and the two estimators are proposed. Results of experimental evaluations are presented in Section 4, where the proposed estimators are compared with classical least-squares based methods and with the one by Gunther.

2. SEMI-BLIND SOURCE SEPARATION MODEL

Within this section, we assume that $a(n)$ does not change in time and has a finite length L. Using a vector notation, (1) can be written as

$$x(n) = s(n) + \mathbf{a}^T\mathbf{r}(n), \tag{3}$$

where $\mathbf{r}(n) = [r(n), r(n-1), \ldots, r(n-L+1)]^T$ and $\mathbf{a} = [a(1), \ldots, a(L)]^T$.

Following [5], the problem to estimate $\mathbf{a}$ can be formulated through an instantaneous mixing model

$$\mathbf{y}(n) = \mathbf{A}\mathbf{u}(n), \tag{4}$$

where $\mathbf{y}(n) = [\mathbf{r}^T(n)\, x(n)]^T$, $\mathbf{u}(n) = [\mathbf{r}^T(n)\, s(n)]^T$, and

$$\mathbf{A} = \begin{pmatrix} \mathbf{I}_L & \mathbf{0}_{L\times 1} \\ \mathbf{a}^T & 1 \end{pmatrix} \tag{5}$$

is an $(L+1) \times (L+1)$ square mixing matrix; $\mathbf{I}_L$ and $\mathbf{0}_{L\times 1}$ denote, respectively, the identity matrix and the zero vector of corresponding dimensions. Here, $\mathbf{u}(n)$ has two independent components $\mathbf{r}(n)$ and $s(n)$, one of which, $\mathbf{r}(n)$, is multidimensional [7]. $\mathbf{A}$ is partially known, so the estimation of its unknown part $\mathbf{a}$ from $\mathbf{y}(n)$ poses a *semi-blind* estimation problem.

As in [5], the block-wise stationary Gaussian model of signals is considered, which was shown appropriate for BSS [6, 8]. This means that $\mathbf{u}(n)$, $n = 1, \ldots, N$ are assumed to be i.i.d. Gaussian vectors having the covariance $\mathbb{R}_{\mathbf{u},n}$, where the covariance is constant over blocks of length N_b. Assume that $M = N/N_b$ is integer corresponding to the number of blocks. The likelihood function of the observed signals $\mathbf{y}(n)$, $n = 1, \ldots, N$, thus reads

$$p(\mathbf{y}|\mathbf{a}) = \prod_{k=1}^{M} (2\pi \det \mathbb{R}_{\mathbf{y},k})^{-N_b/2} \times \exp\left\{ -\frac{1}{2} \sum_{n=(k-1)N_b+1}^{kN_b} \mathbf{y}^T(n)\mathbb{R}_{\mathbf{y},k}^{-1}\mathbf{y}(n) \right\} \tag{6}$$

where $\mathbb{R}_{\mathbf{y},k}$ is the covariance of $\mathbf{y}(n)$ within the kth block. For simplicity, we will omit the block index k from the notation of covariance statistics.

2.1. Covariance structures

According to (4), $\mathbb{R}_\mathbf{y} = \mathbf{A}\mathbb{R}_\mathbf{u}\mathbf{A}^T$ where $\mathbb{R}_\mathbf{u}$ is the covariance of $\mathbf{u}(n)$ having the structure

$$\mathbb{R}_\mathbf{u} = \begin{bmatrix} \mathbb{R}_\mathbf{r} & \mathbf{0}_{L\times 1} \\ \mathbf{0}_{L\times 1}^T & r_s \end{bmatrix}, \tag{7}$$

where $\mathbb{R}_\mathbf{r}$ denotes the $L \times L$ covariance matrix of $\mathbf{r}(n)$ and r_s is the variance of $s(n)$ (both $\mathbb{R}_\mathbf{r}$ and r_s are nuisance parameters). Hence,

$$\mathbb{R}_\mathbf{y} = \begin{bmatrix} \mathbb{R}_\mathbf{r} & \mathbb{R}_\mathbf{r}\mathbf{a} \\ \mathbf{a}^T\mathbb{R}_\mathbf{r} & \mathbf{a}^T\mathbb{R}_\mathbf{r}\mathbf{a} + r_s \end{bmatrix}. \tag{8}$$

The sample-based estimator of $\mathbb{R}_\mathbf{y}$ is given by

$$\mathbf{R}_\mathbf{y} = \frac{1}{N_b}\sum_n \mathbf{y}(n)\mathbf{y}^T(n), \tag{9}$$

where the summation proceeds over the corresponding block. The structure of $\mathbf{R}_\mathbf{y}$ will be parametrized as

$$\mathbf{R}_\mathbf{y} = \begin{bmatrix} \mathbf{R}_\mathbf{r} & \mathbf{r}_{\mathbf{r}x} \\ \mathbf{r}_{\mathbf{r}x}^T & r_x \end{bmatrix}. \tag{10}$$

2.2. Least mean square estimator

The least mean square (LMS) estimator of $\mathbf{a}$, which does not use the BSS model described above, can be written as

$$\mathbf{B} = \langle \mathbf{R}_\mathbf{r} \rangle_M, \qquad \mathbf{q} = \langle \mathbf{r}_{\mathbf{r}x} \rangle_M, \tag{11}$$

$$\mathbf{a} = \mathbf{B}^{-1}\mathbf{q}, \tag{12}$$

where $\langle \cdot \rangle_M$ is the block-averaging operator [8] defined through

$$\langle f \rangle_M = \frac{1}{M}\sum_{k=1}^{M} f_k.$$

2.3. Gunther's estimator

In [5], Gunther derived an estimator of $\mathbf{a}$ based on the maximum likelihood estimation. It starts from an initial value of $\mathbf{a}$. An iteration proceeds as

$$\alpha_k = \mathbf{a}^T\mathbf{R}_\mathbf{r}\mathbf{a} - 2\mathbf{a}^T\mathbf{r}_{\mathbf{r}x} + r_x \tag{13}$$

$$\mathbf{B} = \left\langle \alpha_k^{-1}\mathbf{R}_\mathbf{r} \right\rangle_M, \qquad \mathbf{q} = \left\langle \alpha_k^{-1}\mathbf{r}_{\mathbf{r}x} \right\rangle_M, \tag{14}$$

$$\mathbf{a} = \mathbf{B}^{-1}\mathbf{q} \tag{15}$$

until the convergence, which usually requires only 2 to 5 iterations. By comparing (11) with (14), the estimator could be seen as a weighted LMS estimator.

3. PROPOSED ESTIMATORS

In [9] (page 1045, Appendix A), an approximate minus log-likelihood function

$$-\log p(\mathbf{y}|\mathbf{a}) \approx \frac{N_b}{4}\left\langle \mathtt{tr}(\mathbf{R}_\mathbf{y}^{-1}\Delta\mathbf{R}_\mathbf{y}\mathbf{R}_\mathbf{y}^{-1}\Delta\mathbf{R}_\mathbf{y}) \right\rangle_M + c \tag{16}$$

was derived, where c is a constant, $\mathtt{tr}(\cdot)$ denotes the trace of the argument, and $\Delta\mathbf{R}_\mathbf{y} = \mathbb{R}_\mathbf{y} - \mathbf{R}_\mathbf{y}$. Based on this, we derive the estimator of $\mathbf{a}$ by minimizing a criterion

$$Q(\mathbf{a}) = \left\langle \mathtt{tr}(\mathbf{Q}_k^{-1}\Delta\mathbf{R}_\mathbf{y}\mathbf{Q}_k^{-1}\Delta\mathbf{R}_\mathbf{y}) \right\rangle_M, \tag{17}$$

where $\mathbf{Q}_k = \mathbf{R}_\mathbf{y} + \epsilon_k \mathbf{I}_{L+1}$ are regularized weighting matrices with small positive ϵ_k.

To simplify, we will assume that the nuisance parameters $\mathbb{R}_\mathbf{r}$ and r_s satisfy $\mathbb{R}_\mathbf{r} = \mathbf{R}_\mathbf{r}$ and $r_s = r_x - \mathbf{a}^T \mathbb{R}_\mathbf{r} \mathbf{a}$. Then, it follows that

$$\Delta \mathbf{R}_\mathbf{y} = \begin{bmatrix} \mathbf{0} & \mathbf{R}_\mathbf{r}\mathbf{a} - \mathbf{r}_{\mathbf{r}x} \\ \mathbf{a}^T\mathbf{R}_\mathbf{r} - \mathbf{r}_{\mathbf{r}x}^T & 0 \end{bmatrix} \tag{18}$$

and $Q(\mathbf{a})$ becomes a quadratic function in $\mathbf{a}$. Using the matrix block-inversion lemma,

$$\mathbf{Q}_k^{-1} = \begin{bmatrix} (\mathbf{R}_\mathbf{r} + \epsilon_k \mathbf{I}_L)^{-1} + \kappa_k^{-1}\mathbf{z}_k\mathbf{z}_k^T & -\kappa_k^{-1}\mathbf{z}_k \\ -\kappa_k^{-1}\mathbf{z}_k^T & \kappa_k^{-1} \end{bmatrix}, \tag{19}$$

and

$$\mathbf{Q}_k^{-1}\Delta\mathbf{R}_\mathbf{y} = \begin{bmatrix} -\kappa_k^{-1}\mathbf{z}_k\mathbf{b}_k^T\mathbf{R}_\mathbf{r} & (\mathbf{I}_L + \kappa_k^{-1}\mathbf{z}_k\mathbf{r}_{\mathbf{r}x}^T)\mathbf{b}_k \\ \kappa_k^{-1}\mathbf{b}_k^T\mathbf{R}_\mathbf{r} & -\kappa_k^{-1}\mathbf{r}_{\mathbf{r}x}^T\mathbf{b}_k \end{bmatrix}, \tag{20}$$

where

$$\mathbf{z}_k = (\mathbf{R}_\mathbf{r} + \epsilon_k \mathbf{I}_L)^{-1}\mathbf{r}_{\mathbf{r}x}, \tag{21}$$

$$\kappa_k = \epsilon_k + r_x - \mathbf{z}_k^T\mathbf{r}_{\mathbf{r}x}, \tag{22}$$

$$\mathbf{b}_k = \mathbf{a} - \mathbf{z}_k. \tag{23}$$

Note that $\mathbf{z}_k$ is, in fact, the LMS estimate of $\mathbf{a}$ using only the kth block of data. Inserting (20) into (17), $Q(\mathbf{a})$ simplifies to

$$Q(\mathbf{a}) = 2\left\langle \kappa_k^{-1}\mathbf{b}_k^T(\mathbf{R}_\mathbf{r} + 2\kappa_k^{-1}\mathbf{r}_{\mathbf{r}x}\mathbf{r}_{\mathbf{r}x}^T)\mathbf{b}_k \right\rangle_M. \tag{24}$$

3.1. Batch-offline estimation

Taking the gradient of (24)

$$\frac{\partial}{\partial \mathbf{a}} Q(\mathbf{a}) = 4\left\langle \kappa_k^{-1}(\mathbf{R}_\mathbf{r} + 2\kappa_k^{-1}\mathbf{r}_{\mathbf{r}x}\mathbf{r}_{\mathbf{r}x}^T)\mathbf{b}_k \right\rangle_M \tag{25}$$

and putting it equal to zero, we obtain the first iteration of our estimator

$$\mathbf{B} = \left\langle \kappa_k^{-1}(\mathbf{R}_\mathbf{r} + 2\kappa_k^{-1}\mathbf{r}_{\mathbf{r}x}\mathbf{r}_{\mathbf{r}x}^T) \right\rangle_M, \tag{26}$$

$$\mathbf{q} = \left\langle \kappa_k^{-2}(r_x + \mathbf{z}_k^T\mathbf{r}_{\mathbf{r}x})\mathbf{r}_{\mathbf{r}x} \right\rangle_M, \tag{27}$$

$$\mathbf{a} = \mathbf{B}^{-1}\mathbf{q}. \tag{28}$$

Since (16) is only a second-order Taylor-series expansion of the true log-likelihood function [9], it is worth to perform more iterations. To make (26)-(28) dependent on an initial guess for $\mathbf{a}$, we can replace $\mathbf{z}_k$ in (21) by $\mathbf{a}$ resulting from the previous iteration. This simultaneously avoids the computational burden due to (21). The estimation procedure is summarized in Algorithm 1.

Algorithm 1: Batch-offline estimator of $\mathbf{a}$

Input: $\mathbf{R}_\mathbf{r}$, $\mathbf{r}_{\mathbf{r}x}$ and r_x for $k = 1, \ldots, M$
for $i = 1, \ldots,$ *number of iterations* **do**
  for $k = 1, \ldots, M$ **do**
    if *i=1* **then** $\mathbf{z}_k = (\mathbf{R}_\mathbf{r} + \epsilon_k \mathbf{I}_L)^{-1}\mathbf{r}_{\mathbf{r}x}$;
    else $\mathbf{z}_k = \mathbf{a}$;
    $\kappa_k = \epsilon_k + r_x - \mathbf{z}_k^T\mathbf{r}_{\mathbf{r}x}$
  end
  Compute $\mathbf{B}$, $\mathbf{q}$, and $\mathbf{a}$ according to (26)-(28)
end

3.2. Adaptive estimator

We propose an adaptive algorithm adopting the idea of the standard adaptive RLS [1]. The block-averaging operator is replaced by continuous updates using a forgetting factor $\lambda \in (0, 1)$. This way the sample-covariance estimators $\mathbf{R}_\mathbf{r}$, $\mathbf{r}_{\mathbf{r}x}$ and r_x and $\mathbf{B}$ from (26) and $\mathbf{q}$ from (27) are updated in every step using new data. To lower the computational complexity, the RLS algorithm runs in parallel to compute (21). The steps are summarized in Algorithm 2.

Algorithm 2: Adaptive estimator of $\mathbf{a}$

Initialize: $\mathbf{R}_\mathbf{r} = \mathbf{B} = \mathbf{P} = \mathbf{I}_L$, $\mathbf{r}_{\mathbf{r}x} = \mathbf{q} = \mathbf{z} = \mathbf{0}_{L\times 1}$, $r_x = 0$
for $n = 1, 2, \ldots$ **do**
  $\mathbf{R}_\mathbf{r} \leftarrow \lambda\mathbf{R}_\mathbf{r} + \mathbf{r}(n)\mathbf{r}^T(n)$
  $\mathbf{r}_{\mathbf{r}x} \leftarrow \lambda\mathbf{r}_{\mathbf{r}x} + \mathbf{r}(n)x(n)$
  $r_x \leftarrow \lambda r_x + x(n)^2$
RLS $\mathbf{h} = \mathbf{P}\cdot\mathbf{r}(n)$
RLS $\mathbf{k} = (\lambda + \mathbf{h}^T\mathbf{r}(n))^{-1}\mathbf{h}$
RLS $\xi(n) = x(n) - \mathbf{z}^T\mathbf{r}(n)$ /* the output of RLS */
RLS $\mathbf{z} \leftarrow \mathbf{z} + \xi(n)\mathbf{k}$ /* (21) with $\epsilon = 0$ by RLS */
RLS $\mathbf{P} \leftarrow (\mathbf{P} - \mathbf{k}\,\mathbf{h}^T)/\lambda$ /* the inverse of $\mathbf{R}_\mathbf{r}$ */
  $\kappa = \epsilon + r_x - \mathbf{z}^T\mathbf{r}_{\mathbf{r}x}$ /* according to (22) */
  $\mathbf{B} \leftarrow \lambda\mathbf{B} + \kappa^{-1}(\mathbf{R}_\mathbf{r} + 2\kappa^{-1}(\mathbf{r}_{\mathbf{r}x}\mathbf{r}_{\mathbf{r}x}^T))$ /* by (26) */
  $\mathbf{q} \leftarrow \lambda\mathbf{q} + \kappa^{-2}(r_x + \mathbf{z}^T\mathbf{r}_{\mathbf{r}x})\mathbf{r}_{\mathbf{r}x}$ /* by (27) */
  $\mathbf{a} = \mathbf{B}^{-1}\mathbf{q}$ /* by (28) */
  $\widehat{s}(n) = x(n) - \mathbf{a}^T\mathbf{r}(n)$ /* the output */
end

4. SIMULATION RESULTS

4.1. Echo path estimation

We simulate an AEC situation where $s(n)$ and $r(n)$ are, respectively, continuous man's and woman's utterances taken from TIMIT [11]. Both signals are 8.3 s in length, and the doubletalk is present during the whole recording; the sample rate is 16 kHz. Room Impulse Response (RIR) generator [12] is used to generate $a(n)$, where the room size is $(5\times 7\times 3)$ m, the loudspeaker is placed at [1.5, 0.15, 1] m and the microphone at [0.5, 0.1, 1] m. The length of $a(n)$ is controlled though the reverberation time parameter T_{60}.

In a batch-offline regime, we compute $\widehat{a}(n)$ of length L using the LMS approach (11), the Gunther's estimator (13)-(15), and the proposed estimator (26)-(28). The latter two methods use $M = 10$ and perform 5 iterations.

In experiments, one of three parameters was varied, namely, the input signal-to-echo ratio in $x(n)$ (input SER),

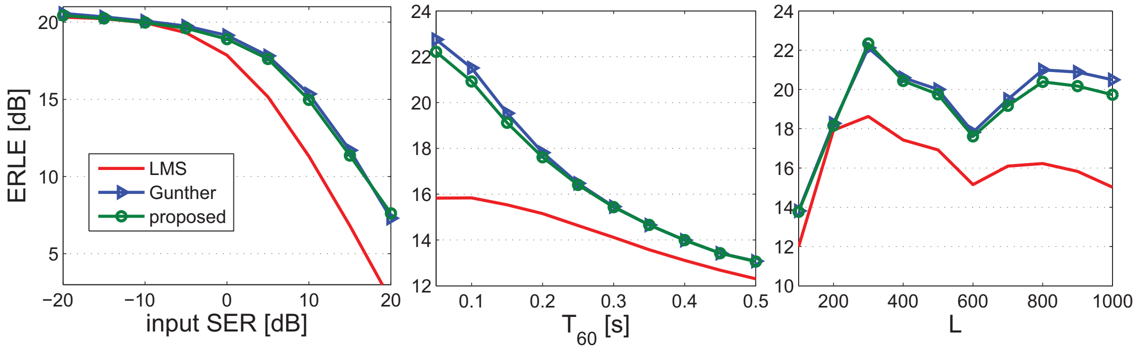

Fig. 1. Results of batch-offline echo cancellation when the input SER in $x(n)$ or T_{60} or L was varied while the other parameters were fixed to 5 dB of the input SER, $T_{60} = 200$ ms and $L = 600$.

T_{60}, and L, while the other ones were fixed. The fixed values are 5 dB of the input SER, $T_{60} = 200$ ms and $L = 600$. Echo Return Loss Enhancement (ERLE) was used for the evaluation, which is the ratio of energies of the echo in $x(n)$ and of the residual echo in $\widehat{s}(n)$ [5].

The results in Fig. 1 show that the proposed estimator yields comparable performance to the Gunther's method, which is in agreement with the fact that both methods consider the same statistical model. LMS is outperformed by the semi-blind methods, especially, when input SER is above -3 dB or T_{60} is below 300 ms or L is sufficiently large compared to the true length of $a(n)$.

4.2. Adaptive Interference Cancellation

A similar situation to AEC occurs when observing a noisy recording of the target signal using a microphone array; we consider two microphones here. If the noise is due to a point source (interferer), the reference signal $r(n)$ could be obtained using a multichannel filter that performs spatial null towards the target source, so-called target-cancellation filter [2, 13, 14]. The situation could be seen as (1) where $x(n)$ is the signal on a selected microphone and $r(n)$ is a filtered version of the interference in $x(n)$. The difference from AEC is that $r(n)$ typically contains a remaining part of the target signal and $a(n)$ need not be causal.

We simulate such scenario using the RIR generator with $T_{60} = 50$ ms. Two microphones are placed 10 cm distant from each other in positions [0.5 0.1 1] m and [0.5 0.2 1] m. The sources, the target (male) and the interferer (female), are about 1 m distant from microphones in positions [1.5 0.15 1] m and [1.5 1.15 1] m, respectively. The reference signal $r(n)$ is obtained using a target-cancellation filter of length 1000 taps, which was derived under noise-free conditions using least-squares [14]. During the testing recording, the filter suppresses the target by about 14 to 37 dB.

The adaptive method proposed in Section 3.2 was applied to cancel the interferer's voice from the signal on microphone 1. To compare, we also apply the standard adaptive

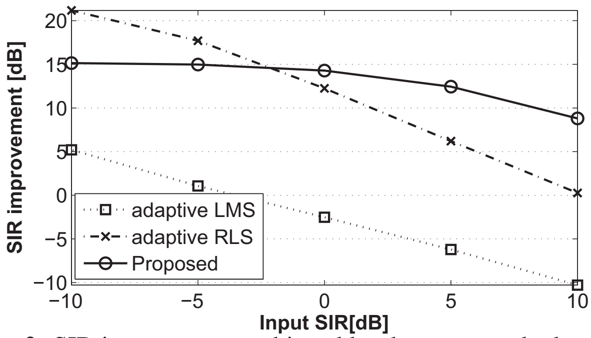

Fig. 2. SIR improvement achieved by the compared adaptive algorithms; $T_{60} = 50$ ms, $L = 100$.

LMS and RLS algorithms [1]. Parameters were adjusted to roughly optimize ERLE: the step-size in adaptive LMS is $\nu = 0.7$ and the forgetting factor in RLS is $\lambda = 0.9999$. In the proposed method, $\lambda = 0.9999$ and $\epsilon = 10^{-4}$. All methods estimate filters of length $L = 100$ taps. The performance is evaluated using the standard Signal-to-Interference ratio (SIR) measured on inputs and outputs of the methods and averaged over the whole processed data.

Results in Fig. 2 show that adaptive LMS fails to converge. RLS works better than LMS thanks to the slow forgetting ($\lambda = 0.9999$). It yields the best performance when the interferer is stronger than the target source, that is, when input SIR≤ -3 dB. Our adaptive algorithm outperforms RLS when input SIR is higher than -3 dB.

5. CONCLUSIONS

The semi-blind methods outperform standard least mean-squares based algorithms, especially, when the target signal is strong compared to the echo/interference. Essentially, a doubletalk detector is not needed. It can be used for further performance improvement by switching between RLS and the proposed method when doubletalk is *not* present. Future research can be also focused on frequency-domain implementations that are computationally cheaper [10, 13].

6. REFERENCES

[1] S. Haykin, *Adaptive Filter Theory*, Fourth Edition, Prentice-Hall, 2002.

[2] S. Gannot, D. Burshtein, and E. Weinstein, "Signal enhancement using beamforming and nonstationarity with applications to speech," *IEEE Trans. on Signal Processing*, vol. 49, no. 8, pp. 1614–1626, Aug. 2001.

[3] K.-H. Lee, J.-H. Chang, N. S. Kim, S. Kang, and Y. Kim, "Frequency domain double-talk detection based on the Gaussian mixture model," *IEEE Signal Process. Lett.*, vol. 17, no. 5, pp. 453-456, May 2010.

[4] H. Buchner, J. Benesty, T. Gansler, and W. Kellermann, "Robust extended multidelay filter and double-talk detector for acoustic echo cancellation," *IEEE Trans. Audio, Speech Lang. Process.*, vol. 14, no. 5, pp. 1633-1644, Sep. 2006.

[5] J. Gunther, "Learning Echo Paths During Continuous Double-Talk Using Semi-Blind Source Separation," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 20, no. 2, Feb. 2012.

[6] D-T. Pham and J-F. Cardoso. "Blind separation of instantaneous mixtures of non stationary sources", *IEEE Trans. Signal Processing*, pp. 1837–1848, vol. 49, no. 9, 2001.

[7] J-F. Cardoso, "Multidimensional independent component analysis," *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 4, pp. 1941–1944, 12-15 May, 1998.

[8] D. Lahat, J. Cardoso, and H. Messer, "Second-Order Multidimensional ICA: Performance Analysis," *IEEE Transactions on Signal Processing*, vol.60, no.9, pp. 4598–4610, Sept. 2012.

[9] P. Tichavský and Z. Koldovský, "Weight adjusted tensor method for blind separation of underdetermined mixtures of nonstationary sources," *IEEE Transactions on Sig. Proc.*, vol. 59, no. 3, pp. 1037–1047, March 2011.

[10] F. Nesta, T.S. Wada, and B.-H. Juang, "Use of decorrelation procedure applied to multi-channel acoustic echo cancellation," *IEEE Audio, Speech, Lang. Process.*, vol. 19, no. 3, pp. 583–599, Mar. 2011.

[11] J. S. Garofolo, et al., "TIMIT Acoustic-Phonetic Continuous Speech Corpus," Linguistic Data Consortium, Philadelphia, 1993.

[12] E. Lehmann and A. Johansson, "Prediction of energy decay in room impulse responses simulated with an image-source model," *Journal of the Acoustical Society of America*, vol. 124, no. 1, pp. 269–277, July 2008.

[13] Y. Takahashi, T. Takatani, K. Osako, H. Saruwatari, K. Shikano, "Blind Spatial Subtraction Array for Speech Enhancement in Noisy Environment," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 4, pp. 650–664, May 2009.

[14] Z. Koldovský, J. Málek, P. Tichavský, and F. Nesta, "Semi-blind Noise Extraction Using Partially Known Position of the Target Source," *IEEE Trans. on Speech, Audio and Language Processing*, vol. 21, no. 10, pp. 2029–2041, Oct. 2013.