Stationary Signal Separation Using Multichannel Local Segmentation

C. Castro-Hoyos, F.M. Grisales-Franco, J.D. Martínez-Vargas,
Carlos D. Acosta-Medina, and Germán Castellanos-Domínguez

Universidad Nacional de Colombia, Sede Manizales,
Signal Processing and Recognition Group
km 7 vía al Magdalena, Colombia
{ccastroh,fmgrisalesf,jmartinezv,cdacostam,cgcastellanosd}@unal.edu.co
http://portal.manizales.unal.edu.co/gta/signal/

Abstract. In this work, we study the influence of locally stationary segments as preprocess stage to separate stationary and non-stationary segments. To this, we compare three different segmentation approaches, namely i)cumulative variance based segmentation, ii)PCA based segmentation, and iii)HMM based segmentation. Results are measured as the true and false detection probabilities, and also as the ratio between the real and estimated number of segments. Finally, to achieve the separation, we use the Analytic Stationary Subspace Analysis (ASSA) and results are measured as the correlation between the true and the estimated stationary sources. In this case, we also compare against the best possible ASSA solution. Results show that inclusion of locally stationary segments could enhance or at least achieve optimal estimation of stationary sources.

1 Introduction

Stationary separation methods from linearly mixed signals are usually required in multiple pattern recognition and digital signal analysis applications like biomedical signal processing, Neurocomputing, and mechanical vibration monitoring systems. Indeed, non–stationary nature of signals certainly affects the extraction of informative components [10]. Therefore, the filtering task commonly assumes a separation model holding some stochastic constraints, so that the more stationary as possible component is extracted. For example, Authors in [2] develop a fast constrained least squares (LS) algorithm that minimizes the Kullback-Leibler divergence to search for stochastic changes. Nevertheless, the LS includes a smoothness condition that expels detection of fast and abrupt local non–stationary changes. Another important aspect of the separation filtering task is the computational burden. To illustrate, stationary source separation from single channel data can be carried out combining empirical mode decomposition and independent component analysis methods [7]. But, intrinsic mode functions demand high computational cost becoming exacerbated with large signals.

Mainly, the use of subspace methods during the separation task allows emphasizing those stationary structures hidden in the underlying random processes. Thus, the analytic stationary subspace analysis is discussed in [1,6] that divides the input non–stationary data into a fixed number of data segments to compute the windowed mean

E. Bayro-Corrochano and E. Hancock (Eds.): CIARP 2014, LNCS 8827, pp. 183–190, 2014.

and covariance matrixes. Although the subspace analysis helps revealing the signal temporal evolution, it assumes local stationarity, but evenly distributed along time; this assumption is mostly not realistic.

Here, we propose to improve techniques of multivariate stochastic segmentation by detecting in advance local stationary time series segments. Testing for locally stationary is carried out using estimators of mean and covariance. Validation of the proposed approach is accomplished on synthetic data holding time series that are generated as linear superposition of stationary and non–stationary sources. Results show that dynamic segmentation serves as an input basis for multivariate decomposition achieving similar results to those obtained with a large number of equally sized segments

2 Methods

2.1 Multivariate Signal Separation Filtering Task

Let matrix $\boldsymbol{X}^s \in \mathbb{R}^{N_c \times N_t}$ denote a multichannel stationary time-series, measured by N_c sensors at N_t time samples. We assume input data to be corrupted by an observed non-stationary multichannel signal $\boldsymbol{X}^n \in \mathbb{R}^{N_c \times N_t}$, so that the measured observation of linearly mixing signals is given by $\boldsymbol{X} = \boldsymbol{X}^s + \boldsymbol{X}^n$. The problem of separability is, by definition, to determine conditions on $\boldsymbol{X}^s$ and $\boldsymbol{X}^n$ such that an estimate of the desired signal $\widehat{\boldsymbol{X}}^s$ can be obtained, from filtered $\boldsymbol{X}$, to a given level of accuracy. Consequently, observed time-series $\boldsymbol{X}$ can be model as a linear superposition of stationary sources $\boldsymbol{S}^s \in \mathbb{R}^{N_s \times N_t}$ and non-stationary sources $\boldsymbol{S}^n \in \mathbb{R}^{N_n \times N_t}$, where N_s and N_n respectively denote the number of stationary and non-stationary sources as follows [1]:

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{S} = [\boldsymbol{A}^s \quad \boldsymbol{A}^n] \begin{bmatrix} \boldsymbol{S}^s \\ \boldsymbol{S}^n \end{bmatrix} \tag{1}$$

The problem of separability is, by definition, to determine conditions on $\boldsymbol{X}^s$ and $\boldsymbol{X}^n$ such that an estimate of the desired signal $\widehat{\boldsymbol{X}}^s$ can be obtained, from filtered $\boldsymbol{X}$, to a given degree of accuracy. Consequently, observed time-series $\boldsymbol{X}$ can be modeled as a linear superposition of stationary sources $\boldsymbol{S}^s \in \mathbb{R}^{N_s \times N_t}$ and non-stationary sources $\boldsymbol{S}^n \in \mathbb{R}^{N_n \times N_t}$, where N_s and N_n respectively denote the number of stationary and non-stationary sources as follows $\boldsymbol{A}^{-1} = \boldsymbol{B} = [\boldsymbol{B}^{s\top} \boldsymbol{B}^{n\top}]^\top$, where $\boldsymbol{S}^s = \boldsymbol{B}^s\boldsymbol{X}$, and $\boldsymbol{S}^n = \boldsymbol{B}^n\boldsymbol{X}$.

Thus, by splitting the N_t time samples into p segments, that is, splitting the time–series $\boldsymbol{X}$ into the set $\{\boldsymbol{\Pi}_i : \forall i \in p\}$ epochs, each one with mean $\boldsymbol{\mu}_i \in \mathbb{R}^{N_c \times 1}$ and covariance matrix $\boldsymbol{\Sigma}_i \in \mathbb{R}^{N_c \times N_c}$, we consider the time series to be stationary in the weak sense *iff* the corresponding values of epoch mean and covariance equal to the average: $\boldsymbol{u}_i = \bar{\boldsymbol{u}}$, and $\boldsymbol{\Sigma}_i = \bar{\boldsymbol{\Sigma}}$, where $\bar{\boldsymbol{u}} = \mathbb{E}\{\boldsymbol{u}_i : \forall p\}$ and $\bar{\boldsymbol{\Sigma}} = \mathbb{E}\{\boldsymbol{\Sigma}_i : \forall p\}$ are the average epoch mean and covariance matrix, respectively. Notation $\mathbb{E}\{\cdot\}$ stands for the expectation operator.

The above explained task can be expressed as the following optimization problem:

$$\min_{\boldsymbol{B}^s} \operatorname{tr}\left(\boldsymbol{B}^s \boldsymbol{S} \boldsymbol{B}^{s\top}\right); \quad \text{s.t.} \quad \boldsymbol{B}^s \bar{\boldsymbol{\Sigma}} \boldsymbol{B}^{s\top} = \boldsymbol{I}_{N_c}, \tag{2}$$

where $\boldsymbol{I}_{N_c}$ is an identity matrix. Reformulation of the optimization problem in terms of the Kullback-Leibler divergence provides the following estimation of $\boldsymbol{\Xi}$ [3]:

$$\boldsymbol{\Xi} = \mathbb{E}\left\{\boldsymbol{u}_i \boldsymbol{u}_i^\top + 2\boldsymbol{\Sigma}_i \bar{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\Sigma}_i\right\} - \bar{\boldsymbol{u}}\bar{\boldsymbol{u}}^\top - 2\bar{\boldsymbol{\Sigma}},$$

Here, the optimization task in Eq. (2) is represented by the generalized eigenvalue problem: $\boldsymbol{\Xi\Phi}{=}\lambda\bar{\boldsymbol{\Sigma}}\boldsymbol{\Phi}$, where solution is given by a set of $\lambda_j{\in}\mathbb{R}, \boldsymbol{\phi}_j{\in}\mathbb{R}^{N_c\times 1}: \forall j{\in}N_c$ generalized eigenvalues and $\bar{\boldsymbol{\Sigma}}$-orthonormal eigenvectors, where the stationary projection $\boldsymbol{B}^s$ is given by the N_s eigenvectors with smallest eigenvalues, $\boldsymbol{B}^s{=}[\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_{N_s}]^\top$, and the non-stationary projection is the remaining eigenvectors.

2.2 Stochastic Multichannel Segmentation

Provided a time series data $\boldsymbol{X}$, the p-segmentation procedure consists of searching a partition set having p non-overlapping segments with similar stochastic dynamics, $\{\boldsymbol{\Pi}_i{:}\forall{\in}p\}$ where each i-th segment $\boldsymbol{\Pi}_i \subseteq \boldsymbol{X}[a_i, b_i]$ goes from the sample a_i to b_i, being $a_1{=}1$ and $b_p{=}N_t$. Consequently, the $p-$th segment is selected when an introduced similarity measure overcomes a given threshold, $\zeta{\in}\mathbb{R}^+$. That is, stochastic segmentation implies detection of state changes produce by abrupt changes in the observed multichannel recordings. To this end, we consider the following known similarity measures:

i) *Cumulative variance-based similarity*: This is variance-based proximity measure between sliding overlapping segments and the mean ensemble, defined as [4]:

$$\mathbb{E}\left\{||\boldsymbol{\pi}_k - \boldsymbol{\mu}_i||^2 : \forall k{\in}\tau_i\right\}\begin{cases} < \zeta, & \text{there is no change,}\\ \geq \zeta, & \text{there is change,}\end{cases}$$

where the support $\tau_i{=}[a_i, b_i]$ and $\boldsymbol{\mu}_i{=}\mathbb{E}\left\{\boldsymbol{\pi}_k : \forall k \in \tau_i\right\}, \boldsymbol{\pi}_k{\in}\mathbb{R}^{N_c\times 1}$ is the $k-$th column of $\boldsymbol{\Pi}_i$, and notation $||\cdot||$ stands for the Euclidean distance.

ii) *PCA-based similarity*: This measure computes the homogeneity between the segment covariance matrix $\boldsymbol{V}_k{\in}\mathbb{R}^{N_c\times N_c}$ and the mean covariance $\boldsymbol{V}{=}\mathbb{E}\left\{\boldsymbol{V}_k{:}\forall k{\in}\tau_i\right\}$. As in [5], we make use of the PCA eigenvectors as $\mathbb{E}\left\{\mathrm{tr}\left(\hat{\boldsymbol{U}}^\top\hat{\boldsymbol{U}}_k\hat{\boldsymbol{U}}_k^\top\hat{\boldsymbol{U}}\right)\right\}/q$, where $\hat{\boldsymbol{U}}$ and $\hat{\boldsymbol{U}}_k{\in}\mathbb{R}^{N_c\times q}$ are the truncated eigenvector matrices of $\boldsymbol{V}$ and $\boldsymbol{V}_k$, respectively. The segmentation procedure using both, cumulative variance and PCA based similarity measures is depicted in Algorithm 1:

Algorithm 1. Sliding Window Cost-based Segmentation

For input time series $\boldsymbol{X} \in \mathbb{R}^{N_c\times N_t}$, initialize cost based model $f(\boldsymbol{\Pi}_i)$ with N_0 time samples, initial segment $a_1 = 1, b_1 = N_0, i = 1$ and set cost threshold value $\zeta \in \mathbb{R}^+$
while $b_i < N_t$ **do**
  Add k new time samples to the $i-$th segment and compute the required function parameters, i.e. mean vector for cumulative variance-based similarity and/or updated covariance matrix for PCA-based approach.
  if $f(\boldsymbol{\Pi}_i) < \zeta$ **then**
    $b_i = b_i + k$
  else
    Increment segment counter $i = i + 1$, update segment boundaries $a_i = b_{i-1} + 1$ and $b_i = b_{i-1} + N_0$.
  end if
end while

iii) *Hidden Markov Models (HMM) based segmentation*: Another measure that uses the posterior probability of a time sample vector $\boldsymbol{\pi}_k$ belonging to a given state [11]. Thus, assuming an HMM of length N_t, a state space dimension $w \in \mathbb{N}$ and the set of hidden state variables $\{z_1, \ldots, z_{N_t}\}$, the full true posterior of the model gives as a result the transition probability matrix $\boldsymbol{\Omega} = P(z_k | z_{k-1})$, with $\boldsymbol{\Omega} \in \mathbb{R}^{w \times w}$. Each matrix element $\left\{\omega_{i,j} \in \boldsymbol{\Omega} : \forall j, i = 1, \ldots, w\right\}$ describes the probability of transition from the i–th to the j–th state within the time interval $k-1$ and k, respectively.

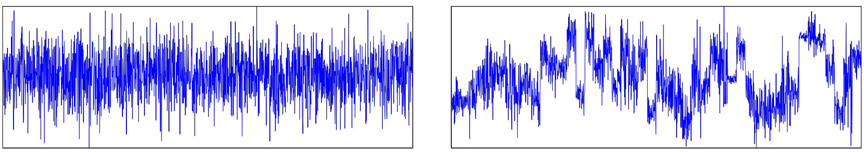
So, to detect abrupt changes in the observed recordings, the most probable a-posterior state at each sample time k is chosen using Viterbi decoding algorithm [9]:

$$u_k = \arg \max_{\forall i \in w} \{P(z_k = i | \boldsymbol{X})\} \tag{3}$$

Therefore, a detected change in the state time course vector, $\boldsymbol{u} = \{u_k : k = 1, \ldots, N_t\}$, provides the single i–th data segment $\boldsymbol{\Pi}_i$.

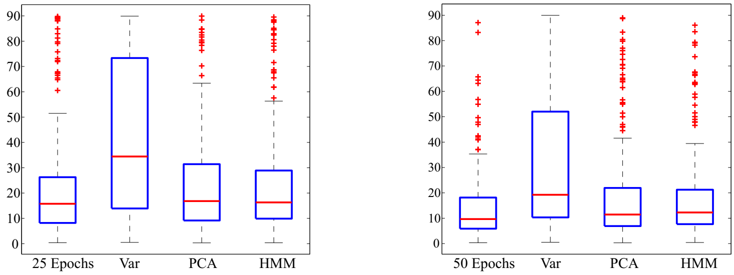
3 Experimental Set-Up

Dataset Description and Performance Measure: According to the model in Eq. (1), all observed time series are generated as linear superposition of simulated stationary and non–stationary sources, using the Stationary Subspace Analysis toolbox [8], publicly available. Stationary sources are generated from the normal distribution $\boldsymbol{X}^s \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$, where $\boldsymbol{I} \in \mathbb{R}^{N_s \times N_s}$ is an identity matrix. Non–stationary sources are generated and correlated with the stationary sources in each segment as: $\boldsymbol{X}_i^n = c_i \boldsymbol{X}_i^s + \boldsymbol{Y}_i^n$, where $\boldsymbol{Y}_i^n$ is created from a normal distribution $\boldsymbol{Y}_i^n \sim \mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$. Additionally $\boldsymbol{\mu}_i \in \mathbb{R}^{N_n \times 1}$ and $\boldsymbol{\Sigma}_i \in \mathbb{R}^{N_n \times N_n}$ are the randomly selected mean vectors and covariance matrices at each epoch, and $c_i \in \mathbb{R}$ is the canonical correlation among stationary and non–stationary sources. All necessary parameters for data generation are randomly selected. Experiments are carried out over 500 multivariate time series and signal length is selected within the interval [750, 2500] samples. The number of epochs ranges from 10 to 20 and the number of stationary and non–stationary sources range from 5 to 12 sources. The additional parameters, as minimum and maximum canonical correlation, source scaling, and others are selected from a normal distribution with zero mean and unitary standard deviation. Fig. 1 shows an example of randomly generated sources with 46 segments.



(a) Stationary Source. (b) Non–Stationary Source.

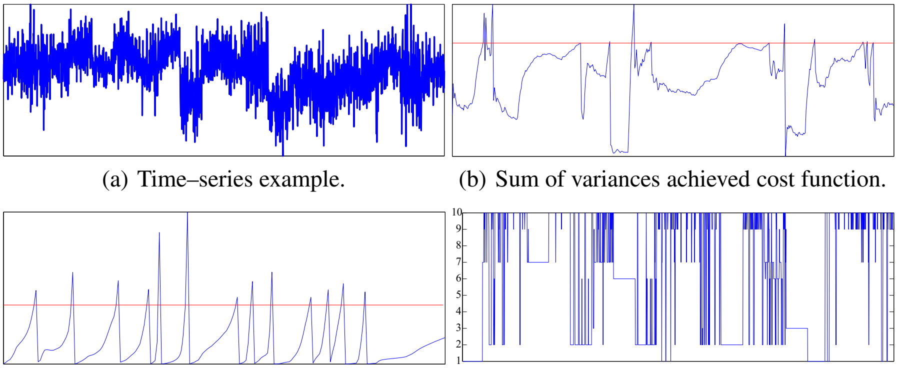
Fig. 1. Examples of one generated stationary and non-stationary real-valued time series



(a) 10 to 15 epochs (b) 16 to 20 epochs signals

Fig. 2. Boxplots of average canonical angle estimated ranging the number of epochs for all the proposed approaches

To evaluate the performance of ASSA under the proposed segmentation algorithms, we carry out the smallest canonical angle between the true non–stationary mixing subspace $\boldsymbol{A}^n$ and the estimated stationary projection subspace $\boldsymbol{B}^s$, that is, $90 - \theta(\hat{\boldsymbol{B}}^{s\top}, \boldsymbol{A}^n)$. The smallest canonical angle becomes zero for the perfect demixing case, so that the stationary projection is orthogonal to the non–stationary subspace. Also, in order to establish the influence on ASSA performance because of the dynamical segmentation, we carry out multivariate decomposition by fixing the optimal number of stationary and non-stationary sources. For testing, we compare against 50 equally sized epochs, where the segmentation parameters are tuned as follows: in the first case, the threshold is fixed (variance-based similarity) as the average variance of the multichannel data, $\zeta_\sigma = \mathbb{E}\left\{\sigma_i^2; \forall i=1,\ldots,N_c\right\}$, where σ_i^2 is the variance of the i–th channel along the time. In the second case (PCA based similarity), the threshold is empirically fixed as $\zeta_U=0.005$, while the number of principal components is set to $q=4$. In the last case (HMM-based segmentation), the embedding dimension space state is set to $w=10$ since this value is the minimum number of epochs to be found over the dataset. During calculation of the achieved performance, we consider separately signals comprising 10 to 15 epochs and 16 to 20 epochs, as explained before. As seen in Fig. 2(a) and Fig. 2(b) showing the obtained average canonical angle obtained, PCA and HMM-based segmentation methods achieve comparable values of canonical angle to the ones reached by the best ASSA solution that demands on a large amount of equally sized epochs. In fact, the angle difference does not exceed, on average, 3 - 5 degrees for either segmentation approach. In contrast, the cumulative variance based segmentation achieves in average low angles, but when signals get higher number of epochs. Yet, the performance of this approach also shows high variance. Fig. 3(a) shows an example of the first channel of a time series, while Fig. 3(b) and Fig. 3(c) show the estimated similarity values of the cumulative variance and PCA based methods. In case of the HMM based segmentation, Fig. 3(d) shows the state in each time-sample. It can be seen that PCA based segmentation is able to find more proper segments accordingly to the time-series. In addition, all segmentation approaches are evaluated in terms of true and false detection probabilities (p_d and p_f, respectively), as well as the ratio between the number of the obtained epochs to the number of epochs as follows:



(a) Time–series example.

(b) Sum of variances achieved cost function.

(c) PCA similarity based cost function.

(d) States transition of data samples.

Fig. 3. Segmentation results on simulated time series example, red lines indicate the threshold value

Table 1. Estimated performance of considered segmentation methods

	p	*Var*	*PCA*	*HMM*
p_d	10 to 15	0.79 ± 0.36	0.48 ± 0.15	0.82 ± 0.19
	16 to 20	0.84 ± 0.33	0.41 ± 0.14	0.86 ± 0.17
p_f	10 to 15	0.64 ± 0.11	0.67 ± 0.067	0.57 ± 0.051
	16 to 20	0.63 ± 0.11	0.71 ± 0.068	0.56 ± 0.038
r_e	10 to 15	18 ± 10	1 ± 0.19	2.6 ± 1.4
	16 to 20	19 ± 10	0.92 ± 0.16	3 ± 1.4

$$p_d = N_D/(N_D + N_M), \qquad p_f = N_F/(N_D + N_F), \qquad r_e = N_e/(\widehat{N}_e)$$

where N_D are the true starting samples, N_F the number of false starting samples (incorrectly labeled time-simple as the starting sample), N_M is the number of missing starting samples, and $\widehat{N}_e$ is estimate of the number of true segments, N_e.

Table 1 shows the average performance obtained by the three considered segmentation methods. Accordingly, the HMM-based segmentation gets the highest true detection probability and lowest false detection probability. As regards the cumulative variance approach, it achieves high segmentation performance, but, at the expense of a high number of segments obtained (high ratio), which is computationally expensive; Thus, the larger the amount of epochs – the higher the number of covariance matrices.

4 Discussion and Conclusion Remarks

In the current paper, we improve the quality of multivariate algorithms for stationary and non-stationary source separation by introducing a segmentation stage detecting intervals holding local stationarity. Experiments are carried out over simulated numerical random signals with multiple number of segments, stationary and non–stationary sources as well as random parameterizations of the data generator. As a concrete method of multivariate projection, we use the analytic stationary subspace analysis assuming the weak–sense stationary conditions upon a given multichannel time series. However, such kind of projections require equally-sized data segments for estimating model parameters, which in real time-series is an unpractical assumption. Instead, we propose an alternative methodology that provides an optimal number of non equally-sized segments.

We compare three local stationarity segmentation approaches as preprocess to the multivariate projection, and also we compare against a high number of equally sized segments, which in theory, may achieve the best possible results for the used separation procedure. The segmentation approaches are: i)segmentation based on cumulative variance over sliding windows, ii)PCA based segmentation, and iii)HMM based segmentation. The first case assumes changes in the temporal evolution of the variance computed in short segments of the signal. However, this approach is highly sensitive to small changes in the signal dynamics, yielding a high number of segments, reflected in the ratio (see Table 1). To overcome this problem, the second approach includes a better estimation of the signal variance, by computing recursively, over short segments, the eigenvalues and eigenvectors of the input data. Such introduced variance estimation improves segmentation results, which are reflected in lower average canonical angles and a lower ratio. Nevertheless, results are highly dependant on the threshold, that is empirically set. For the HMM based segmentation, only the state embedding dimension is required prior to the segmentation, which has an inverse relation with the computational burden of the approach. We set that dimension as 10 based on prior data knowledge. Results show better performance than the other approaches, according to the canonical angle (last columns of Figures 2(a) and 2(b)), and it also achieves the highest values of p_d and lowest values of p_f (Table 1). The main issue of this approach is the computational burden, that, even it is out of the scope of the paper, may pose a high restriction to implement in real time environments. In this work, we compare several approaches for locally stationarity segmentation. This segmentation is used as a preprocess stage to separate stationary and non-stationary sources. Results show that dynamic multivariate segmentation is able to enhance or at least achieve optimal decomposition. Additionally, the internal temporal dynamic evolution is taken into consideration altogether with the weak-sense stationary definitions. Thus, by setting more proper parameters an optimal dynamic segmentation can be found and directly influence ASSA estimations. As future work, we propose to research dynamic thresholding using the non-stationary degree of the data such that higher non-stationary segments are to be penalized stronger than low non-stationary segments. Additionally, we will aim to test the proposed dynamic segmentation with ASSA separation in real data such as electroencephalographic signals where stimuli-based segments are to be found.

Acknowledgments. This work is carried out under COLCIENCIAS grants: Programa Nacional de Formacion de Investigadores 2011 and 2013, and research project *Evaluación asistida de potenciales evocados cognitivos como marcador del transtorno por déficit de atensión e hiperactividad (TDAH).*

References

1. von Bünau, P., Meinecke, F.C., Király, F.C., Müller, K.R.: Finding stationary subspaces in multivariate time series. Phys. Rev. Lett. 103, 214101 (2009)
2. Das, N., Routray, A., Dash, P.: A constrained least squares algorithm for fast blind source separation in a non-stationary mixing environment. In: 2011 International Conference on Energy, Automation, and Signal (ICEAS), pp. 1–6 (2011)
3. Hara, S., Kawahara, Y., Washio, T., von Bünau, P., Tokunaga, T., Yumoto, K.: Separation of stationary and non-stationary sources with a generalized eigenvalue problem. Neural Networks 33, 7–20 (2012)
4. Himberg, J., Korpiaho, K., Mannila, H., Tikanmaki, J., Toivonen, H.: Time series segmentation for context recognition in mobile devices. In: Proceedings of the IEEE International Conference on Data Mining, ICDM 2001, pp. 203–210 (2001)
5. Krzanowski, W.J.: Between-groups comparison of principal components. Journal of the American Statistical Association 74(367), 703–707 (1979)
6. Martínez-Vargas, J., Sepulveda-Cano, L., Travieso-Gonzalez, C., Castellanos-Dominguez, G.: Detection of obstructive sleep apnoea using dynamic filter-banked features. Expert Systems with Applications 39(10), 9118–9128 (2012)
7. Mijovic, B., De Vos, M., Gligorijevic, I., Taelman, J., Van Huffel, S.: Source separation from single-channel recordings by combining empirical-mode decomposition and independent component analysis. IEEE Trans. on Biomedical Engineering 57(9), 2188–2196 (2010)
8. Müller, J.S., van Bünau, P., Meinecke, F.C., Király, F.J., Müller, K.R.: The stationary subspace analysis toolbox. J. Mach. Learn. Res. 12, 3065–3069 (2011)
9. Rezek, I., Roberts, S.: Ensemble hidden markov models with extended observation densities for biosignal analysis. In: Husmeier, D., Dybowski, R., Roberts, S. (eds.) Probabilistic Modeling in Bioinformatics and Medical Informatics. Advanced Information and Knowledge Processing, pp. 419–450. Springer, London (2005)
10. Sepulveda-Cano, L.M., Acosta-Medina, C.D., Castellanos-Dominguez, G.: Finite rank series modeling for discrimination of non-stationary signals. In: Alvarez, L., Mejail, M., Gomez, L., Jacobo, J. (eds.) CIARP 2012. LNCS, vol. 7441, pp. 691–698. Springer, Heidelberg (2012)
11. Woolrich, M.W., Baker, A., Luckhoo, H., Mohseni, H., Barnes, G., Brookes, M., Rezek, I.: Dynamic state allocation for MEG source reconstruction. NeuroImage 77, 77–92 (2013)