SUBSPACE TRACKING OF MULTIPLE SOURCES AND ITS APPLICATION TO SPEAKERS EXTRACTION

Shmulik Markovich Golan[1], *Sharon Gannot*[1] *and Israel Cohen*[2]

[1] School of Engineering
Bar-Ilan University
Ramat-Gan, 52900, Israel
shmulik.markovich@gmail.com; gannot@eng.biu.ac.il

[2] Department of Electrical Engineering
Technion – Israel Institute of Technology
Technion City, Haifa 32000, Israel
icohen@ee.technion.ac.il

ABSTRACT

In this paper we introduce a novel algorithm for extracting desired speech signals uttered by moving speakers contaminated by competing speakers and stationary noise in a reverberant environment. The proposed beamformer uses eigenvectors spanning the desired and interference signals subspaces. It relaxes the common requirement on the activity patterns of the various sources. A novel mechanism for tracking the desired and interferences subspaces is proposed, based on the projection approximation subspace tracking (deflation) (PASTd) procedure and on a union of subspaces procedure. This contribution extends previously proposed methods to deal with multiple speakers in dynamic scenarios.

Index Terms— Subspace tracking, Speakers separation, Beamforming

1. INTRODUCTION

Speech enhancement techniques, utilizing microphone arrays, have attracted the attention of many researchers for the last thirty years, especially in hands-free communication tasks. A summary of several design criteria for beamformers can be found in [1]. While extensive research provided adequate solutions for speakers extraction in a static scenario, dynamic scenarios still pose a challenge.

In a recent contribution Markovich et al. [2] propose a linearly constrained minimum variance (LCMV) based beamformer for extracting the desired signals from multi-microphone measurements in a static scenario. The beamformer satisfies two sets of linear constraints. One set is dedicated to maintaining the desired signals, while the other set is chosen to mitigate both the stationary and non-stationary interferences. The proposed algorithm however is inappropriate for dynamic scenarios.

Affes et al. [3] construct a generalized sidelobe canceler (GSC) beamformer for the multi-source dynamic scenario. The proposed algorithm is based on the PASTd algorithm [4] for tracking the signals' subspace and on the multiple signal classification (MUSIC) algorithm for estimating the steering vectors for the sources. The far field regime and reverberation free environment allow tracking of the steering vectors during multi-speaker scenarios. However, its performance in reverberant scenarios is limited. Affes and Grenier [5] further develop a PASTd based algorithm for tracking changes in the acoustic transfer function (ATF) of a single desired source resulting from small scale movements of the speaker in a reverberant environment. They enhance speech signal that are contaminated by spatially white noise, assuming arbitrary ATFs relate the speaker and the microphone array. The algorithm proves to be efficient in a trading-room scenario, where the direct to reverberant ratio (DRR) is relatively high and the reverberation time is relatively short. Warsitz and Haeb-Umbach [6] use an alternative tracking procedure, based on the gradient ascent method, applied directly to the beamformer filters.

In the current contribution we adopt the PASTd algorithm for tracking non-static scenarios, presented in [2], in which multiple speakers coexist in a reverberant environment. The proposed algorithm is capable of extracting a desired conversation out of many conversations in time-varying and reverberant scenarios, where the expected DRR can be low.

The structure of the work is as follows. In Sec. 2 we formulate the speakers extraction problem. In Sec. 3 we extend the estimation algorithm proposed by Markovich et al. [2], and use an arbitrary subspace spanning the desired ATFs. This relaxes the common requirement for non-overlapping activity patterns of the desired sources. In Sec. 4 we introduce a novel mechanism for tracking the desired and interferences subspaces in a reverberant environment. The proposed speakers extraction algorithm is tested in both simulated and real environments in Sec. 5.

2. PROBLEM FORMULATION

Consider the problem of extracting N_d desired speech signals $s_1^d(n), \ldots, s_{N_d}^d(n)$ uttered by moving speakers contaminated by N_i competing moving speakers $s_1^i(n), \ldots, s_{N_i}^i(n)$ as well as stationary interferences in a reverberant environment. Each of the involved signals undergo filtering before being picked up by M microphones arranged in an arbitrary array. The reverberation effect can be modeled by a finite impulse response (FIR) time-varying filtering. The received signals can be formulated in a vector notation, in the short time Fourier transform (STFT) domain as $\boldsymbol{z}(\ell,k) = \boldsymbol{H}^d(\ell,k)\boldsymbol{s}^d(\ell,k) + \boldsymbol{H}^i(\ell,k)\boldsymbol{s}^i(\ell,k) + \boldsymbol{v}(\ell,k)$ where $\boldsymbol{s}^d(\ell,k) = \begin{bmatrix} s_1^d(\ell,k) & \cdots & s_{N_d}^d(\ell,k) \end{bmatrix}^T$ and $\boldsymbol{s}^i(\ell,k) = \begin{bmatrix} s_1^i(\ell,k) & \cdots & s_{N_i}^i(\ell,k) \end{bmatrix}^T$ are vectors comprising the desired and interfering speech signals, respectively. k denotes the frequency index and ℓ the frame index. $\boldsymbol{H}^d(\ell,k) = \begin{bmatrix} \boldsymbol{h}_1^d(\ell,k) & \cdots & \boldsymbol{h}_{N_d}^d(\ell,k) \end{bmatrix}$ and $\boldsymbol{H}^i(\ell,k) = \begin{bmatrix} \boldsymbol{h}_1^i(\ell,k) & \cdots & \boldsymbol{h}_{N_i}^i(\ell,k) \end{bmatrix}$ are $M \times N_d$ and $M \times N_i$ matrices that involve time-varying ATFs relating the desired and interfering sources and the microphone array. $\boldsymbol{v}(\ell,k)$ denotes stationary noise components of the received signals, consisting of directional as well as spatially white signals.

Assuming the sources and the noise signals are uncorrelated, the

978-1-4244-4296-6/10/$25.00 ©2010 IEEE

correlation matrix of the received signals can be written as:

$$\begin{aligned}\boldsymbol{\Phi}_{zz}(\ell,k) = \boldsymbol{H}^{d}(\ell,k)\boldsymbol{\Lambda}^{d}(\ell,k)\big(\boldsymbol{H}^{d}(\ell,k)\big)^{\dagger} \\ +\boldsymbol{H}^{i}(\ell,k)\boldsymbol{\Lambda}^{i}(\ell,k)\big(\boldsymbol{H}^{i}(\ell,k)\big)^{\dagger} + \boldsymbol{\Phi}_{vv}(\ell,k)\end{aligned} \qquad (1)$$

where $\boldsymbol{\Lambda}^{d}(\ell,k) \triangleq \operatorname{diag}\left(\left[\begin{array}{ccc}(\sigma_1^d(\ell,k))^2 & \ldots & (\sigma_{N_d}^d(\ell,k))^2\end{array}\right]\right)$ and $\boldsymbol{\Lambda}^{i}(\ell,k) \triangleq \operatorname{diag}\left(\left[\begin{array}{ccc}(\sigma_1^i(\ell,k))^2 & \ldots & (\sigma_{N_i}^i(\ell,k))^2\end{array}\right]\right)$ are diagonal matrices with the spectral variances of the desired and interfering sources on their main diagonal respectively. $\boldsymbol{\Phi}_{vv}(\ell,k)$ is the stationary noise correlation matrix. $(\bullet)^{\dagger}$ is the conjugate-transpose operation, and diag $(\bullet)$ is a square matrix with the vector in brackets on its main diagonal. In the following section we derive an algorithm for extracting the desired sources while mitigating the interferences in dynamic environments.

3. SPEAKERS EXTRACTION IN A DYNAMIC ENVIRONMENT

Markovich et al. [2] propose a novel eigenspace based LCMV beamformer, designed for extracting static desired sources. Rather than using the sources' ATFs for constructing the constraints set, they use an arbitrary basis for the interferences subspace and the relative transfer function (RTF)s of the desired sources. They also derive an algorithm for estimating the subspace, that spans the non-stationary interference signals, having an arbitrary activity pattern.

Following [2], define a modified constraints set $\dot{\boldsymbol{C}}^{\dagger}(\ell,k)\boldsymbol{w}(\ell,k) = \boldsymbol{g}(\ell,k)$ where $\dot{\boldsymbol{C}}(\ell,k) = \left[\begin{array}{cc}\tilde{\boldsymbol{H}}^{d}(\ell,k) & \boldsymbol{Q}^{i}(\ell,k)\end{array}\right]$ is the constraints matrix and $\boldsymbol{g}(\ell,k) \triangleq \left[\begin{array}{cc}\underbrace{1\ \ldots\ 1}_{N_d} & \underbrace{0\ \ldots\ 0}_{N_i}\end{array}\right]^{T}$ is the desired response vector. $\boldsymbol{Q}^{i}(\ell,k)$ denotes an orthonormal basis which spans the interferences subspace, i.e. $\boldsymbol{H}^{i}(\ell,k) = \boldsymbol{Q}^{i}(\ell,k)\boldsymbol{\Theta}^{i}(\ell,k)$ where $\boldsymbol{\Theta}^{i}(\ell,k)$ is the projection coefficients matrix. $\tilde{\boldsymbol{H}}^{d}(\ell,k) = \left[\begin{array}{ccc}\tilde{\boldsymbol{h}}_1^{d}(\ell,k) & \cdots & \tilde{\boldsymbol{h}}_{N_d}^{d}(\ell,k)\end{array}\right]$ denotes a matrix of the desired sources' RTFs with respect to reference microphone #1. The RTF of the ith desired source is defined as $\tilde{\boldsymbol{h}}_i^{d}(\ell,k) = \frac{1}{h_{i1}^{d}(\ell,k)}\boldsymbol{h}_i^{d}(\ell,k)$. The closed form beamformer solving this problem is given by:

$$\begin{aligned}\boldsymbol{w}(\ell,k) &= \boldsymbol{\Phi}_{zz}^{-1}(\ell,k)\dot{\boldsymbol{C}}(\ell,k) \\ &\times \left(\dot{\boldsymbol{C}}^{\dagger}(\ell,k)\boldsymbol{\Phi}_{zz}^{-1}(\ell,k)\dot{\boldsymbol{C}}(\ell,k)\right)^{-1}\boldsymbol{g}(\ell,k).\end{aligned} \qquad (2)$$

They further propose the use of the orthogonal triangular decomposition (QRD) procedure to perform the union of basis vectors obtained from several time segments. Estimating the constraint matrix utilizes segments of simultaneously active interference sources, but discards segments of desired signals' double-talk. In the sequel, we further relax the latter requirement, allowing simultaneously active desired sources in the estimation procedure.

Denote by $\boldsymbol{Q}^{d}(\ell,k)$ an orthonormal basis spanning the desired subspace $\boldsymbol{H}^{d}(\ell,k) = \boldsymbol{Q}^{d}(\ell,k)\boldsymbol{\Theta}^{d}(\ell,k)$ where $\boldsymbol{\Theta}^{d}(\ell,k)$ is the projection coefficients matrix. We propose to use the following modified constraints set

$$\tilde{\boldsymbol{C}}(\ell,k) = \left[\begin{array}{cc}\boldsymbol{Q}^{d}(\ell,k) & \boldsymbol{Q}^{i}(\ell,k)\end{array}\right] \qquad (3)$$

$$\tilde{\boldsymbol{g}}(\ell,k) = \left[\begin{array}{cc}\underbrace{\big(Q_{11}^{d}(\ell,k)\big)^{*}\ \ldots\ \big(Q_{N_d1}^{d}(\ell,k)\big)^{*}}_{N_d} & \underbrace{0\ \ldots\ 0}_{N_i}\end{array}\right]^{T} \qquad (4)$$

where we substitute the desired sources' RTFs in the constraints matrix $\dot{\boldsymbol{C}}(\ell,k)$ by the basis $\boldsymbol{Q}^{d}(\ell,k)$, and $\left[\begin{array}{ccc}1 & \cdots & 1\end{array}\right]_{1\times N_d}$ in the desired response vector $\boldsymbol{g}(\ell,k)$ by the first row of $\boldsymbol{Q}^{d}(\ell,k)$.

Let $\tilde{\boldsymbol{w}}(\ell,k)$ be the solution of the LCMV with the modified constraints set. The output of the modified beamformer is given by:

$$\begin{aligned}\tilde{y}_{\mathrm{BF}}(\ell,k) = \tilde{\boldsymbol{w}}^{\dagger}(\ell,k)\boldsymbol{z}(\ell,k) = \sum_{j=1}^{N_d} h_{j1}^{d}(\ell,k)s_j^{d}(\ell,k) \qquad (5) \\ +\tilde{\boldsymbol{g}}^{\dagger}(\ell,k)\big(\tilde{\boldsymbol{C}}^{\dagger}(\ell,k)\boldsymbol{\Phi}_{vv}^{-1}(\ell,k)\tilde{\boldsymbol{C}}(\ell,k)\big)^{-1}\tilde{\boldsymbol{C}}^{\dagger}(\ell,k)\boldsymbol{v}(\ell,k).\end{aligned}$$

Hence, the desired sources as received by the reference microphone are extracted, the non-stationary interferences are mitigated, and the power of the remaining stationary noise is minimized. Although the union based subspace estimation method obtains good performance with static sources, it is rendered useless when they are allowed to move, since the rank of the estimated subspace may excessively grow. Without prior knowledge of the rank, source movement, manifested as ATF change, results in a *birth* of a new basis vector. We circumvent this phenomenon by incorporating a *death* mechanism for the obsolete basis vectors in the estimation procedure. A novel subspace tracking algorithm utilizing birth and death mechanism is introduced in the following section.

4. PROPOSED SUBSPACE TRACKING ALGORITHM

The proposed tracking algorithm is based on the classic PASTd procedure introduced by Yang [4]. The PASTd procedure is a recursive algorithm incorporating a forgetting factor β. The latter results in an inherent memory of $N_\beta \approx \frac{1}{1-\beta}$ frames, contributing to the subspace estimation. The main limitation in applying the PASTd procedure to the problem at hand stems from conflicting memory requirements. On the one hand, we would like to apply PASTd with short memory in order to have fast adaptation time, and to quickly react to birth or death of basis vectors. On the other hand, using short memory, only recently active speakers will be included in the estimated subspace. All other speakers effectively die out. As a consequence, during the adaptation time, desired speakers that resume activity might suffer distortion, and competing speakers that resume activity may not be canceled out.

We propose to settle these contradicting requirements by using a short memory PASTd, allowing for fast adaption of basis vectors. Yet, basis vectors meeting certain conditions are declared *stable* and remain part of the estimated subspace for a predefined *expiry-time*. The stability conditions are explained in Sec. 4.2.

The proposed subspace tracking algorithm consists of three stages. First, a generalized PASTd procedure tracks the current subspace as explained in Sec. 4.1. Second, the expiry time is attributed to stable basis vectors. Third, the current basis vectors and the valid stable basis vectors are combined by using the union operation as explained in Sec. 4.3. A block diagram of the proposed tracking scheme is depicted in Fig. 1.

4.1. PASTd – Subspace Tracking

As we are dealing with two distinct groups of signals (desired and interfering) we apply the tracking algorithm to each group independently. Note that the proposed subspace tracking algorithm can only operate on time-segments in which desired and interfering speakers are mutually inactive. It is assumed that these segments exist and they are used for tracking the respective signal subspaces. Let x denote the active group, where $x \in \{d, i\}$. Define the activity indicator

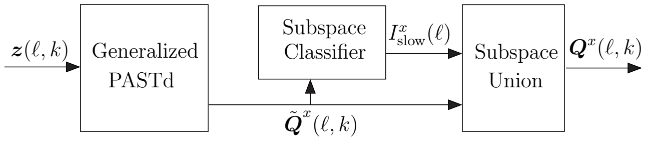

$\boldsymbol{z}(\ell,k)$
Generalized PASTd
Subspace Classifier
$I^x_{\text{slow}}(\ell)$
Subspace Union
$\boldsymbol{Q}^x(\ell,k)$
$\tilde{\boldsymbol{Q}}^x(\ell,k)$

Fig. 1. Block diagram of the proposed tracking algorithm

of the xth group

$$I^x(\ell) = \begin{cases} 1 & \text{only sources of the } x\text{th group are active} \\ 0 & \text{otherwise} \end{cases} . \qquad (6)$$

We assume that this activity indicator is available to the algorithm. Note, that a group x is declared active if at least one of its signals is active. The activity indicator $I^x(\ell)$ is regulating the subspace tracking algorithm.

PASTd estimation method is only suitable for tracking the signal subspace in a spatially white noise environment. Therefore, a whitening procedure should precede the activation of the tracking algorithm. Denote the whitened microphone signals as $\boldsymbol{z}_w(\ell,k) = \boldsymbol{\Phi}^{-1}_{vv,L}(\ell,k)\boldsymbol{z}(\ell,k)$, where $\boldsymbol{\Phi}_{vv,L}(\ell,k)$ is the lower triangular matrix obtained by the Cholesky decomposition of the stationary noise covariance matrix, $\boldsymbol{\Phi}_{vv}(\ell,k) = \boldsymbol{\Phi}_{vv,L}(\ell,k)\boldsymbol{\Phi}^{\dagger}_{vv,L}(\ell,k)$. The noise covariance matrix $\boldsymbol{\Phi}_{vv}(\ell,k)$ can be estimated by any conventional noise estimation procedure. The resulting covariance matrix of the whitened microphone signals is therefore given by:

$$\boldsymbol{\Phi}_{z_w z_w}(\ell,k) = \boldsymbol{\Phi}^{-1}_{vv,L}(\ell,k)\boldsymbol{\Phi}_{zz}(\ell,k)\big(\boldsymbol{\Phi}^{-1}_{vv,L}\big)^{\dagger}(\ell,k). \qquad (7)$$

The PASTd procedure tracks $N_u \leq M$ major eigenvectors of the two groups of the whitened sources $\{\boldsymbol{u}^x_r(\ell,k)\}_{r=1}^{N_u}$ and their corresponding eigenvalues $\{d^x_r(\ell,k)\}_{r=1}^{N_u}$. It is proven by Yang [4] that the estimated subspace converges to an orthonormal basis of the signals subspace.

A basis that spans the signal subspace of the original measurements $\boldsymbol{z}(\ell,k)$ is given by:

$$\tilde{\boldsymbol{u}}^x_r(\ell,k) = \sqrt{d^x_r(\ell,k)}\boldsymbol{\Phi}_{vv,L}(\ell,k)\boldsymbol{u}^x_r(\ell,k) \qquad (8)$$

where we scaled the basis vectors by their corresponding eigenvalues.

Note that this representation is no longer orthogonal. To obtain an orthogonal representation the following steps are applied. Define, $\tilde{\boldsymbol{U}}^x(\ell,k) \triangleq \begin{bmatrix} \tilde{u}^x_1(\ell,k) & \cdots & \tilde{u}^x_{N_u}(\ell,k) \end{bmatrix}$. Next, a QRD is applied to $\tilde{\boldsymbol{U}}^x(\ell,k)$. Finally, the required orthogonal basis $\tilde{\boldsymbol{Q}}^x(\ell,k)$ is obtained by selecting the dominant vectors spanning $\tilde{\boldsymbol{U}}^x(\ell,k)$ scaled by their corresponding energy.

4.2. Classification of Subspace Stability

The basis $\tilde{\boldsymbol{Q}}^x(\ell,k)$ defined in the previous section spans the subspace of the currently active sources in group x. Recall that this basis is always valid for at least N_β frames, due to the inherent memory of the PASTd technique. In a static scenario these basis vectors should remain unaltered. Based on this property, we propose a classification criterion for subspace stability. We define an indicator function

$$I^x_{\text{stable}}(\ell) \triangleq \begin{cases} 1 & \{\tilde{\boldsymbol{Q}}^x(\ell,k)\}_{k=0}^{N_{\text{DFT}}-1} \text{ is stable} \\ 0 & \text{otherwise} \end{cases} .$$

Each subspace that is valid for more than N_{stable} frames will be declared stable. Define the projection matrix to $\tilde{\boldsymbol{Q}}^x(\ell,k)$, the signal subspace, by

$$\boldsymbol{P}_{\tilde{Q}^x}(\ell,k) \triangleq \tilde{\boldsymbol{Q}}^x(\ell,k)\left(\big(\tilde{\boldsymbol{Q}}^x(\ell,k)\big)^{\dagger}\tilde{\boldsymbol{Q}}^x(\ell,k)\right)^{-1}\big(\tilde{\boldsymbol{Q}}^x(\ell,k)\big)^{\dagger}. \qquad (9)$$

The energy of the projection of the received signals in frame ℓ' to the current basis $\{\tilde{\boldsymbol{Q}}^x(\ell,k)\}_{k=0}^{N_{\text{DFT}}-1}$ is given by:

$$E_{\tilde{Q}^x}(\ell',\ell) \triangleq \sum_{k=0}^{N_{\text{DFT}}-1} \alpha^x(\ell,k)\|\boldsymbol{P}_{\tilde{Q}^x}(\ell,k)\boldsymbol{z}(\ell',k)\|^2 \qquad (10)$$

where $\alpha(\ell,k) \triangleq 1 - \frac{N_{\tilde{Q}^x(\ell,k)}}{M}$ is a compensation factor for high signal subspace rank. Hence, the aggregated projection energy over N_{stable} frames is given by:

$$E_{\tilde{Q}^x}(\ell) = \sum_{j=0}^{N_{\text{stable}}-1} E_{\tilde{Q}^x}(\ell - N_\beta - j, \ell).$$

Finally, we set $I^x_{\text{stable}}(\ell) = 1$ if $\frac{E_{\tilde{Q}^x}(\ell)}{E^x(\ell)}$ is higher than a predefined threshold, where $E^x(\ell)$ is the aggregated energy of the received signal over N_{stable} frames. Subspaces that are declared stable are attributed with an expiry-time. The expiry-time provides a mechanism for forgetting unused basis vectors.

4.3. Subspaces Union

To guarantee that basis vectors common to the current subspace and the stable subspaces are not counted more than once they should be collected by the union operator (see an analogue discussion in [2]). The union operator can be implemented in many ways. Here we chose to use the QRD. The required orthonormal basis $\boldsymbol{Q}^x(\ell,k)$ for group x is obtained by selecting the dominant vectors spanning the collection of valid subspaces. Note that the rank of the signal subspace is estimated from the received data and therefore the knowledge of N_d and N_i is not required.

5. EXPERIMENTAL STUDY

The proposed algorithm is tested with simulated signals as well as with real signals recorded in our acoustics lab. We examine a scenario in which two desired speakers and two interfering speakers are moving around in a reverberant noisy environment. The dimensions of the simulated room are $3m \times 4m \times 2.7m$. The reverberation time is set to $0.3s$ in both environments. The acoustics lab and the simulated room are depicted in Figs. 2, 3, respectively. The microphone array comprises 9 microphones and is arranged in a non-uniform linear array with total length of $0.64m$. The signal to interference ratio (SIR) (with respect to the non-stationary interferences) and signal to noise ratio (SNR) (with respect to the stationary interference) are 0dB and 30dB respectively. The sonogram and the waveform of the signal received by a reference microphone, in the acoustics lab scenario, are depicted in Fig. 4(a). The respective output of the proposed algorithm is depicted in Fig. 4(b). Comparing both signals, it is clearly seen that the interference signals are significantly attenuated, especially in high frequency bands. The SIR improvement in the acoustics lab, using the proposed algorithm, is 7.5dB, while in the simulated environment is 9.7dB.

Fig. 2. The acoustics lab at Bar-Ilan University premises.

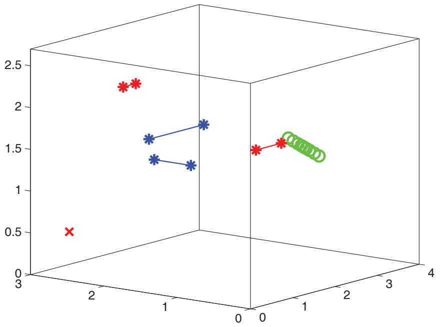


Fig. 3. The simulated scenario: Green circles denote the microphones. Blue and red stars denote desired and interfering sources, respectively. A line connecting two stars denotes the route of the source's movement. A red × denotes a stationary interference.

6. CONCLUSIONS

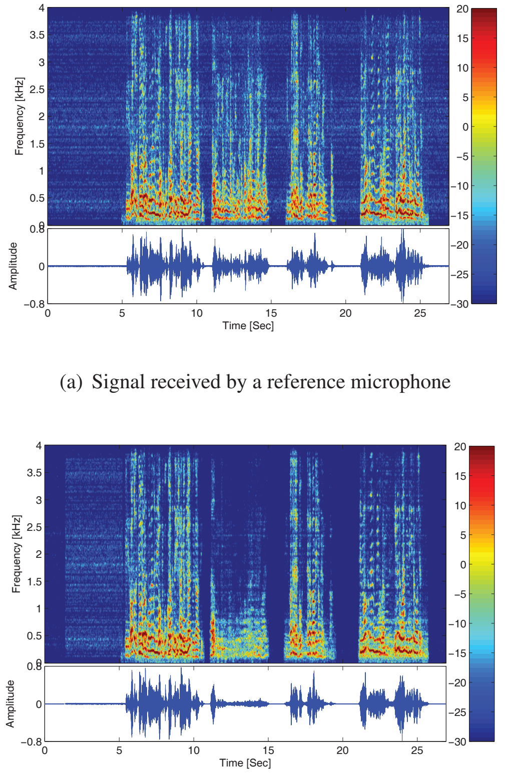
A novel algorithm for tracking signal subspaces has been introduced. The algorithm tracks the current subspace using the PASTd algorithm and classifies certain subspaces as stable. An expiry-time is then attributed to the stable subspaces. The union operator implemented by the QRD is used for collecting valid basis vectors, independently for the desired and the interfering groups of signals. The resulting signals subspaces are used to construct a beamformer for extracting desired sources in a dynamic environment. The proposed tracking algorithm relaxes limiting requirements on sources activity (common to other algorithms), and allows for simultaneous source activity within the groups. The novel algorithm is shown to yield good results both in real and simulated environments.

7. REFERENCES

[1] B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 5, no. 2, pp. 4–24, Apr. 1988.

[2] S. Markovich, S. Gannot, and I. Cohen, "Multichannel eigenspace beamforming in a reverberant environment with multiple interfering speech signals," *IEEE Trans. Audio, Speech and Language Processing*, vol. 17, no. 6, pp. 1071–1086, 2009.

[3] S. Affes, S. Gazor, and Y. Grenier, "An algorithm for multisource beamforming and multi-target tracking," *IEEE Trans. Signal Processing*, vol. 44, no. 6, pp. 1512–1522, Jun. 1996.



(a) Signal received by a reference microphone

(b) The output of the beamformer

Fig. 4. Received signal and the beamformer output in a real environment with moving sources

[4] B. Yang, "Projection approximation subspace tracking," *IEEE Trans. Signal Processing*, vol. 43, no. 1, pp. 95–107, Jan. 1995.

[5] S. Affes and Y. Grenier, "A signal subspace tracking algorithm for microphone array processing of speech," *IEEE Trans. Speech and Audio Processing*, vol. 5, no. 5, pp. 425–437, Sep. 1997.

[6] E. Warsitz and R. Haeb-Umbach, "Acoustic filter-and-sum beamforming by adaptive principal component analysis," *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 4, pp. 797–800, Mar. 2005.