Signal Processing 107 (2015) 185–196

Contents lists available at ScienceDirect

Signal Processing

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

Blind compensation of interchannel sampling frequency mismatch for ad hoc microphone array based on maximum likelihood estimation

Shigeki Miyabe [a,*], Nobutaka Ono [b], Shoji Makino [a,1]

[a] Life Science Center of Tsukuba Advanced Research Alliance, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan
[b] Principles of Informatics Research Division, National Institute of Informatics/Department of Informatics, School of Multidisciplinary Sciences, The Graduate University for Advanced Studies (SOKENDAI), 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan

ARTICLE INFO

Article history:
Received 17 February 2014
Received in revised form
10 September 2014
Accepted 11 September 2014
Available online 22 September 2014

Keywords:
Ad hoc microphone array
Drift
Sampling frequency
Maximum likelihood estimation
Blind source separation

ABSTRACT

In this paper, we propose a novel method for the blind compensation of drift for the asynchronous recording of an ad hoc microphone array. Digital signals simultaneously observed by different recording devices have drift of the time differences between the observation channels because of the sampling frequency mismatch among the devices. On the basis of a model in which the time difference is constant within each short time frame but varies in proportion to the central time of the frame, the effect of the sampling frequency mismatch can be compensated in the short-time Fourier transform (STFT) domain by a linear phase shift. By assuming that the sources are motionless and have stationary amplitudes, the observation is regarded as being stationary when drift does not occur. Thus, we formulate a likelihood to evaluate the stationarity in the STFT domain to evaluate the compensation of drift. The maximum likelihood estimation is obtained effectively by a golden section search. Using the estimated parameters, we compensate the drift by STFT analysis with a noninteger frame shift. The effectiveness of the proposed blind drift compensation method is evaluated in an experiment in which artificial drift is generated.

© 2014 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/).

1. Introduction

Microphone array signal processing is a framework for analyzing spatial information of a sound field observed with multiple microphones to perform speech enhancement, source separation, source localization, and so forth, which are difficult by the processing of single-channel observations [1]. Microphone arrays are used in various applications, including teleconferencing, hands-free speech recognition, hearing aids, acoustic monitoring, spatial audio, and computer games. While conventional microphone array signal processing assumes that multichannel signals are observed by a unified analog-to-digital converter (ADC), recently increasing attention has been focused on an extension of the microphone array framework, the so-called ad hoc microphone array, where a combination of observations by independent recording devices is treated as multichannel recording [2]. The non-necessity of wired channels to achieve synchronization enables the downsizing of the recording devices, which is an important attribute for hearing aids [3]. In addition, this framework is suitable for recording meetings because of the easy construction of the microphone array by the combination of widely available portable recording devices, such as cell phones, IC recorders, and video cameras, and the freedom of the microphone arrangement, which enables recording with a

* Corresponding author. Tel.: +81 29 853 6566; fax: +81 29 853 7387.
E-mail address: miyabe@tara.tsukuba.ac.jp (S. Miyabe).
[1] The following co-author is a member of EURASIP: Shoji Makino.

http://dx.doi.org/10.1016/j.sigpro.2014.09.015
0165-1684/© 2014 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/).

high SNR by setting each device close to each speaker [2,4]. Also, considerable effort has been made to develop wireless acoustic sensor networks (WASNs), where the recording devices are connected by wireless networks [5].

However, the increased freedom of ad hoc microphone arrays raises various issues that do not arise in conventional array signal processing. For example, the array geometry is unknown [2,6–8], the recording devices have different unknown gains [2], each device starts recording independently [7,8], and the sampling frequencies are not common among the observation channels [9–15]. Also, in WASNs, the efficiencies of communication and distributed computation are important issues to achieve array signal processing with a limited bandwidth and array nodes with low computational power [5].

Among these issues of ad hoc microphone arrays, one of the most important is the mismatch of sampling frequencies. Since each ADC is not synchronized with the others, the individual variability of clocks results in a slight mismatch of the sampling frequencies, causing a change in the time difference between channels due to the constant skew, the so-called *drift*. Array signal processing generally assumes that the sources do not move and utilizes the phase differences inherent to the positions of the sources. However, drift causes the phase differences to constantly change as if the sources are moving, preventing the use of array signal processing to analyze phase differences assuming static sources [10,11]. Also, the asynchronous recording causes offsets of the recording start time. Estimation and compensation of the sampling frequency mismatch and recording start offset are indispensable as preprocesses in array signal processing.

While there are two types of methods for estimating the sampling frequency mismatch, i.e., supervised and unsupervised, our research focuses on unsupervised estimation. The former, supervised estimation, mainly estimates the relation between the clock time of the ADC and the absolute time by using the time stamp received from satellites or wireless networks [9,13–15]. While the advantage of this approach is that the recording start offset can be estimated in addition to the sampling frequency mismatch, the disadvantage is the constraint that the recording devices must have the ability to receive the time stamp. Another problem is that the accuracy of the time stamp is generally much lower than that required in array signal processing, and training involving the long observation of time stamps is required for accurate estimation. Unsupervised estimation cannot obtain a precise estimate of the recording start offset without prior information. Inaccurate compensation of the recording start offset is problematic with particular classes of supervised array signal processing which receive time differences of arrival (TDOAs) of positions of specific sources as prior information because the offset changes the relation between the TDOAs and the positions. However, in a blind scenario of array signal processing such as blind source separation (BSS) [16], where only the observation is given, a rough compensation of the recording start offset is sufficient. Therefore, in this paper we focus on the accurate compensation of the drift and the rough compensation of the recording start offset in an unsupervised manner.

To the best of our knowledge, there have been few works on the blind estimation of sampling frequency mismatch. Liu et al. proposed a method of estimation involving the iteration of independent component analysis (ICA) and evaluation of the correlation between the estimated independent components by utilizing the property that ICA can extract uncorrelated independent sources only when the drift is well compensated [10]. However, its applicability is limited to determined systems with equal numbers of sources and microphones so that the independent components can be extracted. Markovich-Golan et al. estimated the sampling frequency mismatch as the rate of change of the phase in the interchannel correlation of a noise observation [12]. The modeling of the sampling frequency mismatch using the phase is very similar to our proposed framework, which is discussed later. However, the scenario of this method, which assumes voice activity detection (VAD), is different from our fully blind estimation scenario.

In this paper, we propose a novel method for blind drift compensation by maximum likelihood estimation of the sampling frequency mismatch in the short-time Fourier transform (STFT) domain. The basic idea of the maximum likelihood estimation is published in a previous conference proceedings paper [17]. The optimization algorithm is followed by resampling as a modification of the STFT analysis with a noninteger frame shift, which we proposed in another conference proceedings paper [18]. In addition, we newly propose an iterative algorithm for estimating the recording start offset and sampling frequency mismatch, which enables the accurate compensation of drift even with a long observation. We model the drift in the STFT domain as a linear phase by ignoring the drift inside each short time frame. By assuming that the sources are unmoving and have stationary amplitudes, the sound wave to be observed is regarded as stationary regardless of the number of sources. Since the stationarity collapses with the pseudo-movement of the sources caused by drift, the stationarity can be a cue to estimate the drift. Thus, we derive a likelihood function in the STFT domain to measure the stationarity and evaluate the compensation of the sampling frequency mismatch. The maximum likelihood estimate is searched for efficiently by performing a golden section search. We also show that the likelihood function evaluates the coherence between the channels. To compensate the recording start offset, we shift the observed signal in the time domain to maximize the interchannel correlation of the signals with the sampling frequency mismatch compensated. Since the accuracies of the estimation of the offset and the compensations of the drift are mutually dependent, particularly when the observation is long, we iterate these procedures. We evaluate the effectiveness of the proposed method of blind drift compensation in an experiment to emulate the asynchronous recording of an ad hoc microphone array by giving an artificial sampling frequency mismatch.

The rest of the paper is organized as follows. In Section 2, we formulate the asynchronous observation of the ad hoc microphone array. In Section 3, we describe our modeling of the drift in the STFT domain. In Section 4, we derive the likelihood function used to estimate the

Download English Version:

https://daneshyari.com/en/article/6959618

Download Persian Version:

https://daneshyari.com/article/6959618

Daneshyari.com