Contents lists available at ScienceDirect
Digital Signal Processing
www.elsevier.com/locate/dsp

Acoustic direction finding using single acoustic vector sensor under high reverberation

Metin Aktas [a,*,1], Huseyin Ozkan [b]

[a] Transportation, Security, Energy and Automation Systems Business Sector (UGES) at ASELSAN Inc., Ankara, Turkey
[b] Faculty of Engineering and Natural Sciences, Sabancı University, Istanbul 34956 Turkey

A R T I C L E I N F O

Article history:
Available online xxxx

Keywords:
Acoustic vector sensor
Direction of arrival estimation
Source localization
Under-determined
High reverberation
Time-frequency analysis

A B S T R A C T

We propose a novel and robust method for acoustic direction finding, which is solely based on acoustic pressure and pressure gradient measurements from single Acoustic Vector Sensor (AVS). We do not make any stochastic and sparseness assumptions regarding the signal source and the environmental characteristics. Hence, our method can be applied to a wide range of wideband acoustic signals including the speech and noise-like signals in various environments. Our method identifies the “clean” time frequency bins that are not distorted by multipath signals and noise, and estimates the 2D-DOA angles at only those identified bins. Moreover, the identification of the clean bins and the corresponding DOA estimation are performed jointly in one framework in a computationally highly efficient manner. We mathematically and experimentally show that the false detection rate of the proposed method is zero, i.e., none of the time-frequency bins with multiple sources are wrongly labeled as single-source, when the source directions do not coincide. Therefore, our method is significantly more reliable and robust compared to the competing state-of-the-art methods that perform the time-frequency bin selection and the DOA estimation separately. The proposed method, for performed simulations, estimates the source direction with high accuracy (less than 1 degree error) even under significantly high reverberation conditions.

1. Introduction

Direction of arrival (DOA) estimation [1,2] for an acoustic source is an important research topic in digital signal processing applications such as source separation [3,4], audio surveillance [5,6] and video conferencing [7]. The major challenge in these applications is the acoustic signal distortions due to the reflection, diffraction, and scattering by the objects in the transmission medium [8]. As a result of such distortions, the transmitted signal is often received via multiple paths. The total number of these paths and the signal power at each path are directly related to the positions and the material properties of the objects in the medium, which determine the reverberation strength, i.e., the persistence of the sound after it is produced [9]. Hypothetically, each path is considered to generate a virtual source that is highly correlated with the actual source [10]. Each virtual source conceals the actual source and hence makes the DOA estimation complicated. Moreover, as the reverberation strength increases, the total number of signals including the real and virtual sources is more likely to exceed the number of sensors. Thus, the source direction estimation becomes an under-determined problem due to the degenerate undesired solutions and therefore, it is generally addressed with strong assumptions about the signal source and the environmental conditions [11–18]. However, in this paper, we study the DOA estimation problem under high reverberation, i.e., in the presence of many degenerate solutions, without any assumptions on the stochastic properties of the source signal and the environmental conditions. Our technique performs a time-frequency decomposition such that the signal portions that are not distorted with multipath signals and noise are identified. In this process, we exploit the characteristics (that is independent with the frequency content of the signals) of the directional amplitude response of an Acoustic Vector Sensor (AVS) to guarantee that the multipath and noise distortions are completely eliminated. Then, the solution corresponding to the actual source is successfully separated. Therefore, our technique produces reliable solutions for the DOA estimation even under high reverberation. In this sense, our approach is highly novel and robust to environmental conditions.

When the total number of sources (including the virtual ones) exceeds the number of acoustic sensors, i.e., when the DOA es-

* Corresponding author.
E-mail addresses: maktas@aselsan.com.tr (M. Aktas), hozkan@sabanciuniv.edu (H. Ozkan).
[1] Member of EURASIP.

https://doi.org/10.1016/j.dsp.2018.01.001
1051-2004/

timation is an under-determined problem, the DOA estimation performance of the approaches [11–13], which only consider over-determined systems, seriously degrades [14,15]. On the other hand, several methods [16–18] are proposed to specifically handle the under-determined DOA estimation with certain assumptions about the signal and environmental characteristics such as uncorrelated and limited number of source signals [16], [5] and known or accurately estimated room impulse response [17,18]. We note that these methods [16–18] have restricted conditions that hardly hold in challenging real communication environments, cf. [19].

To mitigate these restricted conditions, other techniques [20–22] exploit the sparsity of the non-stationary signals in the time-frequency domain. However, these studies heavily depend on W-disjoint orthogonality of the source signals [20]. Since the assumption of the W-disjoint orthogonality of the received signals corresponds to assuming only one active dominant source at each time-frequency bin, these methods estimate the DOA angles at each time-frequency bin separately. In this case, the under-determined DOA estimation problem is handled as multiple overdetermined DOA estimation problems and hence, more sources can be localized with less number of sensors. Nevertheless, the DOA estimation performance of the methods using W-disjoint orthogonality are significantly sensitive to the several factors such as noise and reverberation [14,23]. As the noise power and reverberation increases, the source sparseness in the time-frequency domain decreases [23]. In that case, at some time-frequency bins, multiple sources (which is undesired) become dominant and the accuracy of DOA estimates at those bins decreases due to the invalidness of W-disjoint orthogonality. The work in [14] shows with the simulation that increasing reverberation strength increases the variance of the DOA estimate spread around the real DOA angles.

Instead of estimating the DOA at each time-frequency bin by relying on W-disjoint orthogonality, it is proposed [24–28] to select the time-frequency bins with only one single active source and estimate the DOAs at only those bins, which turns the under-determined problem into a bin-specific over-determined problem. The most critical part common to these methods is then to correctly determine those specific bins. To this end, in [24] and [25], only the most powerful source is assumed to be active at each time-frequency bin and thus the DOA estimations is performed at the bins bearing high local power. In [26], subspace based time-frequency selection procedure is proposed with the assumption that signals and noise are statistically independent, which is—on the contrary—clearly violated in most of the realistic scenarios, where the high degree of reverberation creates coherent, i.e., highly correlated, sources. A three-step time-frequency bin selection method is proposed in [27] by jointly using the methods in [24] and [26]. The time-frequency bins with sufficiently high local power are selected and then, "onset detection" is performed by marking a sudden rise in energy in the frequency bands. In the last step, the coherence test of [26] is applied to the time-frequency bins selected in the onset detection step. Although selecting the time-frequency bins with higher local power [24] is an effective method for eliminating the noise, it leads to incorrect decisions under high reverberation. The reason is that the time-frequency bins with power above a certain threshold may be resulted from not only a single source but from simultaneously active sources at those bins, for example, a typical failure scenario can be realized due to the overlapped portions of the direct path and multipath signals. Similarly, detecting a sudden rise in energy in time-frequency bins [27] is not a sufficient condition to conclude that only the new sound source is active at those bins. When the energy of the direct path signal does not degrade so rapidly as expected, the signals from other paths may overlap with the direct path signal. In another study [28], a smoothing operation over frequency bins is proposed to overcome the false detections for coherent sources (to enhance the method in [26]) by using spherical array with 32 sensors. The basic assumption for this method is that the array steering matrix is frequency independent and the coherent source signals are not linearly dependent over the frequency. However, this assumption becomes invalid when the same signal from different reflection paths arrive approximately at the same time.

On the contrary, we propose a novel time-frequency selection method for a single source without any assumption about the signal source statistics, power distribution and environmental characteristics. Our approach performs the DOA estimation and the time-frequency bin selection jointly in one framework, which is first introduced in our previous paper [37] for perfect collocated sensors [33–37]. This framework consists of single source detection and clustering the detected sources for direct path identification. Based on this approach, we first consider that there is only one single active source in the medium and estimate the DOA at each time-frequency bin. Then, we model the amplitude array response of the AVS from the estimated DOA and select the time-frequency bins, where the observed signal is consistent with the modeled array response. We emphasize that this current paper describes a more thorough analysis of our approach and introduces an additional new methods for addressing the realistic scenarios in which the perfect collocation of sensors is not possible.

Due to the physical constraints, the microphones on an AVS cannot be placed at the same point in space in real scenario. Moreover, placing microphones very close to each other may generate reflected waves distorting the received signals. Based on these practical considerations, in this work we handle the realistic scenario of microphones placed on the AVS with offset in 3D space, in contrast to the studies that assume perfect collocation of microphones [33–37]. Note that, the DOA estimation algorithms derived for perfect collocated microphone systems can not be used effectively in practical cases, since the offset between microphones invalidates the assumption that there are only amplitude differences between microphone signals. To handle this situation, in this paper, we propose novel DOA estimation and single source identification methods to able to use the framework introduced in our previous work [37] in realistic AVS structures. In our proposed method, we do not need to know the exact offset between microphones to find the direction of arrival (DOA) estimation; instead we use the resulting time shifts which we mathematically as well as experimentally analyze. These time shifts introduce the spatial phase factors in the array response of the AVS and give rise to ambiguity in the DOA estimation. To this end, the proposed method handles this ambiguity in direction of arrival estimation problem under high reverberation and practical AVS constraints by exploiting the directional amplitude structure of the AVS array response [29,30]. Through this approach, we guarantee that there is only one active source at the selected time-frequency bins unlike other approaches such as [24]. We mathematically prove and experimentally show that the proposed approach does not label any time-frequency bins with multiple sources wrongly as a single-source bin in any condition. Therefore, our method is more reliable compared to the existing methods that perform the time-frequency bin selection and DOA estimation separately [24–28], i.e., not jointly as proposed in this paper. Since the identified time-frequency bins are multipath free (source signal is received through the direct path only, and not from reflecting surfaces), we also do not need to know the environmental characteristics such as room impulse response to find the DOA estimates. Our proposed method does not rely on W-disjoint orthogonality assumption and can be applied to any wideband acoustic signal, not limited to speech signals. 2-D DOA angle estimation at each time-frequency bin is performed directly from data without an exhaustive search as in the MUSIC algorithm

Please cite this article in press as: M. Aktas, H. Ozkan, Acoustic direction finding using single acoustic vector sensor under high reverberation, Digit. Signal Process. (2018), https://doi.org/10.1016/j.dsp.2018.01.001