Signal Processing 96 (2014) 310–314

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Fast Communication

Approximation of the null distribution of the multiple coherence estimated with segment overlapping

Radoslav Bortel *, Pavel Sovka

Faculty of Electrical Engineering, Czech Technical University, Technická 2, 16627 Prague, Czech Republic

ARTICLE INFO

Article history:
Received 4 February 2011
Received in revised form
26 September 2013
Accepted 1 October 2013
Available online 11 October 2013

Keywords:
Multiple coherence
Null distribution
Segment overlapping

ABSTRACT

In this fast communication we suggest an approximation of the null distribution of the multiple coherence (MC) estimated with segment overlapping. The approximation is based on the formulas known for the non-overlapped segmentation, but the parameter corresponding to the number of segments is altered. The suggested approximation is statistically tested through a Monte Carlo simulation, and it is shown that its precision is quite high for a considerable range of MC parameters.

1. Introduction

In the multivariate signal analysis we often need to find if a given signal $y[n]$ is dependent on a group of other signals $x_i[n]$, $i=1,\ldots,K$. Such dependance can be examined with a statistic termed as the multiple coherence (MC), which is a measure of how well $y[n]$ can be linearly predicted from $x_i[n]$. The MC has found numerous applications, e.g. in multiple channel detection [17,5], biological signal analysis [11,8,13,9,10,16], mechanical system analysis [19,7] and other various fields [14].

In theory, if there is some linear dependance between $y[n]$ and $x_i[n]$, the true MC will be greater than zero. In practice, the true MC has to be estimated, and the resulting MC estimate has to be tested for its significance with statistical tests. For this purpose the statistical distribution of the MC estimate was presented e.g. in [6,17].

This statistical distribution, however, is limited by some assumptions. First, the analyzed signals are assumed to be multivariate Gaussian. Second, the spectral densities used for the MC estimation are assumed to be computed by averaging of *independent* signal segments. The latter assumption is unpleasantly limiting if we want to estimate the MC from a single record of a random process. In such case the spectral densities can be estimated with Welch's method [18]; however, the signal segmentation has to be performed without segment overlapping. But this is known to provide less precise spectral estimates than the segmentation with overlapping.[1] Therefore, it would be beneficial if we knew the statistical distribution of the MC estimate even when the MC is estimated from signal segments which are dependent due to their overlapping. The most interesting would be the null distribution (i.e. the distribution for the case where $y[n]$ is independent of $x_i[n]$ for each i), which allows to test if the sought dependance is present or not.

In this fast communication we will suggest an approximation of the null distribution of the MC estimated with segment overlapping.

[1] The usefulness of the overlapped segmentation was illustrated on the estimates of the magnitude squared coherence (MSC), which is a special case of the MC for $K=1$. In [4], it was shown that a proper use of the overlapped segmentation can decrease the variance and bias of the MSC estimate by the factor of 2. In [2] we also noted that a required MSC estimate variance can be achieved with half the data, when the overlapped segmentation is employed. Therefore, the use of the overlapped segmentation is quite beneficial.

* Corresponding author. Tel.: +420 737918312.
E-mail addresses: bortelr@feld.cvut.cz (R. Bortel), sovka@fel.cvut.cz (P. Sovka).

0165-1684/$ - see front matter
http://dx.doi.org/10.1016/j.sigpro.2013.10.003

2. Methods

2.1. MC definition

If $S_y(\Omega)$ is the power spectral density of $y[n]$, $S_{xy,i}(\Omega)$ are the cross spectral densities of $y[n]$ and $x_i[n]$, $S_{xx,ij}(\Omega)$ are the cross spectral densities of $x_i[n]$ and $x_j[n]$, and we define the matrices

$$\mathbf{S}_{xy}(\Omega) = [S_{xy,1}(\Omega), \ldots, S_{xy,K}(\Omega)]^T, \qquad \mathbf{S}_{xx}(\Omega) = [S_{xx,ij}(\Omega)]_{\substack{i=1\ldots K \\ j=1\ldots K}}, \tag{1}$$

then the MC can be defined as [3]

$$|\gamma_{xy}(\Omega)|^2 = \frac{\mathbf{S}_{xy}^H(\Omega)\mathbf{S}_{xx}^{-1}(\Omega)\mathbf{S}_{xy}(\Omega)}{S_y(\Omega)}, \tag{2}$$

where ^{H} denotes the complex conjugate.

Throughout the rest of this paper we will assume that $S_y(\Omega) > 0$, and $\mathbf{S}_{xx}(\Omega)$ is invertable, so that $|\gamma_{xy}(\Omega)|^2$ exists.

2.2. MC estimation

If $y[n]$ and $x_i[n]$ are N samples long records of a random stationary process, the MC can be estimated in the following way. First, $y[n]$ and $x_i[n]$ are segmented

$$\begin{aligned} x_{li}[n] &= x_i[(l-1)(M-P)+n] \\ y_l[n] &= y[(l-1)(M-P)+n] \end{aligned}, \quad n = 1, \ldots, M, \tag{3}$$

where M denotes the segment length, P denotes the segment overlap, and $l=1, \ldots, L$, where L is the number of segments. Then, we form the matrices

$$\boldsymbol{\mathcal{X}}(\Omega) = [X_{li}(\Omega)]_{\substack{l=1\ldots L \\ i=1\ldots K}}, \quad \boldsymbol{\mathcal{Y}}(\Omega) = [Y_1(\Omega), \ldots, Y_L(\Omega)]^T, \tag{4}$$

where $X_{li}(\Omega) = \mathcal{F}\{w[n]x_{li}[n]\}$ and $Y_l(\Omega) = \mathcal{F}\{w[n]y_l[n]\}$, where $\mathcal{F}$ denotes the discrete time Fourier transform [12, Eq. 2.134], and $w[n]$ is a chosen weighting window. Now, we can compute the estimates of the cross and power spectral densities

$$\hat{\mathbf{S}}_{xx}(\Omega) = \frac{1}{L}\boldsymbol{\mathcal{X}}^H(\Omega)\boldsymbol{\mathcal{X}}(\Omega), \quad \hat{\mathbf{S}}_{xy}(\Omega) = \frac{1}{L}\boldsymbol{\mathcal{X}}^H(\Omega)\boldsymbol{\mathcal{Y}}(\Omega),$$
$$\hat{S}_y(\Omega) = \frac{1}{L}\boldsymbol{\mathcal{Y}}^H(\Omega)\boldsymbol{\mathcal{Y}}(\Omega), \tag{5}$$

with which the MC estimate $|\hat{\gamma}_{xy}(\Omega)|^2$ will be given as

$$\begin{aligned} |\hat{\gamma}_{xy}(\Omega)|^2 &= \frac{\hat{\mathbf{S}}_{xy}^H(\Omega)\hat{\mathbf{S}}_{xx}^{-1}(\Omega)\hat{\mathbf{S}}_{xy}(\Omega)}{\hat{S}_y(\Omega)} \\ &= \frac{\boldsymbol{\mathcal{Y}}^H(\Omega)\boldsymbol{\mathcal{X}}(\Omega)(\boldsymbol{\mathcal{X}}^H(\Omega)\boldsymbol{\mathcal{X}}(\Omega))^{-1}\boldsymbol{\mathcal{X}}^H(\Omega)\boldsymbol{\mathcal{Y}}(\Omega)}{\boldsymbol{\mathcal{Y}}^H(\Omega)\boldsymbol{\mathcal{Y}}(\Omega)}. \end{aligned} \tag{6}$$

2.3. Statistical distribution of MC estimate

If $y[n]$ and $x_i[n]$ are multivariate Gaussian, and if the segmentation (3) is done without segment overlapping (i.e. $P=0$), then the statistical distribution of the MC estimate (6) is [6,17]

$$\begin{aligned} f(|\hat{\gamma}_{xy}|^2 | K, L, |\gamma_{xy}|^2) &= \frac{\Gamma(L)}{\Gamma(N-1)\Gamma(L-K+1)}(1-|\gamma_{xy}|^2)^L \\ &\cdot (|\hat{\gamma}_{xy}|^2)^{K-2}(1-|\hat{\gamma}_{xy}|^2)^{L-K}_2 F_1(L, L; K-1; |\hat{\gamma}_{xy}|^2|\gamma_{xy}|^2), \end{aligned} \tag{7}$$

where $f(|\hat{\gamma}_{xy}|^2 | K, L, |\gamma_{xy}|^2)$ denotes the probability density function (PDF) of $|\hat{\gamma}_{xy}(\Omega)|^2$ (the argument (Ω) was dropped to avoid clutter), $\Gamma(.)$ denotes the Gamma function, and ${}^1_2F(.,.;.;.)$ denotes the generalized hypergeometric function [1, Eq. 15.1.1].

In the case of the null distribution (i.e. when $|\gamma_{xy}(\Omega)|^2 = 0$), (7) reduces to the Beta distribution [1, Eq. 26.1.33] with parameters K and $L-K$

$$f_0(|\hat{\gamma}_{xy}(\Omega)|^2 | K, L) = b(|\hat{\gamma}_{xy}(\Omega)|^2 | K, L-K), \tag{8}$$

$$F_0(|\hat{\gamma}_{xy}(\Omega)|^2 | K, L) = B(|\hat{\gamma}_{xy}(\Omega)|^2 | K, L-K), \tag{9}$$

where $f_0(.|.,.)$ and $F_0(.|.,.)$ denote the PDF and the cumulative density function (CDF) of the null distribution of the MC estimate, and $b(.|.,.)$ and $B(.|.,.)$ are the PDF and CDF of the Beta distribution, respectively. Consequently, the MC estimate $|\hat{\gamma}_{xy}(\Omega)|^2$ is significantly greater than zero at the significance level α if it exceeds a threshold c given as the $1-\alpha$ quantile of the Beta distribution

$$c = F_0^{-1}(1-\alpha | K, L) = B^{-1}(1-\alpha | K, L-K). \tag{10}$$

In this paper we will adopt the nomenclature used with the magnitude squared coherence (MSC), and will refer to the threshold c as the confidence limit [2].

2.4. Generalization for overlapped segmentation

If $y[n]$ and $x_i[n]$ are multivariate Gaussian, but the segmentation (3) is done with overlapping (i.e. $P > 0$), then the MC estimate becomes more precise, but the statistical distribution (7), (8), and (9) and the confidence limit formula (10) are no longer valid.

To allow statistical evaluation of this more precise MC estimate, we will now attempt to approximate its statistical distribution. We will try to base this approximation on formulas (9) and (8), but with their parameters altered.

There are two parameters, K and L, which decide the shape of (8) and (9). K is given by the number of signals, and will probably not need changing, when we use the overlapped segmentation. On the other hand, the parameter L, which is given by the number of independent segments, will most likely need increasing when the overlapping is introduced.

To specify how much L should be increased, we can use the fact that for $K=1$, the MC reduces to the magnitude squared coherence (MSC). The distribution of the MSC estimated with the overlapped segmentation was already approximated in [2] using the same principle (i.e. the reuse of the statistical distribution known for the non-overlapped segmentation), so we can try to draw the new value of L from [2].

Specifically, if

(i) M-1 zeros are padded before and after the examined signals,

(ii) the segment overlap is no smaller than the segment overlap used for the short time Fourier transform that uses the window $w[n]$ (e.g. this is 75% overlap for the Hamming, Hanning and Bartlett windows, 84% overlap for the Blackman window, and 85% overlap for the Kaiser window with parameter ($\beta=10$),