Signal Processing 105 (2014) 338–349

Contents lists available at ScienceDirect

Signal Processing

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

Multi-channel post-filtering based on spatial coherence measure

Jwu-Sheng Hu *, Ming-Tang Lee

Department of Electrical Engineering, National Chiao-Tung University, Hsinchu, Taiwan

ARTICLE INFO

Article history:
Received 7 October 2013
Received in revised form
15 March 2014
Accepted 17 April 2014
Available online 15 May 2014

Keywords:
Coherence
Multi-channel post-filtering
Microphone array
Multi-rank signal model

ABSTRACT

A multi-channel post-filtering algorithm using the proposed spatial coherence measure is derived. The spatial coherence measure evaluates the similarity between the measured signal fields using power spectral density matrices. In the proposed post-filter, the assumption of homogeneous sound fields is relaxed. Besides, multi-rank signal models can be easily adopted. Under this measure, the bias term due to the similarity of the desired signal field and the noise field is further investigated and a solution based on bias compensation is proposed. It can be shown that the compensated solution is equivalent to the optimal Wiener filter if the bias or the noise power spectral density matrix is perfectly measured. Simulations with incoherent, diffuse, and coherent noise fields and a local scattered desired source were conducted to evaluate the algorithms. The results demonstrate the superiority of the proposed bias compensated post-filter across different types of noise fields with a more accurate signal model.

1. Introduction

Multi-channel speech enhancement has attracted much attention in recent years. In the real world, desired speech signals are often corrupted by background noises, speech interferences, and reverberation. For more than two microphones, there are two main categories of speech enhancement approach: beamforming and multi-channel post-filtering. Beamforming has been applied to several narrow- or wide-band signals processes, which can be defined by a *filter-and-sum* process [1] in the conventional sense. A well-known designing strategy is to preserve the signal from the direction of interest while attenuating others, which can be achieved by the minimum variance distortionless response (MVDR) algorithm [2,3]. The MVDR beamforming is optimal in the mean square error (MSE) sense when the interference-plus-noise power spectral density (PSD) matrix can be obtained and there is no mismatch on the presumed steering vector. Typically, adaptive filtering techniques are applied to estimate the PSD matrix and additional training processes or *a priori* information of signal presence is needed for offline or online implementation [1–4]. On the other hand, the multi-channel post-filtering, which considers both the spatial information and the signal-to-noise ratio (SNR), can be designed in a more general way. Simmer et al. [5] show that the optimal minimum mean square error (MMSE) solution can be decomposed into an MVDR beamformer followed by a single-channel Wiener filter. This solution is also called a *multi-channel Wiener filter*.

Most post-filtering algorithms aim to enhance the single-channel Wiener filter by a more accurate estimation of SNR. The SNR estimation for speech enhancement can be implemented based on the minimum statistics for the stationary noise [6–8], or the spatially pre-processed power [9]. Most of them are energy-based. Alternatively, the phase information of a microphone pair has already

* Corresponding author.
E-mail addresses: jshu@cn.nctu.edu.tw (J.-S. Hu), lhoney.ece97g@g2.nctu.edu.tw (M.-T. Lee).

http://dx.doi.org/10.1016/j.sigpro.2014.04.020
0165-1684/

been used in blind source separation (BSS) [10] as well as the computational auditory scene analysis (CASA) [11]. Aarabi et al. [12–14] provide a different view of the SNR from the phase error perspective for the dual-channel case. In their work, the relationship between the phase error and the SNR was derived [12]. However, the idea of phase error can only be applied to the case of two-microphone. In addition to the SNR estimation, some post-filtering algorithms directly estimate the spectral densities [15–17]. Like the case of phase error, the cross-spectral density is usually defined between two microphones. For more than two microphones, the common practice is to perform average among all distinct microphone pairs [15,16]. Although this might enhance the robustness of the estimation, there is still no formal proof regarding its effectiveness. In particular, it does not consider the spatial arrangement of microphones, i.e., the advantages of using more than two microphones is not fully explored. In this paper, a new spatial measure is defined on a microphone array which leads to a novel post-filtering algorithm (named *spatial coherence based post-filter*, SCPF). The post-filter belongs to the class of spectral densities estimators (which is inherent in the estimation of the input PSD matrix), while it is guaranteed to lie in the range of [0, 1]. Further, the proposed spatial coherence measure can be easily extended to multi-rank signal models encompassing incoherently scattered source, etc. Multi-rank signal models or rank relaxation has been widely used in sensor array localization [18–21], beamforming [22–25], or quadratic optimization problems [25,26]. It is more convenient to consider various design requirements than previous methods using microphone array.

However, a bias term due to the similarity of the desired signal field and the noise field deteriorates the noise reduction performance. As a result, a bias compensated method is proposed (called *bias compensated spatial coherence based post-filter*, BC-SCPF). It can be shown that the BC-SCPF is equivalent to the optimal Wiener filter if the bias or the noise PSD matrix is perfectly measured. Three kinds of noise fields were used with a local scattered source for analysis: incoherent, diffuse, and coherent. Three ITU-T standards were computed to evaluate the perceptual quality and the noise reduction performance. The simulation results show the superiority of the proposed BC-SCPF with a more accurate signal model in all noise fields comparing with various methods proposed before.

The paper is organized as follows. Section 2 states the objective and reviews some related works. In Section 3, a trace inequality is introduced and a coherence measure is defined based on it. The SCPF and BC-SCPF are proposed in Section 4. The simulation setup and results with three noise fields are presented in Section 5, and Section 6 gives the conclusion.

2. Problem formulation and prior works

2.1. Problem formulation

Consider a linear array with M omni-directional microphones. The observation vector is given by

$$\mathbf{x}(t) = \mathbf{s}(t) + \mathbf{n}(t) \tag{1}$$

where $\mathbf{s}(t)$ and $\mathbf{n}(t)$ are the desired signal and noise. Both of them can be multi-dimensional. By assuming locally time-invariant transfer functions and applying the short-time Fourier transform (STFT), the observations are divided in time into overlapping frames by the application of a window function and analyzed in the time–frequency domain as,

$$\mathbf{x}(\omega, k) = \mathbf{s}(\omega, k) + \mathbf{n}(\omega, k) \tag{2}$$

where ω and k are discrete frequency and frame indices respectively.

A beamforming method aims to find a spatial filter $\mathbf{w}$ to estimate the desired source by

$$y(\omega, k) = \mathbf{w}^H(\omega, k)\mathbf{x}(\omega, k) \tag{3}$$

A post-filtering method aims to find a gain function (or mask) to suppress the undesired noise, which can be multiplied on the beamformer output as

$$\hat{s}(\omega, k) = G(\omega, k) \cdot y(\omega, k) \tag{4}$$

2.2. Multi-channel post-filtering based on noise field coherence

McCowan et al. [16] proposed a multi-channel post-filter as a modification of the Zelinski post-filter [15]. In their systems, the microphones have to pass a time alignment module to adjust the propagation of the desired source between microphones before the post-filter estimation, which is equivalent to the information in the presumed steering vector $\mathbf{a}_s(\omega)$. That is, the pre-processed input vector $\tilde{\mathbf{x}}(\omega, k)$ after the time alignment module can be written as

$$\tilde{\mathbf{x}}(\omega, k) = \mathbf{x}(\omega, k) \circ \mathbf{a}_s(\omega) \tag{5}$$

where $\circ$ denotes the Schur–Hadamard (elementwise) matrix product. If the desired signal is a point source, the presumed steering vector $\boldsymbol{a}_s(\omega)$ can be equivalent to the truncated impulse response $\boldsymbol{h}(\omega)$ if the magnitudes and time delays of the source to the microphones are exactly measured. However, the steering vector is not sufficient to describe general cases, which will be discussed in detail in Section 4.1.

Compared to the Zelinski post-filter, the work in [16] considered a generalized coherence function to describe the characteristics of the noise field on the aligned inputs. Noises between sensors can be coherent (or correlated). The noise coherence function of the time aligned inputs is defined as

$$\tilde{\Gamma}_{n_i n_j}(\omega) = \tilde{\phi}_{n_i n_j}(\omega) \Big/ \sqrt{\tilde{\phi}_{n_i n_i}(\omega) \cdot \tilde{\phi}_{n_j n_j}(\omega)} \tag{6}$$

where $\tilde{\phi}_{n_i n_j}(\omega)$ is the cross-spectral density between the noises at the i-th and j-th microphones. Note that the diagonal terms of $\tilde{\mathbf{\Gamma}}_n(\omega)$ are 1 and its trace equals to M. In their works, the homogeneous sound fields are assumed. That is, the sources have the same power spectrum at each sensor. Based on this assumption, the spectral densities of the aligned inputs are expressed as [16]

$$\tilde{\phi}_{x_i x_i}(\omega) = \tilde{\phi}_s(\omega) + \tilde{\phi}_n(\omega) \tag{7}$$

$$\tilde{\phi}_{x_j x_j}(\omega) = \tilde{\phi}_s(\omega) + \tilde{\phi}_n(\omega) \tag{8}$$