NOISE CANCELLATION USING TWO CLOSELY SPACED MICROPHONES: EXPERIMENTAL STUDY WITH A SPECIFIC MODEL AND TWO ADAPTIVE ALGORITHMS

Mohamed DJENDI [1], André GILLOIRE [2] and Pascal SCALART [1]
[1] University of Rennes - IRISA / ENSSAT, 6 Rue de Kerampont, B.P. 80518, 22305 Lannion Cedex, France,
Email: {mohamed.djendi, pascal.scalart}@enssat.fr

[2] France Télécom- TECH/SSTP, 2 Avenue Pierre Marzin, 22307 Lannion Cedex, France
Email: andre.gilloire@francetelecom.com

ABSTRACT

We consider the speech enhancement problem in a moving car through a blind source separation scheme involving two closely spaced microphones. We propose the use of a double fast Newton transversal filter algorithm (DFNTF) to estimate and suppress coherent noise components from speech, and a model of signal mixtures able to represent correctly the effect of microphones spacing. We also consider the realistic case where the noises at the sensor inputs contain non-coherent components. The simulation results show that the DFNTF algorithm, when controlled by a Voice Activity Detector (VAD), is able to fully cancel the correlated noise components from speech.

1. INTRODUCTION

Blind source separation (BSS) methods aim at estimating N_s source signals $u_i(n)$ from N_0 observed signals $p_i(n)$, which are mixtures of these source signals. In the so-called convolutive mixture model, each propagation path from source *j* to sensor (observation) *i* is represented by a linear filter, whose impulse response is denoted $\boldsymbol{h}_{ij}$ hereafter. Many investigations have been performed in the case when $N_0 = N_s$ so that the mixing matrix is square and assumed invertible [1],[2]. In this paper, we consider an extension of this idealized case: we observe at the outputs of two microphones convolutive mixtures of signals issued from two spatially localized (point) sources, plus additive noises assumed uncorrelated between the two microphones. One of the two point sources is speech (the useful signal), and the second one can represent either the car engine noise or far-end speech that we want to cancel. The additive noise components represent the non-coherent part of the diffuse acoustic (background) noise in the vicinity of the microphones. Adaptive noise cancellers based on the source separation principle, as described in [1], use two adaptive filters arranged in a feed-forward or a feedback symmetric structure. In this paper, we propose a feed-forward implementation based on what we call the double fast Newton transversal filter (DFNTF) algorithm. We also present a specific experimental model to evaluate the performance of the proposed algorithm; this model is able to correctly represent the effect of close sensors on the performance of the noise canceller. Experimental results obtained from simulations and a comparison between the double NLMS (DNLMS) and the proposed DFNTF algorithms are also given and discussed in this paper.

2. MODEL ADAPTED TO THE PROBLEM

The considered basic configuration involves two convolutive mixtures of two uncorrelated sources, defined as:

$$p_1(n) = \boldsymbol{h}_{11} * \boldsymbol{s}(n) + \boldsymbol{h}_{21} * \boldsymbol{b}(n) + n_1(n)$$
$$p_2(n) = \boldsymbol{h}_{22} * \boldsymbol{b}(n) + \boldsymbol{h}_{12} * \boldsymbol{s}(n) + n_2(n)$$

where $\boldsymbol{h}_{11}$ and $\boldsymbol{h}_{22}$ represent the impulse responses of each channel separately, and $\boldsymbol{h}_{12}$ and $\boldsymbol{h}_{21}$ represent the cross-coupling effects between the channels; n_1 and n_2 represent the additive background noises. * represents convolution. In Fig.1, $\boldsymbol{h}_{11}$ and $\boldsymbol{h}_{22}$ are assumed to be identity; this assumption does not impact the practical usefulness of the model: we may assume as well that the near-end speaker is not very far from the microphones and that we have no *a priori* information on the noisy point source.

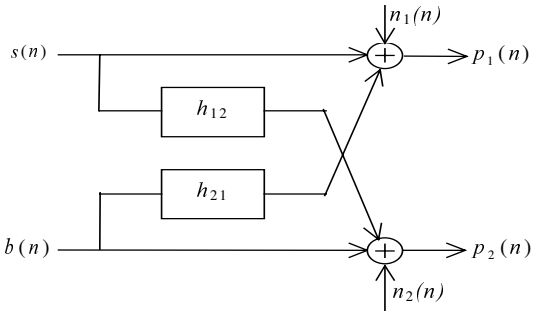

Fig. 1. Assumed mixing model with background noise

3. ADAPTIVE FILTERING ALGORITHMS USED

It has been proved in [1] that minimizing the correlation of the output is the same as in the least square (LS) case; hence we propose to use the DFNTF and the DNLMS algorithms.

3.1. The NLMS Algorithm

The NLMS algorithm updates the coefficients of the impulse response $\boldsymbol{h}_L = [\boldsymbol{h}_0 \ldots \boldsymbol{h}_{L-1}]^T$ of a FIR filter so as to minimize the mean-square error (MSE) between the filter output and a desired-response signal $d(n)$. The updating rule is as follows [3]: $e(n) = d(n) - \boldsymbol{h}_L^T(n)\,\boldsymbol{x}_L(n)$

$$\boldsymbol{h}_L(n) = \boldsymbol{h}_L(n-1) + \frac{\mu\, e(n)\, \boldsymbol{x}_L(n)}{\boldsymbol{x}_L{}^T(n)\,\boldsymbol{x}_L(n)}$$

where $\boldsymbol{x}_L(n) = [x(n), x(n-1) \ldots x(n-L+1)]^T$ is the input signal vector and μ is the adaptation gain or step size, which must be chosen between 0 and 2 to achieve convergence.

3.2. The FNTF Algorithm

This algorithm [4],[5] is an efficient implementation of the exponentially weighted recursive LS (RLS) algorithm; it is based on the minimisation of the cost function:

$$J(n) = \sum_{i=0}^{t} \lambda^{n-i} \left[y(n) - \boldsymbol{h}_L^T(n)\,\boldsymbol{x}_L(n) \right]^2 \qquad (1)$$

It is well established that time-recursive minimization of (1) leads to the following RLS set of equations:

$\boldsymbol{h}_L(n) = \boldsymbol{h}_L(n-1) - c_L(n)\, e(n)$, $e(n) = y(n) - \hat{y}(n)$

$\hat{y}(n) = \boldsymbol{h}_L^T(n-1)\,\boldsymbol{x}_L(n)$, $c_L(n) = \boldsymbol{R}_L{}^{-1}(n)\,\boldsymbol{x}_L(n)$

where $\boldsymbol{R}_L(n)$ is the $L \times L$ covariance matrix of the input signal. In the above RLS equations, the update of the gain vector $c_L(n)$ requires the update of the inverse covariance matrix. In the FNTF algorithm, it is obtained by the use of two prediction parts of order $N << L$; moreover, instead of updating $c_L(n)$, the FNTF version that we use updates the so-called dual Kalman gain $\tilde{c}_L(n) = [\gamma_L(n) c_L(n)]^{-1}$ where $\gamma_L(n)$ is the likelihood variable defined as $\gamma_L(n) = 1 + c_L^T(n)\,\boldsymbol{x}_L(n)$. For stabilization of the FNTF, we have used a technique recalled in [4],[5].

4. DOUBLE FNTF ALGORITHM (DFNTF)

Fig.2 shows the forward implementation of the noise canceller. In this figure, the evident theoretical solution of the problem when $n_1 = n_2 = 0$ is given by setting $\boldsymbol{w}_{21} = \boldsymbol{h}_{21}$ and $\boldsymbol{w}_{12} = \boldsymbol{h}_{12}$ [1]. The proposed double fast Newton transversal filter (DFNTF) algorithm minimizes the following *a priori* errors: $\boldsymbol{u}_1(n) = p_1(n) - p_2(n) * \boldsymbol{w}_{21}(n)$ and $\boldsymbol{u}_2(n) = p_2(n) - p_1(n) * \boldsymbol{w}_{12}(n)$. For each input p_1 and p_2, we calculate the dual Kalman gain by the propagation of two prediction parts and an extrapolation. We have used the algorithm in [4] to compute these components. Introducing the two Kalman gain vectors $\tilde{c}_L^1(n)$ and $\tilde{c}_L^2(n)$ with length L associated with the filters $\boldsymbol{w}_{12}$ and $\boldsymbol{w}_{21}$, we get the filter updates:

$\boldsymbol{w}_{12}(n) = \boldsymbol{w}_{12}(n-1) - \tilde{c}_L^1(n)\, u_2(n)$

$\boldsymbol{w}_{21}(n) = \boldsymbol{w}_{21}(n-1) - \tilde{c}_L^2(n)\, u_1(n)$

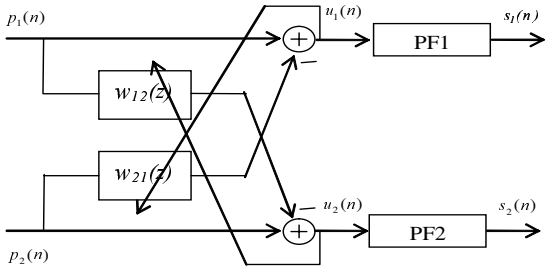

Fig. 2. Forward implementation of the ANC

In the structure shown Fig.2, PF1 and PF2 are two post filters, which are not used in the simulations. The two outputs u_1 and u_2 inside this structure, in the presence of the noise components n_1 and n_2 are given as:

$\boldsymbol{u}_1(n) = \boldsymbol{s}(n) * [\delta(n) - \boldsymbol{h}_{12} * \boldsymbol{w}_{21}(n)] + \boldsymbol{n}_1(n) - \boldsymbol{n}_2(n) * \boldsymbol{w}_{21}(n)$

$\boldsymbol{u}_2(n) = \boldsymbol{b}(n) * [\delta(n) - \boldsymbol{w}_{12}(n) * \boldsymbol{h}_{21}] + \boldsymbol{n}_2(n) - \boldsymbol{n}_1(n) * \boldsymbol{w}_{12}(n)$

This structure allows elimination of the noise component $\boldsymbol{b}(n) * \boldsymbol{h}_{21}$ from the output $\boldsymbol{u}_1(n)$ and of a part of speech signal component $\boldsymbol{s}(n) * \boldsymbol{h}_{12}$ from output $\boldsymbol{u}_2(n)$. The LS solution in the frequency domain in the presence of the decorrelated diffuse noises n_1 and n_2 is given as:

$$w_{12}(\omega) = F_1^{-1}(\omega)\left[S_{bb}(\omega) h_{21}(\omega) + S_{ss}(\omega) h_{12}(\omega)\right] \qquad (2)$$

$$F_1(\omega) = h_{21}^2(\omega) S_{bb}(\omega) + S_{ss}(\omega) + S_{n_1 n_1}(\omega) \qquad (3)$$

$$w_{21}(\omega) = F_2^{-1}(\omega)\left[S_{ss}(\omega) h_{12}(\omega) + S_{bb}(\omega) h_{21}(\omega)\right] \qquad (4)$$

$$F_2(\omega) = h_{12}^2(\omega) S_{ss}(\omega) + S_{bb}(\omega) + S_{n_2 n_2}(\omega) \qquad (5)$$

where S_{ss}, S_{bb}, $S_{n_1 n_1}$, $S_{n_2 n_2}$ are respectively, the power spectral densities of the speech, noise and diffuse noises.

5. VAD-CONTROLLED ADAPTIVE SCHEME

It is well known that the signals at the outputs of the symmetric structure shown in Fig. 2 are obtained within a permutation. Nevertheless, one can get the useful signal at the appropriate output by taking advantage of the non-stationarity of speech, which is basically an intermittent signal [2],[6]. We have used a voice activity detector (VAD) to control the adaptation of the filters: i.e., the filter $\boldsymbol{w}_{21}$ is adapted during noise-only periods, whereas the filter $\boldsymbol{w}_{12}$ is adapted only during voice activity periods. As shown experimentally in the sequel, this VAD-controlled adaptive scheme yields de-noised speech at the output s_1, and achieves good convergence of the adaptive algorithms. We note that in noise-only periods, the structure controlled by the VAD behaves as an ANC, as described in [3].

6. EXPERIMENTAL STUDY

6.1. Experimental Model and Simulation Conditions

We propose a specific implementation of the mixing model shown Fig.1, consistent with the physics of the problem and able to represent appropriately the effect of the distance between the two microphones on the characteristics of the signals, while complying with the assumed unit transfer functions from each point-source signal to each direct path sensor. In this implementation, we assume that the cross-coupling impulse responses $\boldsymbol{h}_{12}$ and $\boldsymbol{h}_{21}$ are made of two parts:

1) A unit pulse δ(n) localized at the beginning of the impulse response, which represents the direct acoustic path from each source to the cross-coupled microphone.

2) An exponentially weighted tail $\boldsymbol{h}'$ representing the room effect. This tail was obtained by weighting simulated random noise sequences according to the weighting function: $f(n) = A.e^{-Bn}$, where A is a scale factor (taken equal to 1) and B is a damping factor, which models the absorption of the sound waves on the walls of the car and is therefore linked to the reverberation time n_r: $B = 3\ln(10)/n_r = 6.9078/n_r$

The simulated impulse responses have been constructed as follows: $\boldsymbol{h}_{12} = \delta(n) + \boldsymbol{h}'_{12}$ and $\boldsymbol{h}_{21} = \delta(n) + \boldsymbol{h}'_{21}$. The amplitudes of the tails $\boldsymbol{h}'_{12}$ and $\boldsymbol{h}'_{21}$ have been adjusted according to the variance of the random noise. If the two microphones are made coincident, the cross-coupled filters become identical to the direct channels $\boldsymbol{h}_{11}$ and $\boldsymbol{h}_{22}$. Fig.3 shows examples of such impulse responses: the random noise variance is equal to 0.5, which corresponds to relatively spaced microphones; the damping factor is $B = 0.028$, with a sampling period $T_s = 125\mu s$, the corresponding reverberation time is 30.8 ms. The size of the impulse responses is $L = 100$.

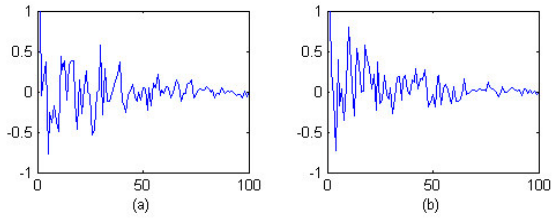

Fig. 3. Examples of simulated impulse responses *(a): h_{12}, (b): h_{21}*

All the simulations have been performed at the sampling frequency f_s = 8 kHz. The speech signal is a sentence of about 4s and the point-source noise signal is stationary with average speech spectrum (USASI noise). Note that in all the simulations carried out with this structure, we did not use the post-filters. The simulation parameters of the DFNTF algorithm are: $L = 100$, $N = 10$, $\lambda = 0.9967$. For fair comparison, the step size of the DNLMS has been adjusted so as to obtain the same asymptotic error as the DFNTF [5], i.e. $\lambda = 1 - 1/pL$ and $\mu = 1/p$ where $p = 3$. The SNR (speech-to-noise ratio) at the inputs of the forward structure is 3dB.

6.2. Simulations with Spaced Microphones and No Background Noise

The impulse responses have been constructed with a random noise variance of 0.5. The obtained signals are shown in Fig.4.

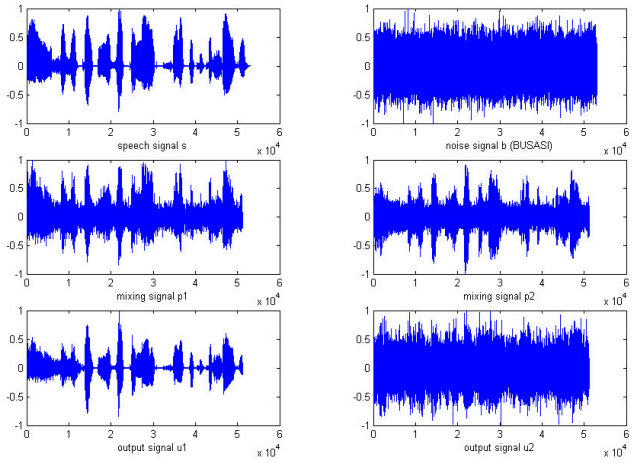

Fig. 4. Source signals (top), mixtures (middle) and noise canceller outputs (bottom) obtained with the DFNTF algorithm

One can see from inspection of Fig.4 that the DFNTF algorithm performs well with the feed-forward structure and that the speech output is completely de-noised. Note that this result was obtained thanks to the use of the VAD, as explained before. Fig.5 shows the system mismatch $\|\boldsymbol{h}_{21} - \boldsymbol{w}_{21}\| / \|\boldsymbol{h}_{21}\|$ for the DFNTF and DNLMS algorithms. The highly superior convergence speed of the DFNTF appears clearly in the figure: this result comes from the spectral shaping of the noise; one can expect even higher differences in real cases.

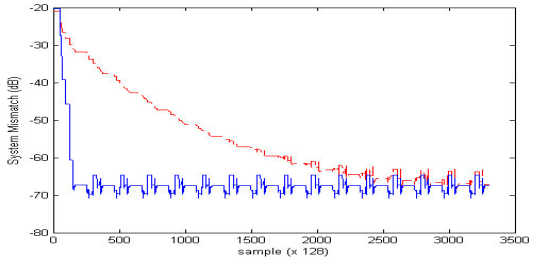

Fig. 5. Comparison of system mismatches obtained by DNLMS (dotted) and DFNTF (solid) algorithms. The periodic pattern comes from recycling the same speech and noise files to achieve sufficient simulation length.

6.3. Simulations with Close Microphones and No Background noise

The impulse responses have been constructed with random noise variance of 0.025. The signals obtained with the DFNTF algorithm are shown in Fig.6.

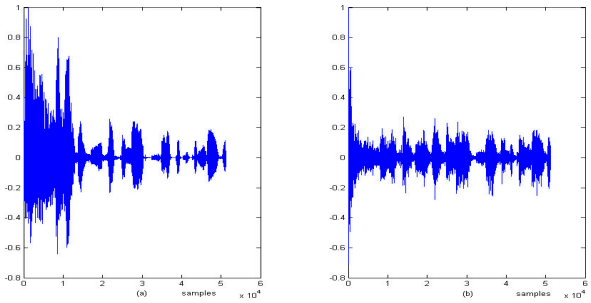

Fig. 6. Outputs obtained with the DFNTF algorithm in the case of close microphones. (a): output u_1 ; (b): output u_2 .

In this experiment, the two adaptive filters $\boldsymbol{w}_{12}$ and $\boldsymbol{w}_{21}$ are close to $\delta(n)$, hence close to $\boldsymbol{h}_{11}$ and $\boldsymbol{h}_{22}$ respectively. One can observe that the output signals are attenuated because the post filter was not used; its theoretical form is $1/[1 - \boldsymbol{h}_{12}(n) * \boldsymbol{w}_{21}(n)]$[1], which yields ideal reconstruction of the source signals. Therefore, one can deduce that correct source signal amplitudes can be recovered through the use of the post-filters, which take high gain values in the case of close sensors. In the simulations reported in Sect. 5.2 with farther spaced microphones, the outputs are not attenuated – although somewhat modified. This is due to the significant values of the product $\boldsymbol{w}_{12}$*$\boldsymbol{w}_{21}$. We conclude that the proposed specific model highlights the physical phenomenon, which represents the behaviour of the forward structure in real situations. Similar results were obtained with the DNLMS algorithm. We also note that behaviour of the system mismatch similar to Fig.5 was obtained in these simulations with the DFNTF and DNLMS algorithms.

6.4. Simulations with Spaced Microphones and Background Noise

Recall that the appropriate LS solution (without post-filters) is given by (2) and (4). The power ratio between the point source noises and the diffuse noises on each microphone was equal to 5 dB. The signals obtained with the DFNTF are shown in Fig.7.

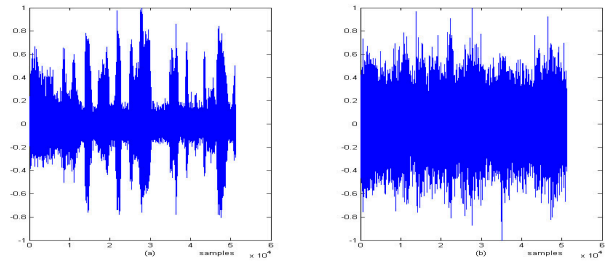

Fig. 7. Outputs obtained with the DFNTF algorithm in the presence of the diffuse background noises.

Fig.7 shows that some noise is present in the two outputs, which is a combination of the two diffuse noises (and of components coming from the point-source noise). The SNR enhancement obtained is 6.5 dB. The same experiment has also been carried out with the DNLMS algorithm and we obtained a similar result. We further observed that the system mismatch in this experiment goes up to about -22 dB, which illustrates that the diffuse noises disturb the filter identification.

7. CONCLUDING REMARKS

We have presented a noise cancelling system based on a blind source separation structure and on the double fast Newton transversal filter (DFNTF) algorithm, and we have shown the superiority of the DFNTF algorithm compared to the double NLMS algorithm. We have shown that the control of the adaptive filters by a VAD allows full cancellation of the coherent noise components from speech. We have also proposed a new model to study the separation problem with two closely-spaced microphones, which is physically consistent, as it appears from simulations with different sensor spacings. Considering the realistic case where the noises at the sensors inputs contain non-coherent components, we note that bringing the sensors closer should decrease those components; nevertheless, the post-filters derived from the blind source separation scheme may lead to large noise amplification. Since the spectrum of the non-coherent noise is localised at high frequencies, one can predict that its impact on the overall performance of the system should be limited. An alternative solution to this problem was proposed in [6] and [7] by adding a noise reduction filter at the speech output of the BSS stage.

8. REFERENCES

[1] S.Van Gerven and D. Van Compernolle," Signal separation by symmetric adaptive decorrelation: stability, convergence, and uniqueness," *IEEE Trans. Signal Proc.,* vol. 43, no. 7, pp. 1602-1612, July 1995.

[2] E. Weinster, M. Feder and A. Oppenheim," Multi-channel signal separation by decorrelation," *IEEE Trans. Speech and Audio Proc.*, vol. 1, no. 4, pp. 405-413, Oct. 1993.

[3] B. Widrow, J.R. Glover, J.M. McCool, J. Kaunitz, C.S. Williams, R.H. Hearn, J.R. Zeidler, E. Dong and R.C. Goodlin, "Adaptive noise cancelling: principles and applications", *Proc. IEEE*, vol. 63, pp. 1692-1716, Dec. 1975.

[4] M. Djendi, M. Rahim, A. Guessoum, M. Bouchard, and D. Berkani, "Comparative study of new version of the Newton type adaptive filtering algorithm," In *Proc. IEEE ICASSP,* 2004, pp. 677-680.

[5] T. Petillon, A. Gilloire and S. Theodoridis, "The fast Newton transversal filters: An efficient scheme for acoustic echo cancellation in mobile radio," *IEEE Trans. Signal Processing*, vol.42, n°.3, pp. 509-518. March 1994.

[6] E. Visser, and T.W. Lee, "Application of blind source separation in speech processing for combined interference removal and robust speaker detection using a two-microphone setup," In *Proc. 4th Int. Symp. On Independent Component and Blind Signal Separation,* ICAS2003, 2003, pp. 325-329.

[7] L. Parra and C. Spence, "Convolutive blind separation of non stationary sources," *IEEE Trans. Speech and Audio Proc.*, vol.8, no. 3, pp.320-327, May 2000.