

ON THE NON-UNIQUENESS PROBLEM AND THE SEMI-BLIND SOURCE SEPARATION

Francesco Nesta[†], Ted S. Wada[‡], Shigeki Miyabe[§], Biing-Hwang (Fred) Juang[‡]

[†] Fondazione Bruno Kessler-Irst (Trento, Italy), Università di Trento (Italy)

[‡] Center for Signal and Image Processing, Georgia Institute of Technology, Atlanta, GA, USA

[§] Department of Information Physics and Computing, Graduate School of Information Science and Technology, University of Tokyo, Japan
 e-mail: [†]nesta@fbk.eu, [‡]{twada,juang}@ece.gatech.edu, [§]miyabe@hil.t.u-tokyo.ac.jp

ABSTRACT

Semi-blind source separation (SBSS) is a special case of the well-known source separation problem when some partial knowledge of the source signals is available to the system. In particular, a batch-wise adaptation in the frequency domain based on the independent component analysis (ICA) can be effectively used to jointly perform source separation and multi-channel acoustic echo cancellation (MCAEC) without double-talk detection. However, the non-uniqueness problem due to the correlated far-end reference signals still affects the SBSS approach. In this paper, we analyze the structure of the SBSS de-mixing matrix and the behavior of a batch on-line adaptation algorithm under two most common far-end mixing conditions. We show that with a proper constraint on the de-mixing matrix, high echo reduction can be achieved just as the misalignment remains relatively low even for the worst-case scenario of single far-end talker and also without any pre-processing procedure to decorrelate the far-end signals.

Index Terms— Blind source separation, multi-channel acoustic echo cancellation, semi-blind source separation

1. INTRODUCTION

It was shown in [1] that blind source separation (BSS) and stereophonic acoustic echo cancellation (SAEC) can be effectively combined together by applying independent component analysis (ICA) in the frequency domain. Such an approach is referred to as the semi-BSS (SBSS) since the reference signals, i.e., mixture of far-end source signals, are known *a priori* and can be used directly for the adaptation of the separation filter [2]. Although the double-talk detection is no longer necessary due to the effectiveness of the ICA and the batch-wise (i.e., off-line) adaptation, the so-called non-uniqueness problem still exists when the modeling filter is equal to or longer in length than the far-end room impulse response [3]. Such a condition is rare since in reality the impulse response length is infinite. However, the ill-conditioning of the mixing system does occur frequently, e.g., when only one far-end source is active, resulting in highly correlated reference signals from the far end. Thus some pre-processing procedure before playback at the near end to decorrelate the signals becomes necessary at the cost of degraded signal quality perceived by the near-end listeners.

In this paper, we analyze the structure of the SBSS de-mixing matrix to see how the multi-channel acoustic echo cancellation (MCAEC) performance can be improved. We also study the behavior of a combination of batch-wise and on-line adaptations to possibly take advantage of both types of learning. We will show through two different far-end mixing conditions that with a proper constraint on the de-mixing matrix, both high echo reduction and relatively low misalignment can be achieved even for the worst-

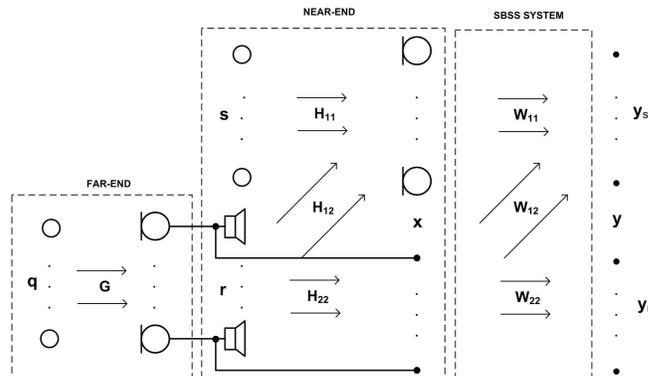


Figure 1: Model of the near-end and the far-end mixing systems and the SBSS system.

case scenario and without any pre-decorrelation procedure performed on the far-end signals.

2. SBSS MODEL

We consider a time-invariant mixing model in the frequency domain where the number of microphones are assumed to be greater than or equal to the number of sources. As illustrated in Figure 1, a set of sources, represented by a vector \mathbf{q} , is recorded by microphones at the far end, where the corresponding mixing system is represented by a frequency response matrix \mathbf{G} . A set of near-end sources \mathbf{s} is multiplied by the frequency response \mathbf{H}_{11} , and a set of reference signals \mathbf{r} , after being played through the near-end loudspeakers, is multiplied by the frequency response \mathbf{H}_{12} and recorded by the near-end microphones. Then a set of observations \mathbf{x} used at the input of the SBSS system is

$$\mathbf{x}(\omega) = \mathbf{H}(\omega) \begin{pmatrix} \mathbf{s}(\omega) \\ \mathbf{r}(\omega) \end{pmatrix}, \quad (1)$$

$$\mathbf{H}(\omega) = \begin{bmatrix} \mathbf{H}_{11}(\omega) & \mathbf{H}_{12}(\omega) \\ \mathbf{O} & \mathbf{I} \end{bmatrix}, \quad (2)$$

where \mathbf{H} is the response matrix of the entire near-end mixing system, \mathbf{H}_{22} is naturally assumed to be the identity matrix \mathbf{I} , and \mathbf{O} is a matrix with all elements equal to 0. The purpose of the SBSS system is to perform the estimation of the near-end sources by using a de-mixing matrix \mathbf{W} :

$$\mathbf{y}(\omega) = \mathbf{W}(\omega)\mathbf{x}(\omega) = \begin{pmatrix} \mathbf{y}_s(\omega) \\ \mathbf{y}_r(\omega) \end{pmatrix} \simeq \begin{pmatrix} \mathbf{s}(\omega) \\ \mathbf{r}(\omega) \end{pmatrix}, \quad (3)$$

where we generalize the structure of \mathbf{W} as

$$\mathbf{W}(\omega) = \begin{bmatrix} \mathbf{W}_{11}(\omega) & \mathbf{W}_{12}(\omega) \\ \mathbf{O} & \mathbf{W}_{22}(\omega) \end{bmatrix}. \quad (4)$$

We note that in the SBSS, we are not interested in recovering the signals played through the loudspeakers since we already have them as the reference signals. Then \mathbf{y}_r can be any linear combination of \mathbf{r} , so the form of the sub-matrix \mathbf{W}_{22} can be controlled to optimize the SBSS performance appropriately.

3. SOLUTION OF THE SBSS

The near-end echo paths, represented by the response matrix \mathbf{H}_{12} , can be identified through the SBSS by estimating the de-mixing matrix \mathbf{W} that maximizes the statistical independence of the output signals in \mathbf{y} . Any generic ICA algorithm can be used for the estimation of \mathbf{W} , but from now on the natural gradient algorithm will be considered, where \mathbf{W} is updated by iterating over the following formulas:

$$\mathbf{y}(\omega) = \mathbf{W}_{(n)}(\omega)\mathbf{x}(\omega), \quad (5)$$

$$\mathbf{W}_{(n+1)}(\omega) = \mathbf{W}_{(n)}(\omega) + \eta(\mathbf{I} - E[\Phi(\mathbf{y}(\omega))\mathbf{y}(\omega)^H])\mathbf{W}_{(n)}(\omega), \quad (6)$$

where η is the adaptation step-size, $\Phi(\cdot)$ is a non-linear function and $E[\cdot]$ is the expectation operator that can be approximated by averaging over time. A search for the solution converges when the gradient $\mathbf{I} - E[\Phi(\mathbf{y})\mathbf{y}^H]$ becomes null, i.e., when what we refer to as the *generalized covariance matrix* $E[\Phi(\mathbf{y})\mathbf{y}^H]$ becomes an identity matrix. By a Taylor expansion of the non-linear function $\Phi(\cdot)$, such a condition is achieved by minimizing each generic cross-moment of order α :

$$E[y_a(\omega)^\alpha y_b(\omega)^*] = 0, \quad \forall \alpha. \quad (7)$$

That is, the statistical independence for two source outputs y_a and y_b is achieved when the generalized covariance $E[\Phi(y_a)y_b^*]$ is null, as two zero-mean random variables can be considered independent if all of the high-order cross-cumulants are null [4].

Lets for the moment consider statistical independence between the separated sources vector \mathbf{y}_s associated with the near-end system and the separated sources vector \mathbf{y}_r associated with the reference signals. By applying (1), (2), (3), and (4) to (7), we obtain

$$E[\mathbf{y}_s(\omega)^\alpha \mathbf{y}_r^H(\omega)] = E[(\mathbf{W}_{11}(\omega)\mathbf{H}_{11}(\omega)\mathbf{s}(\omega) + (\mathbf{W}_{11}(\omega)\mathbf{H}_{12}(\omega) + \mathbf{W}_{12}(\omega))\mathbf{r}(\omega))^\alpha \mathbf{r}^H(\omega)\mathbf{W}_{22}^H(\omega)] = \mathbf{O} \quad (8) \quad \forall \alpha,$$

where \mathbf{y}_s^α indicates the raising of each element of the vector \mathbf{y}_s to the power α and \mathbf{y}_s^H denotes the Hermetian (conjugate) transpose of \mathbf{y}_s (i.e., the scalar sources y_a and y_b were substituted with the vectors of the factorized sources \mathbf{y}_s and \mathbf{y}_r). By applying the binomial expansion, we can rewrite (8) as

$$E[(\mathbf{W}_{11}(\omega)\mathbf{H}_{11}(\omega)\mathbf{s}(\omega))^\alpha \mathbf{r}^H(\omega)\mathbf{W}_{22}^H(\omega)] + E[(\mathbf{W}_{11}(\omega)\mathbf{H}_{12}(\omega) + \mathbf{W}_{12}(\omega))\mathbf{r}(\omega)]^\alpha \mathbf{r}^H(\omega)\mathbf{W}_{22}^H(\omega)] + E[(\sum_{k=1}^{\alpha-1} \frac{(\alpha-1)!}{k!(\alpha-1-k)!} (\mathbf{W}_{11}(\omega)\mathbf{H}_{11}(\omega)\mathbf{s}(\omega))^{\alpha-1-k} \odot ((\mathbf{W}_{11}(\omega)\mathbf{H}_{12}(\omega) + \mathbf{W}_{12}(\omega))\mathbf{r}(\omega))^k) \mathbf{r}^H(\omega)\mathbf{W}_{22}^H(\omega)] = \mathbf{O} \quad (9) \quad \forall \alpha,$$

where \odot indicates is the Hadamard (element-wise) product. By using the multinomial expansion to further expand the additive terms with powers α , $\alpha - 1 - k$, and k , it is possible to demonstrate that

if \mathbf{r} and \mathbf{s} are statistically independent from each other, the first and the third terms in (9) are null. In fact, all the matrix elements would be factorized as a sum of moments $E[s_i^\alpha r_j^\beta]$ that are zero for each α if s_i and r_j are zero-mean and mutually independent. It means the solution for \mathbf{H}_{12} that satisfies (8) does not depend on the near-end sources, and the optimization is possible even though both the near-end and the far-end sources are active at the same time (i.e., the double-talk situation). Then (9) can be simplified as

$$E[(\mathbf{W}_{11}(\omega)\mathbf{H}_{12}(\omega) + \mathbf{W}_{12}(\omega))\mathbf{G}(\omega)\mathbf{q}(\omega)]^\alpha \mathbf{q}^H(\omega) \mathbf{G}^H(\omega)\mathbf{W}_{22}^H(\omega)] = \mathbf{O}, \quad (10)$$

where we substituted the reference signals vector \mathbf{r} by considering the far-end mixing system \mathbf{G} and the far-end sources vector \mathbf{q} . Since the far-end sources are assumed to be statistically independent from each other (see Appendix), we can rewrite (10) as

$$[(\mathbf{W}_{11}(\omega)\mathbf{H}_{12}(\omega) + \mathbf{W}_{12}(\omega))\mathbf{G}(\omega)]^\alpha E[\mathbf{q}(\omega)^\alpha \mathbf{q}^H(\omega)] \mathbf{G}^H(\omega)\mathbf{W}_{22}^H(\omega) = \mathbf{O}, \quad (11)$$

where $E[\mathbf{q}^\alpha \mathbf{q}^H]$ is the generalized (high-order) autocorrelation matrix of the far-end sources that has a full rank since all of the sources are assumed to be statistically independent. If \mathbf{W}_{22} and \mathbf{G} are not singular, then (11) is satisfied when

$$\mathbf{W}_{11}(\omega)\mathbf{H}_{12}(\omega) + \mathbf{W}_{12}(\omega) = \mathbf{O}. \quad (12)$$

Finally, assuming that \mathbf{W}_{11} is known and invertible, \mathbf{H}_{12} can then be estimated as

$$\hat{\mathbf{H}}_{12}(\omega) = -\mathbf{W}_{11}(\omega)^{-1}\mathbf{W}_{12}(\omega). \quad (13)$$

By (10) we assume that there is always a solution $\mathbf{W}_{12} = -\mathbf{W}_{11}\mathbf{H}_{12}$ that maximizes the statistical independence of the output signals in \mathbf{y}_s . However, the exact echo path identification is only possible if \mathbf{W}_{11} , \mathbf{W}_{22} and \mathbf{G} are fully ranked. The singularity or the ill-conditioning of \mathbf{W}_{11} and \mathbf{W}_{22} is a rare occurrence if we assume spatial diversity for the near-end loudspeakers and talkers. The ill-conditioning of \mathbf{G} is a more serious problem than that of \mathbf{W}_{11} or \mathbf{W}_{22} since it occurs when the far-end talkers are located at the same position or, equivalently, if only one source is active at a time.

The dependence on the far-end mixing conditions hampers a stable on-line adaptation, which ultimately affects the identifiability of the echo paths. Thus a proper constraint to limit the fluctuation of the solution $\hat{\mathbf{H}}_{12}$ during an iterative optimization procedure becomes necessary.

4. EFFECT OF THE CONSTRAINT ON $\mathbf{W}(\omega)$

We assume that the far-end impulse response is always longer than the near-end modeling filter in the time domain. However, due to sparse and ill-representation of signals and mixing system in the frequency domain, the non-uniqueness problem may still exist in practice with respect to the number of far-end sources and microphones. Therefore, we need to consider two different mixing conditions at the far end separately to discuss the effect of a matrix constraint:

- (A) The number of active sources is equal to the number of microphones (2 far-end sources for the SAEC case).
- (B) The number of active sources is less than the number of microphones (1 far-end source for the SAEC case).

4.1. Case A

If all of the far-end sources are simultaneously active, the reference signals in \mathbf{r} should be linearly independent, although they are still statistically correlated according to the impulse responses corresponding to an individual source. Hence, the generalized covariance matrix of the ICA adaptation has a full rank, and the iterative update of \mathbf{W} can converge to a unique solution independently from a constraint on \mathbf{W}_{22} (e.g., $\mathbf{W}_{22} = \mathbf{I}$). From the point of view of the maximization of the statistical independence, the intrinsic decorrelation of the reference signals by the ICA would increase the convergence speed of the optimization procedure such that no constraints should be applied to \mathbf{W}_{22} . However, without any constraint, the matrix may accidentally approach to the singularity, which hampers the inversion of the matrix \mathbf{W} needed by the minimal distortion principle (MDP) to reduce the intrinsic scaling ambiguity of ICA. Since we are not interested in the final output components corresponding to the decorrelated reference signals, we can avoid the inversion of the entire de-mixing matrix \mathbf{W} and apply the MDP only for the separation of the near-end sources as:

$$\bar{\mathbf{W}}_{11}(\omega) = \text{diag}(\mathbf{W}_{11}^{-1}(\omega))\mathbf{W}_{11}(\omega). \quad (14)$$

4.2. Case B

The optimal solution for \mathbf{H}_{12} may not be unique when there are less number of sources than the number of microphones at the far end. Such a case corresponds to the near-singularity of the far-end response matrix \mathbf{G} or equivalently to the rank deficiency of the generalized autocovariance matrix $E[\Phi(\mathbf{r})\mathbf{r}^H]$. Nevertheless, although exploiting the higher-order statistics (HOS) cannot directly solve the non-uniqueness problem, the likelihood that the gradient of the ICA optimization cost would point towards a specific region in the solution space during a gradient-descent adaptation is strongly related to the structure of the de-mixing matrix and to the characteristics of the far-end impulse responses.

For example, if the far-end microphones are sufficiently spaced apart as in a realistic situation (e.g., 10 to 20 cm), the far-end impulse responses are already sparse in the time domain. The sparsity is not necessarily inherited from the time domain at each frequency in the frequency domain, but the frequency responses are likely to be only slightly correlated across frequency. Then we can approximate (after dropping ω for notational convenience)

$$[(\mathbf{W}_{11}\mathbf{H}_{12} + \mathbf{W}_{12})\mathbf{G}]^\alpha \simeq (\mathbf{W}_{11}\mathbf{H}_{12} + \mathbf{W}_{12})^\alpha \mathbf{G}^\alpha, \quad (15)$$

which can be derived as in Appendix by considering \mathbf{W}_{11} , \mathbf{H}_{12} and \mathbf{W}_{12} to be constant matrices and \mathbf{G} a matrix of zero-mean independent random variables, taking the expectation of $[(\mathbf{W}_{11}\mathbf{H}_{12} + \mathbf{W}_{12})\mathbf{G}]^\alpha$, and estimating $E[\mathbf{G}^\alpha]$ by \mathbf{G}^α . Also, since the far-end sources are assumed to be independent, the generalized covariance matrix $E[\mathbf{q}^\alpha \mathbf{q}^H]$ is expected to be diagonal. Thus we can approximate

$$\mathbf{D} = \mathbf{G}^\alpha E[\mathbf{q}^\alpha \mathbf{q}^H] \mathbf{G}^H \simeq \text{diag}\{E[q_i^\alpha q_i^*] \sum_j g_{ij}^\alpha g_{ij}^*\}, \quad (16)$$

where \mathbf{D} is a diagonal matrix and $*$ denotes the complex conjugation. Therefore, by assuming for simplicity $\mathbf{W}_{11} = \mathbf{I}$ (i.e., no near-end source separation is performed) and using the constraint $\mathbf{W}_{22} = \mathbf{I}$, (11) reduces to

$$E[\mathbf{y}_s \mathbf{y}_r^H] \simeq (\mathbf{H}_{12} + \mathbf{W}_{12})^\alpha \mathbf{D}. \quad (17)$$

It then becomes clear that with such a constraint on \mathbf{W}_{22} , the elements of the matrix \mathbf{W}_{12} are independently optimized. In other

words, the update direction during the gradient-descent optimization procedure for each element of \mathbf{W}_{12} is less likely to be affected by other elements that are related to different echo paths. Hence, the effect of the non-uniqueness problem is alleviated through the reduction in the ambiguity of physically allowed solution for the near-end echo paths.

We should point out that the diagonal constraint cannot completely solve the true non-uniqueness problem since (16) is only an approximation. Nonetheless, the constraint tends to globally bind the solution space of the time-domain filters related to \mathbf{W}_{12} , which consequently reduces the overall misalignment. In addition, fixing the diagonal elements of \mathbf{W}_{22} may introduce a divergence problem due to the norm of its gradient, in which case the problem is solved through the scaled natural gradient [5]. In such a version of the natural gradient algorithm, the de-mixing matrix is scaled at each iteration by a factor c in order to impose a *posterior* unit-norm constraint on $E[\Phi(\mathbf{y})\mathbf{y}^H]$. To apply the scaled natural gradient and still force \mathbf{W}_{22} to be diagonal, we need to impose the constraint $\Delta \mathbf{W}_{22} = 0$ (i.e., keep \mathbf{W}_{22} constant) and the initialization $\mathbf{W}_{(0)} = \mathbf{I}$.

5. EXPERIMENTAL RESULT

The SBSS algorithm was evaluated for the SAEC case, where the data were simulated in order to generate the worst case scenario from the misalignment's point of view: two far-end sources q_1 and q_2 alternate in activity, each being active for a long time (25 s), and do not change in position during the time. Impulse responses were simulated using different distances between the microphones. The simulated far-end impulse responses have $T_{60} = 300$ ms, and the filters for \mathbf{G} was truncated to 4096 taps. The simulated near-end impulse responses have a filter length of 3200 taps. The short-time Fourier transform (STFT) was applied to signals sampled at $f_s = 16$ kHz with Hanning windowing of 4096 taps with 75% overlap. The step-size and the non-linear function for the ICA were $\eta = 0.1$ and $\Phi(\cdot) = \tanh(10 \cdot |\mathbf{x}|) \exp(j\phi(\mathbf{x}))$, respectively, where \mathbf{x} is the the observed signal vector.

For evaluating the performance of the SBSS with or without the de-mixing matrix constraint, we considered the case of just one near-end source and microphone since we were interested only in the effect of the constraint on \mathbf{W}_{22} . We implemented a batch on-line adaptation, where each block b of \mathbf{x} is transformed into a time-frequency representation by the STFT, and the SBSS is applied independently for each frequency using a certain number of iteration *iter*. The estimate $\hat{\mathbf{W}}$ was averaged across blocks by an autoregressive model

$$\hat{\mathbf{W}}^{(b)}(\omega) = \gamma \cdot \hat{\mathbf{W}}^{(b-1)}(\omega) + (1 - \gamma) \cdot \mathbf{W}^{(b)}(\omega) \quad (18)$$

with a fixed step-size $\gamma = 0.9$, where $\mathbf{W}^{(b)}$ is the de-mixing matrix obtained in the block b . At each new block, a previously estimated $\hat{\mathbf{W}}$ was used to initialize the ICA. For each batch implementation, $\mathbf{W}^{(b)}$ was computed using non-overlapped blocks of 1 second in order to avoid more than one source being present within a same block to maintain the worst-case scenario of \mathbf{G} being always singular. The *true* echo return loss enhancement (tERLE), i.e., ERLE calculated after removing the near-end source signals, and the misalignment were computed as in [1]. For the computation of the tERLE, the filters obtained from the inversion of the de-mixing matrix at each frequency were transformed from non-causal to casual by a circular shifting of 2048 taps. Sample signals and results can be found at [6].

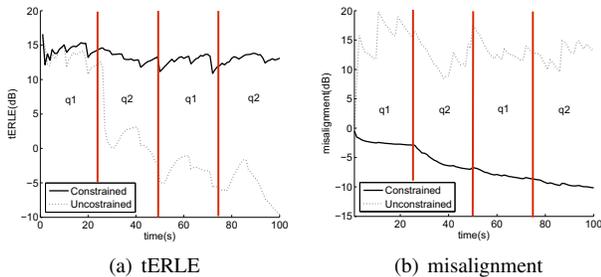


Figure 2: Comparison between constrained SBSS and unconstrained SBSS (iter=20).

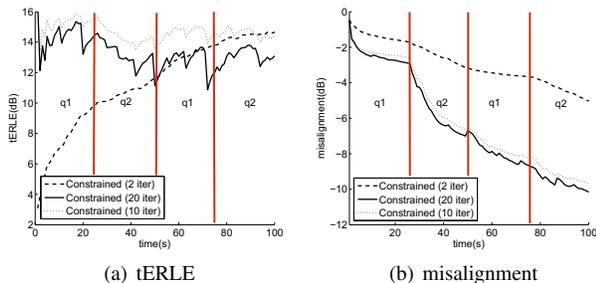


Figure 3: Performance comparison for different number of ICA iterations for each block.

Figure 2 shows the comparison between the constrained and the unconstrained SBSS for a far-end microphone spacing of 0.2 m. We note that for the constrained SBSS, the tERLE does not improve continuously during the adaptation but remains relatively stable around a value of 13-14 dB while the misalignment slowly decreases. When an active far-end talker switches to another every 25 s, the unconstrained SBSS evidently has a large degradation in the tERLE because $\hat{\mathbf{W}}_{12}$ depends on the far-end conditions that change at those times. A better interpretation can be obtained by analyzing the behavior of the unconstrained SBSS, for which we note that the misalignment is considerably high even though the tERLE is just as high as for the constrained SBSS during the first 25 s (only q_1 is active). It means that the ICA converges to a solution that is strongly dependent on the far-end mixing system, which generally corresponds to the estimate of non-causal filters and cannot be interpreted in the same way with the echo path that has a physical meaning at the near-end. Such a solution is then not any more valid for the talker q_2 than for the talker q_1 and explains such a degradation in the echo reduction when \mathbf{G} changes.

Another interesting comparison is shown in Figure 3, where only the constrained SBSS with a different number of the ICA iterations for each block is considered. We observe that as the number of iterations is increased, the tERLE converges quickly to a good solution but with a high variance during the time. In fact, since the non-uniqueness problem is not completely solved by the \mathbf{W}_{22} constraint, the variance of the convergence point in its solution space is directly proportional to the number of iterations. In other words, $\mathbf{W}^{(b)}$ depends less on its starting point as we increase the number of iterations, and the time-smoothed adaptation in (18) becomes less effective. On the other hand, by moving from a batch-wise to an on-line adaptation (i.e., 2 iter), we increase the stability of the adaptation but at the cost of a very slow convergence speed. Consequently in a batch on-line implementation, a trade-off between

stability and the convergence speed must be made.

6. CONCLUSION

We discussed the effect of the non-uniqueness problem in the semi-blind source separation (SBSS). We showed that the misalignment can be reduced by a proper constraint on the de-mixing matrix in a batch on-line ICA adaptation. Experimental results show that in the worst-scenario case of single far-end talker, a stable adaptation is possible without the need of any decorrelation procedure on the reference signals before being played through the near-end loudspeakers. Future investigation will consider studying the effect of the separation of multiple local sources and the interaction between the local sources and the reference signals.

APPENDIX

Assuming that $\mathbf{x} = \{x_j\}$ is a vector of zero-mean random variables of length N and that $\mathbf{A} = \{a_{ij}\}$ is an $N \times N$ matrix with constant elements, the statistical moment of order α for the i^{th} element of \mathbf{Ax} is given by

$$E \left[\left(\sum_{j=1}^N a_{ij} x_j \right)^\alpha \right] \quad \forall i. \quad (19)$$

By using the multinomial expansion, (19) can be rewritten as

$$E \left[\sum_{\substack{l_1, l_2, \dots, l_N \geq 0 \\ l_1 + l_2 + \dots + l_N = \alpha}} \frac{\alpha!}{l_1! \dots l_N!} \prod_{j=1}^N (a_{ij} x_j)^{l_j} \right] \quad \forall i. \quad (20)$$

If the elements in \mathbf{x} are mutually independent, (20) reduces to

$$E \left[\sum_j (a_{ij} x_j)^\alpha \right] = \sum_j a_{ij}^\alpha E[x_j^\alpha] \quad \forall i. \quad (21)$$

We can then generalize that

$$E[(\mathbf{Ax})^\alpha] = \mathbf{A}^\alpha E[\mathbf{x}^\alpha], \quad (22)$$

where \mathbf{x}^α and \mathbf{A}^α indicate the raising of each element of the vector \mathbf{x} and of the matrix \mathbf{A} to the power α . By the property of the covariance of linear combinations of variables, we know that if the random variables in \mathbf{x} are independent, then given the $N \times N$ matrices \mathbf{A} and \mathbf{B} , we have

$$E[\mathbf{Ax} \mathbf{x}^H \mathbf{B}] = \mathbf{A} E[\mathbf{x} \mathbf{x}^H] \mathbf{B}. \quad (23)$$

By using (20) and following the derivation of (22), it is possible to generalize (23) for higher-order moments as

$$E[(\mathbf{Ax})^\alpha \mathbf{x}^H \mathbf{B}] = \mathbf{A}^\alpha E[\mathbf{x}^\alpha \mathbf{x}^H] \mathbf{B}. \quad (24)$$

7. REFERENCES

- [1] T. S. Wada, S. Miyabe, and B.-H. Juang, "Use of decorrelation procedure for source and echo suppression," in *Proceedings of IWAENC*, Seattle, USA, Sept. 2008.
- [2] M. Joho, H. Mathis, and G. S. Moschytz, "Combined blind/nonblind source separation based on the natural gradient," *IEEE Signal Process. Letters*, vol. 8, no. 8, pp. 236–238, Aug. 2001.
- [3] J. Benesty, D. R. Morgan, and M. M. Sondhi, "A better understanding and an improved solution to the specific problems of stereophonic acoustic echo cancellation," *IEEE Trans. on Speech and Audio Process.*, vol. 6, no. 2, pp. 156–165, Mar. 1998.
- [4] A. Cichocki and S.-I. Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*. New York, NY, USA: John Wiley & Sons, Inc., 2002. [Online]. Available: <http://portal.acm.org/citation.cfm?id=863120>
- [5] S. Douglas and M. Gupta, "Scaled natural gradient algorithms for instantaneous and convolutive blind source separation," in *Proceedings of ICASSP*, vol. II, Apr. 2007, pp. 637–640.
- [6] [Online]. Available: <http://shine.fbk.eu/people/nesta>