Blind synchronization in wireless acoustic sensor networks

Dani Cherkassky, *Student member, IEEE,* and Sharon Gannot, *Senior member, IEEE*

The challenge of blindly resynchronizing the data acquisition processes in a wireless acoustic sensor network (WASN) is addressed in this paper. The sampling rate offset (SRO) is precisely modeled as a time scaling. The applicability of a wideband correlation processor for estimating the SRO, even in a reverberant and multiple source environment, is presented. An explicit expression for the ambiguity function, which in our case involves time scaling of the received signals, is derived by applying truncated band-limited interpolation. We then propose the recursive band-limited interpolation (RBI) algorithm for recursive SRO estimation. A complete resynchronization scheme utilizing the RBI algorithm, in parallel with the SRO compensation module, is presented. The resulting resynchronization method operates in the time domain in a sequential manner and is thus capable of tracking a potentially time-varying SRO. We compared the performance of the proposed RBI algorithm to other available methods in a simulation study. The importance of resynchronization in a beamforming application is demonstrated by both a simulation study and experiments with a real WASN. Finally, we present an experimental study evaluating the expected SRO level between typical data acquisition devices.

***Index Terms*—Distributed microphone array, Wireless acoustic sensor network, Blind synchronization, Sampling rate offset, Wideband correlation processing**

I. Introduction

MICROPHONE arrays are used in a wide range of acoustical signal processing applications, e.g., localization, speech enhancement, and blind source separation [1],[2],[3],[4],[5]. A traditional microphone array consists of several microphones connected to a data acquisition center equipped with a processing unit. Such a configuration supports coherent acquisition of the signals and facilitates the application of optimal array processing methods [6],[7] to extract the spatial information. In recent years, the concept of wireless acoustic sensor networks (WASNs) has attracted the attention of the speech processing community [8]. Together with the appealing advantages offered by WASNs, some new challenges arise. One of the challenges is the lack of synchronization between the WASN nodes, which results in incoherent acquisition of the acoustical scene and thus prevents the straightforward application of the optimal array processing methods.

In contrast to a microphone array, where all signals are sampled with the same clock, the sampling process in each WASN node relies on its local clock oscillator, rendering sampling rate offsets (SROs) inevitable. The general challenge of synchronizing the data acquisition processes in a WASN, in particular in communication applications, has been addressed in the literature [9], [10], [11]. The vast majority of traditional synchronization methods are founded on an exchange of time stamp information between the network nodes [12], [13].

In the field of audio signal processing, the task of resynchronizing WASN signals can be accomplished by either processing the audio signals or distributing a synchronization signal in the network. An early work by Wehr et al. [14] addressed the synchronization problem in a distributed beamforming for blind source separation. They proposed an algorithm for estimating the SRO based on a modulated reference signal that is broadcast in the WASN. In [15], the authors proposed an SRO estimation methodology based on the communication layer, in which time stamps between the network nodes are broadcast. The time stamps are utilized by a Kalman filter to adjust and control the sampling rate in each node.

The methods aimed at resynchronizing a WASN solely by utilizing the audio signals are frequently referred to in the literature as *blind* synchronization methods. Blind synchronization involves two processing stages: SRO estimation and SRO compensation. In general, blind synchronization dictates an exchange of audio signals between the independent nodes, which prima facie seems unfavorable in terms of the resulting communication bandwidth. However, a broadcast of the audio signals in the network is typically required from the application perspective, e.g., a distributed beamforming [8]. Accordingly, blind WASN synchronization does not necessarily impose an excessive communication bandwidth. In this paper, we address the challenge of blindly resynchronizing a WASN.

An early work by Pawig et al. [16] considered blind SRO estimation between input and output channels in a single channel echo cancellation system. They proposed a combined time-recursive algorithm for tracking the room impulse response and the SRO simultaneously. In this method, the SRO compensation is achieved by applying a non-integer sampling rate conversion method [17]. In general, the SRO can be easily compensated once the offset is determined by resampling the discrete-time signals. Lagrange polynomial interpolation [18] and band-limited interpolation [19] are well-known and commonly used techniques for resampling discrete-time data streams.

In recent years, several studies addressed blind SRO estimation. In [20], the authors proposed approximating the SRO phenomenon by using linear phase drift between the signals in time. In this method, the SRO is estimated assuming a spatially coherent and statistically stationary environment. The linear phase drift approximation was also considered in [21], where the authors proposed a maximum likelihood SRO estimator achieved by modeling the microphone signals in the short-time Fourier transform (STFT) domain with a multi-variant

D. Cherkassky and S. Gannot are with the Faculty of Engineering, Bar-Ilan University, Ramat-Gan, Israel (e-mail: dani.cherkassky@gmail.com; sharon.gannot@biu.ac.il).

Gaussian distribution. The linear phase drift approximation was also utilized in [22],[23] and [24] for the same purpose. Recently, Wang and Doclo utilized a similar approximation in [25] and proved that the correlation coefficient between two signals tends to present the highest value when their sampling is synchronized. Accordingly, they proposed a two-stage exhaustive search for maximizing the correlation coefficient between two asynchronous signals and thus deduced their SRO.

In [26], we proposed an alternative approach for modeling the SRO and blindly estimating it. The SRO was precisely modeled as a time scaling and its equivalence to the Doppler effect was shown. We also demonstrated that the wideband correlation processor [27] is an applicable SRO estimator, even in a reverberant and multiple speaker scenario. In this method, the maximization of the wideband correlation between two asynchronous signals is implemented by utilizing an exhaustive search. At each grid point, one of the signals is time-scaled and translated. The effectiveness of the proposed method was validated in a simulation study, as well as in a real-life application. However, the computation load associated with the aforementioned procedure makes it impractical for a typical real-time application.

In the method described in the current paper, we adopt the time-scaling model and utilize the wideband correlation preprocessor for estimating the SROs. First, we derive an explicit expression of the *ambiguity function*, which in our case, involves time scaling of the received signals. The time scaling of the signals is implemented by applying truncated band-limited interpolation [28]. We propose an iterative batch algorithm for maximizing the ambiguity function, based on the well-known Newton-Raphson root finding method. Additionally, we utilize stochastic approximation in order to implement the batch iterative algorithm in a recursive fashion. The main contribution of this work is a complete resynchronization scheme utilizing the recursive band-limited interpolation (RBI) algorithm, in parallel with the SRO compensation module. The resulting resynchronization method operates in the time domain in a sequential manner and is thus capable of tracking a potentially time-varying SRO.

The paper is organized as follows. Section II is dedicated to introducing the preliminaries of blind WASN synchronization. The time-scaling model of the SRO is presented. The performance of the wideband correlation preprocessor for estimating the SRO and the delay in sampling start (DISS) is analyzed. In Section III, the proposed RBI framework for SRO estimation and resynchronization is presented. An empirical and simulative study is described in Section IV. The first part of the study is aimed at evaluating the performance of the RBI synchronization method in comparison with existing algorithms. In the second part, we address the distributed beamforming task and demonstrate the importance of synchronization in this application, by both simulation and a laboratory experiment. The capability of the proposed recursive algorithm to track a time-varying SRO is also demonstrated by simulation. Section V concludes with a brief summary. In Appendix A, we present an empirical study examining the expected SRO level between typical asynchronous data acquisition devices.

II. Preliminaries

The following section is divided into two parts. In the first part, we present the time-scaling model of the SRO. In the second part, the wideband correlation processor and its applicability to the SRO estimation in a reverberant and multi speaker scenario is addressed. The model and the estimator were already presented in [26]. Nevertheless, since the model and the estimator are the core of the proposed RBI algorithm, hereafter we present the main results.

A. Sampling rate offset model

Consider N_{s} coherent speech sources $\{s_n(t)\}_{n=1}^{N_{\mathrm{s}}}$ propagating in a reverberant environment and captured by an array of microphones. The microphones' signals are further corrupted by a low-level, spatially uncorrelated sensor noise $v(t)$. Accordingly, the continuous-time signal at the output of the mth microphone is given by

$$x_m(t) = \sum_{n=1}^{N_{\mathrm{s}}} h_m^n(t) * s_n(t) + v_m(t), \tag{1}$$

where t is the continuous-time index, $*$ represents the linear convolution operator, and $h_m^n(t)$ is the acoustic impulse response relating $s_n(t)$ and the mth microphone.

Since the sampling process in each WASN node relies on its local clock source, SROs are inevitable. For the sake of simplicity and without loss of generality, let us assume in the following that each node comprises a *single* microphone. The sampling rate of the mth microphone is defined in terms of the sampling rate of the reference microphone, f_s, as $f_s^m = f_s/a_m$. Without loss of generality, we set the first microphone to be the reference by setting $a_1 = 1$.

We demonstrated in [26] that the effect induced by the SRO is similar to the *Doppler effect*. Accordingly, we can formulate an equivalent problem, in which the sampling process in all nodes is synchronized, but the continuous-time signal at the mth microphone is time scaled by a_m and translated by d_m with respect to (w.r.t.) the first microphone, while d_m indicates the DISS between the mth microphone and the first microphone. Under this model, the equivalent continuous-time signal received by the mth microphone is given by

$$\widetilde{x}_m(t) = \sum_{n=1}^{N_{\mathrm{s}}} x_m^n\left(a_m\, t + d_m\right) + v_m(t), \tag{2}$$

where $a_1 = 1$, $d_1 = 0$ Sec and $x_m^n(t)$ denote the component of $s_n(t)$ at the mth microphone.

The performance of any data-dependent array processing algorithm depends on the coherence level between the array signals [6]. It is clear that the coherence between the signals $\{\widetilde{x}_m(t)\}_{m=1}^{M}$ drops when the difference between the scaling factors, $\{a_m\}_{m=2}^{M}$ and a_1 increases. In the following section, we present the wideband correlation processor for SRO (i.e., $\{a_m\}_{m=2}^{M}$) estimation. In order to analyze the properties of the estimator, we use the equivalent microphone signals (2). Based on this analysis, we derive a recursive SRO correction

$$\mathrm{WC}_{\widetilde{x}_1\widetilde{x}_m}(\alpha,\tau)=\sqrt{|\alpha|}\int_{-\infty}^{\infty}\widetilde{x}_m(t)\widetilde{x}_1\left(\alpha t-\tau\right)dt=\sqrt{|\alpha|}\left[W^1(\alpha,\tau)+W^2(\alpha,\tau)\right] \tag{3}$$

where:

$$W^1(\alpha,\tau)\triangleq\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[C_m^1(a_m,d_m-r_1,\tau+r_2,\alpha)h_m^1(a_m r_1)h_1^1(a_m r_2)]dr_1dr_2 \tag{4a}$$

$$W^2(\alpha,\tau)\triangleq\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[C_m^2(a_m,d_m-r_1,\tau+r_2,\alpha)h_m^2(a_m r_1)h_1^2(a_m r_2)]dr_1dr_2 \tag{4b}$$

$$C_m^1(a_m,d_m,\tau,\alpha)\triangleq\int_{-\infty}^{\infty}s_1\left(a_m t+d_m\right)s_1\left(\alpha t-\tau\right)dt \tag{4c}$$

$$C_m^2(a_m,d_m,\tau,\alpha)\triangleq\int_{-\infty}^{\infty}s_2\left(a_m t+d_m\right)s_2\left(\alpha t-\tau\right)dt. \tag{4d}$$

algorithm directly applied to the original measurements (1) in Section III.

B. Wideband correlation processor

Correlation processing is a very common technique in radar and sonar signal processing, aimed at identifying and localizing objects by means of time difference of arrival (TDOA) and scale difference of arival (SDOA) estimation [29]. In [26], we demonstrated that the wideband correlation processor can be utilized for SRO estimation in the problem at hand. Since the wideband correlation processor is the core of the algorithm proposed in Section III, we present the main result of [26] in the following.

For the sake of brevity and clearness, let us consider a scenario with $N_{\mathrm{s}}=2$. However, all the results are valid for more complex scenarios, with a larger number of coherent sources, as validated experimentally in Section IV. The only limitation that should be considered is the assumption that the sources are uncorrelated. Applying the wideband correlation processor to the *equivalent* signals, assuming that $s_1(t)$ and $s_2(t)$ are uncorrelated, the noise is spatially incoherent, and some straightforward algebra leads to the expression (3).

It is clear from (4c) and (4d) that C_m^1 and C_m^2 are the time-scaled auto-correlation functions of $s_1(t)$ and $s_2(t)$, respectively, and that they obtain their maximum values at $\tau=-d_m$ and $\alpha=a_m$. However, as $W^1(\alpha,\tau)$ and $W^2(\alpha,\tau)$ are related to C_m^1 and C_m^2, respectively, by a (double) linear convolution, a pronounced peak at $\tau=-d_m$ cannot be guaranteed. This is due to the TDOA axis being smeared by the convolution of C_m^1 and C_m^2 with the acoustical impulse responses, as evident from (4a) and (4b). On the other hand, since the SDOA axis is not smeared by the convolution, a pronounced extremum of $W^1(s,\lambda)$ and $W^2(s,\lambda)$, and hence also of their sum (3), can be expected in the SDOA axis.

It is therefore concluded that the maximum of the wideband correlation function w.r.t. the SDOA establishes a suitable SROs estimator, while its maximum w.r.t. the TDOA establishes a biased DISS estimator. Since $\mathrm{WC}_{\widetilde{x}_1\widetilde{x}_m}$ vanishes when both $s_1(t)$ and $s_2(t)$ are inactive, it is meaningful to apply the wideband correlation processor when at least one coherent source is active. Accordingly, a detection procedure for source inactivity is required for facilitating a robust estimation. The design of such a detection procedure is, however, beyond the scope of this study.

III. Synchronization Algorithm

Let us revert to the SRO estimation challenge and derive the RBI synchronization algorithm.

The principal idea of using the discrete-time version of the wideband correlation processor (3) for SRO estimation in a WASN was explored by the authors in [26]. Explicitly, we demonstrated that the SRO between the mth and the first microphone can be estimated by solving the following maximization problem:

$$\hat{a}_m=\underset{\alpha}{\operatorname{argmax}}\left\{Q(\alpha,\tau)\right\} \tag{5a}$$

$$Q(\alpha,\tau)=\sum_{n=1}^{N}x_m[n]x_1^{\alpha\tau}[n], \tag{5b}$$

where $x_m[n],\forall n=1,\ldots,N$ are the samples of $x_m(t)$ and $x_1^{\alpha\tau}[n]=x_1(\alpha t-\tau)_{|t=\frac{n}{f_s}},\forall n=1,\ldots,N$ are the samples of a time-scaled and time-translated continuous-time signal $x_1(\alpha t-\tau)$. In [26], an exhaustive grid search was applied for establishing the maximum of the ambiguity function Q. In the following, we derive the RBI algorithm for computing $x_1^{\alpha\tau}[n]$, optimizing the ambiguity function Q and resynchronizing the network.

A. Approximate ambiguity function

As demonstrated in Appendix A, the expected SROs in a WASN are in the order of tens of PPM, where PPM $=10^{-6}$. Accordingly, an optimization of Q by an exhaustive search has to be implemented over a very fine grid and thus requires an enormous amount of computation power, specifically in real-time applications. To circumvent this drawback, we propose in the following a recursive method for solving (5a) based on truncated band-limited interpolation [19].

Given a sequence of samples $x_1[l],-L\leq l<L$, we can reconstruct the corresponding continuous-time signal $x_1(t)$ by applying band-limited interpolation [28]:

$$x_1(t)=\sum_{l=-L}^{L}x_1[l]\,\mathrm{sinc}\left(\left(t-\frac{l}{f_s}\right)f_s\right), \tag{6}$$

where $\mathrm{sinc}(a) = \sin(\pi a)/(\pi a)$. The reconstructed continuous-time signal can be time-scaled by α and translated by τ:

$$x_1(\alpha t - \tau) = \sum_{l=-L}^{L} x_1[l] \,\mathrm{sinc}\left(\left(\alpha t - \tau - \frac{l}{f_s}\right) f_s\right). \quad (7)$$

The scaled continuous-time signal $x_1(\alpha t - \tau)$ can be sampled with the nominal sampling rate f_s:

$$x_1^{\alpha\tau}[n] = \sum_{l=-L}^{L} x_1[l] \,\mathrm{sinc}\left(\alpha n - \tau f_s - l\right). \quad (8)$$

It should be noted that in general, in the case where $\alpha > 1$, the procedure described in (6)-(8) results in a sampling frequency reduction. Accordingly, to avoid aliasing, a lowpass filter should be applied before sampling the scaled continuous-time signal. However, in our specific case, we consider modest SROs, of a fraction of Hertz at maximum, as presented in Table II. Thus, the lowpass filtering was not applied for the sake of simplicity.

We can now substitute the scaled discrete-time signal (8) into (5b) and write the explicit form of the ambiguity function Q in terms of the available samples only:

$$Q(\alpha, \tau) = \sum_{n=1}^{N} \sum_{l=-L}^{L} x_m[n] x_1[n] \,\mathrm{sinc}\left(\alpha n - \tau f_s - l\right). \quad (9)$$

Although perfect band-limited interpolation requires $L \to \infty$, practically $\mathrm{sinc}(a) \approx 0$ for $|a| \gg 1$. Thus, we can approximate Q by considering only the terms satisfying $|\alpha n - \tau f_s - l| \leq L_0$, where L_0 is a design parameter controlling the approximation accuracy:

$$Q(\alpha, \tau) \approx \sum_{n=1}^{N} \sum_{l=p-L_0}^{p+L_0} x_m[n] x_1[n] \,\mathrm{sinc}\left(\alpha n - \tau f_s - l\right), \quad (10)$$

where $p = \lfloor \alpha n - \tau f_s \rfloor$. The approximate ambiguity function (10) can be substituted into (5a), constituting an explicit optimization problem that has to be solved for estimating the SRO of the mth microphone w.r.t. the first microphone. We present the solution of the optimization problem in the following section.

B. Ambiguity function optimization

Finding an analytical solution for $\boldsymbol{\theta} = [\alpha, \tau]^T$ that maximizes the ambiguity function Q is a cumbersome task. However, we are able to derive analytic expressions for the first- and second-order derivatives of Q w.r.t. $\boldsymbol{\theta}$. Thus, we can optimize Q by applying a gradient descent search method. Let us rewrite (10) using $\boldsymbol{\theta}$:

$$Q(\boldsymbol{\theta}) = \sum_{n=1}^{N} \sum_{l=p-L_0}^{p+L_0} x_m[n] x_1[n] \,\mathrm{sinc}\left(\boldsymbol{\theta}^T \mathbf{r}[n] - l\right), \quad (11)$$

where $\mathbf{r}[n] = [n, -f_s]^T$. Applying the Newton-Raphson iterative optimization method [30] yields the following iteration rule:

$$\hat{\boldsymbol{\theta}}^{i+1} = \hat{\boldsymbol{\theta}}^{i} + \left[\boldsymbol{H}^i\right]^{-1} \boldsymbol{\nabla}^i, \quad (12)$$

where i is the iteration index, $\boldsymbol{H}$ is the Hessian matrix, and ∇ is the gradient vector. The gradient and the Hessian of the ambiguity function Q can be straightforwardly calculated:

$$\boldsymbol{\nabla}^i \triangleq \frac{\partial Q(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{n=1}^{N} \sum_{l=p-L_0}^{p+L_0} d'[n,l] x_m[n] x_1[l] \mathbf{r}[n], \quad (13)$$

$$\boldsymbol{H}^i \triangleq \frac{\partial^2 Q(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} = \sum_{n=1}^{N} \sum_{l=p-L_0}^{p+L_0} d''[n,l] x_m[n] x_1[l] \mathbf{r}[n] \mathbf{r}^T[n], \quad (14)$$

$$d'[n,l] = \begin{cases} 0, & \text{if } \varphi = 0 \\ \frac{\pi^2 \varphi \cos(\pi\varphi) - \pi \sin(\pi\varphi)}{(\pi\varphi)^2}, & \text{otherwise} \end{cases}, \quad (15)$$

$$d''[n,l] = \begin{cases} -\frac{1}{3}\pi^2, & \text{if } \varphi = 0 \\ \frac{(2 - \pi^2\varphi^2)\sin(\pi\varphi) - 2\pi\varphi\cos(\pi\varphi)}{\pi\varphi^3}, & \text{otherwise} \end{cases} \quad (16)$$

where $d'[n,l]$ and $d''[n,l]$ are the first- and second-order derivatives of the sinc function, respectively, and $\varphi = \boldsymbol{\theta}^T \mathbf{r} - l$ is the argument of the sinc function in (11).

The iterative-batch SRO estimation algorithm, as defined in (12), requires the use of an analysis window with a length equal to N, over which the SRO between the signals is assumed to be constant. This assumption does not hold in practice, as the independent clock sources in the nodes typically fluctuate over time. To circumvent the piecewise constant SRO assumption, we suggest a sequential SRO estimation algorithm. To obtain a sequential algorithm, we utilize the *stochastic approximation* methodology, with the iteration index is substituted with the time index:

$$\hat{\boldsymbol{\theta}}[n] = \hat{\boldsymbol{\theta}}[n-1] + \hat{\boldsymbol{H}}^{-1}[n] \cdot \hat{\boldsymbol{\nabla}}[n], \quad (17)$$

where $\hat{\boldsymbol{H}}[n]$ and $\hat{\boldsymbol{\nabla}}[n]$ are the estimates of the Hessian matrix and the gradient vector, accordingly. For calculating $\hat{\boldsymbol{H}}[n]$ and $\hat{\boldsymbol{\nabla}}[n]$, we incorporate a forgetting factor $0 \leq \lambda \leq 1$ and apply the well-known recursive exponential smoothing procedure:

$$\begin{aligned} \hat{\boldsymbol{\nabla}}[n] &= \sum_{n=1}^{N} \lambda^{n-i} \sum_{l=p-L_0}^{p+L_0} d'[n,l] x_m[n] x_1[l] \mathbf{r}[n] \\ &= \lambda \hat{\boldsymbol{\nabla}}[n-1] + \mathbf{r}[n] x_m[n] \sum_{l=p-L_0}^{p+L_0} d'[n,l] x_1[l], \end{aligned} \quad (18)$$

$$\begin{aligned} \hat{\boldsymbol{H}}[n] &= \sum_{i=1}^{N} \lambda^{n-i} \sum_{l=p-L_0}^{p+L_0} d''[n,l] x_m[n] x_1[l] \mathbf{r}[n] \mathbf{r}^T[n] \\ &= \lambda \hat{\boldsymbol{H}}[n-1] + x_m[n] \mathbf{r}[n] \mathbf{r}^T[n] \sum_{l=p-L_0}^{p+L_0} d''[n,l] x_1[l]. \end{aligned} \quad (19)$$

The framework defined by (17)-(19) constitutes a recursive method for estimating a (potentially) time-varying SRO.

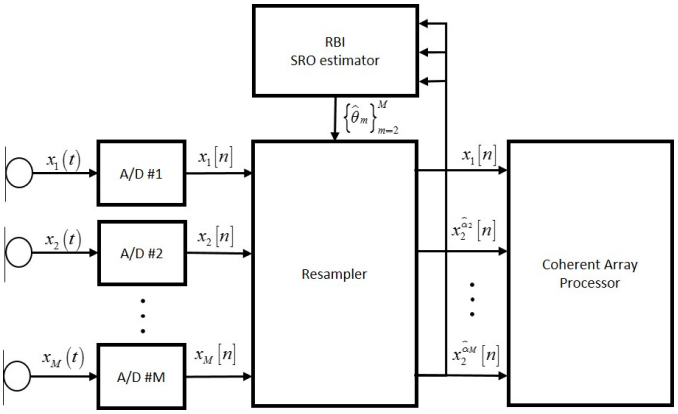

Fig. 1: Recursive band-limited interpolation synchronization algorithm

C. Practical considerations

The complete synchronization algorithm is presented in Fig. 1. It comprises two major building blocks: the SROs estimator and resampler. Several practical aspects related to the implementation of the proposed scheme should be considered.

1) Delay in sampling start estimation

As demonstrated in Sec. II-B, the TDOAs at the output of the proposed algorithm $\{\hat{\tau}_m\}_{m=2}^{M}$ are biased estimates of the DISS between the nodes $\{d_m\}_{m=2}^{M}$. This is due to the TDOA axis being smeared by the convolution of C_m^d and C_m^i with the acoustical impulse responses, as evident from (4c) and (4d). Accordingly, $\{\hat{\tau}_m\}_{m=2}^{M}$ cannot be used for accurate compensation of the DISS. However, under the plausible assumption that $|\hat{\tau}_m - d_m|, m = 2, \ldots, M$ are smaller than the analysis window of the array processor, the bias effect can be absorbed into the estimation of the acoustic impulse responses $h_m^d(t), h_m^i(t)$ [1].

2) Recursive band-limited interpolation initialization

Since in the general case the objective function $Q(\boldsymbol{\theta})$ is not unimodal, the search method (17) could become trapped in a local maximum, depending on the initial conditions. It is therefore important to initialize the search algorithm in close proximity to the global maximum. As reflected in Table II, the typical values of a_m are very close to 1.0, and thus initializing the search with $\{\alpha_m[0] = 1.0\}_{m=1}^{M}$ should be sufficient for avoiding local maximum on the SDOA axis, in most cases. It should be stressed that the last argumentation can by no means serve as a proof that initializing $\{\alpha_m[0] = 1.0\}_{m=1}^{M}$ will guarantee avoiding local maximum. A further study of the variation of the objective function w.r.t. the $\{\alpha_m\}_{m=1}^{M}$ is required in order to suggest an optimal initialization technique. However, this is beyond the scope of the current study. The initializing of $\{\tau_m[0]\}_{m=1}^{M}$ close to the true DISS seems to be an even more difficult task. In the case where $\{d_m\}_{m=2}^{M}$ are small, their coarse estimates can be obtained from cross-correlating $x_1[n]$ and $x_m[n]$. Let us define L_1 as a number of samples used for estimating the cross-correlation. L_1 is one of the design parameters of the proposed RBI algorithm. However, since the nodes are fully independent, the delay in the sampling start is arbitrary and unbounded. Thus, in the most general case, when no prior information about the DISS between the nodes is available, the signals have to be coarsely aligned using side information that might be available in the network, for example by exchanging time stamps or by broadcasting an acoustic signature upon start of work. In general, the coarse alignment mechanism should compensate the DISS such that the residual delay is much smaller than L_1.

3) Resynchronization

In order to resynchronize the network, we have to utilize the estimators $\{\hat{\alpha}[n], \hat{\tau}[n]\}_{m=2}^{M}$ for compensating their effects. In our proposed scheme shown in Fig. 1, the resampling module is independent of the estimation module. Thus, principally, any time-domain resampling technique is applicable. For example, considering WASN resynchronization, in [25] and [20] it was proposed that a fourth-order Lagrange interpolation be used; in [26], we utilized splines interpolation. However, in the scope of this work, it is natural to utilize band-limited interpolation for resynchronizing the network. Given the estimators $\{\hat{\alpha}[n], \hat{\tau}[n]\}_{m=2}^{M}$, the signal at the m$th$ node is resampled by

$$x_m^{\hat{\alpha}\hat{\tau}}[n] = \sum_{l=-L_2}^{L_2} x_m[l]\,\mathrm{sinc}\left(\frac{1}{\hat{\alpha}}n + \hat{\tau} f_s - l\right), \qquad (20)$$

where L_2 is a design parameter controlling the tradeoff of the accuracy and complexity of the resampling.

D. Computational complexity

Let us consider the computational complexity of the proposed RBI algorithm. The algorithm contains two modules, SRO estimation and SRO compensation (resynchronization). In order to estimate the SRO of the mth microphone (17)-(19) have to be solved each time a new sample is acquired. The resynchronization of the mth microphone is achieved by computing (20) each time a new sample is acquired.

The precise computational power required for implementing the RBI algorithm strongly depends on the manner in which the $\sin(\cdot), \cos(\cdot)$ and $\mathrm{sinc}(\cdot)$, functions are computed. However, for our needs, let as assume that all the related trigonometric functions are stored as tables and their value can be evaluated in $O(1)$.

The resynchronization (20) requires $O(2L_2)$ multiplications per sample. The implementation of (17) requires an inversion of a 2×2 matrix, multiplication of a 2×2 matrix with 2×1 vector, and addition of two 2×1 vectors; the resulting computational load is $O(1)$. An implementation of (18) and (19) requires $O(2L_0)$ multiplications each. Accordingly, the computational power required for implementing the proposed resynchronization scheme is $O(4L_0 + 2L_2)$ per sample.

IV. Experimental Study

In the following sections we compare the performance of the proposed RBI method to some state of the art alternative algorithms and demonstrate the necessity of resynchronizing the WASN prior to applying the beamforming algorithms.

A. Setup and definitions

The proposed algorithm was tested both in a simulated and a real $6\times6\times2.4$ m room with $T_{60}\approx200$ msec. A uniform linear array comprising M microphones with 2 cm inter-spacing was positioned in the center of the room. Three acoustical sources were positioned at a distance of 2 m from the array, with an angle of arrival (AOA) equal to $60°, 90°, 120°$ w.r.t. the array axes. Namely, two directional speech signals $s_{\rm i}(t)$, and $s_{\rm d}(t)$ impinged the array with an AOA equal to $60°, 90°$, respectively, and a stationary fan noise $v(t)$ impinged the array with an AOA equal to $120°$. The powers of the sources are defined as $\sigma_{\rm i}^2, \sigma_{\rm d}^2$, and $\sigma_{\rm v}^2$, respectively. In the simulation experiments, the impulse responses between the sources and the microphones were simulated using [31] and the microphone signal was further corrupted by a spatially uncorrelated sensor noise with a power equal to $\sigma_{\rm n}^2$. In the following, we utilize the following definitions to characterize the setups and the results.

$$\text{SIR} \triangleq \frac{\sigma_{\rm d}^2}{\sigma_{\rm i}^2},\ \text{SNR} \triangleq \frac{\sigma_{\rm d}^2}{\sigma_{\rm v}^2},\ \gamma \triangleq \frac{\sigma_{\rm d}^2+\sigma_{\rm i}^2}{\sigma_{\rm v}^2},\ \rho \triangleq \frac{\sigma_{\rm d}^2+\sigma_{\rm i}^2+\sigma_{\rm v}^2}{\sigma_{\rm n}^2}. \tag{21}$$

In the following, the stationarity and the microphones' signals are manifested by the value of γ. When $\gamma\to\infty$, the microphones' signals are dominated by the speech components, which are by definition non-stationary, while when $\gamma\to0$ the microphones' signals are dominated by the stationary noise. The amount of spatial coherence between the microphones' signals is manifested by the value of ρ. The signals are dominated by the coherent sources when $\rho\to\infty$ and by the spatially uncorrelated noise when $\rho\to0$.

The linearly constrained minimum variance (LCMV) beamformer [32] was chosen as a test algorithm for the beamforming experiments. The constraint vector of the LCMV was set to impose a distortionless response toward $s_{\rm d}(t)$ and a null towards $s_{\rm i}(t)$. The response toward $v(t)$ was unconstrained. In the first part of the beamforming experiments, the sources were active in a non-overlapping pattern, while in the second part all the sources were concurrently active. The signals acquired during the first part of the experiments were utilized for estimating the covariance matrix of $v(t)$, as well as for estimating the relative transfer functions (RTFs) of $s_{\rm d}(t)$ and $s_{\rm i}(t)$. The RTFs were estimated by eigenvalue decomposition of the received signals' covariance matrix [33]. The LCMV beamformer was applied to the signals acquired in the second part of the experiments. We characterize the LCMV performance by three attributes:

$$\text{SIR Gain} \triangleq \frac{\text{SIR}_{\rm output}}{\text{SIR}_{\rm input}},\quad \text{SNR Gain} \triangleq \frac{\text{SNR}_{\rm output}}{\text{SNR}_{\rm input}},$$
$$\text{Distortion} \triangleq \sum_{n=1}^{N_{\rm T}} \left(d[n]-y[n]\right)^2, \tag{22}$$

where $d[n]$ and $y[n]$ are the desired and the actual outputs of the beamformer, respectively, and $N_{\rm T}$ is the length of the considered signals in samples. The subscripts *input* and *output* describe the parameter at the input and the output of the beamformer, respectively.

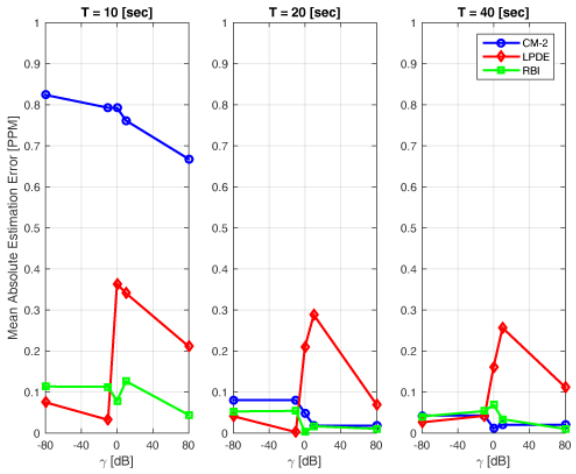

Fig. 2: Summary of sampling rate offset estimation performance. The true sampling rate offset is set to 5.27 PPM in this experiment.

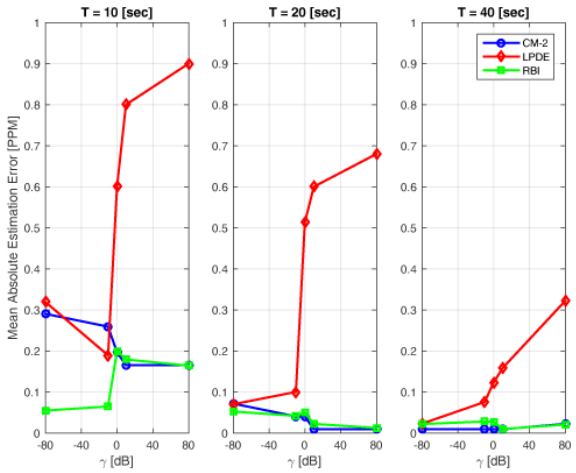

Fig. 3: Summary of sampling rate offset estimation performance. The true sampling rate offset is set to 30.74 PPM in this experiment.

The RBI design parameters were fixed to $L_0=10, L_1=1000, L_2=30, \lambda=0.999$ during the following experiments.

B. Sampling rate offset estimation benchmark

We turn now to a simulation study aimed at comparing the performance of the RBI algorithm proposed in Section III-B with that of the correlation maximization (CM-2) algorithm proposed in [25] and the linear-phase drift estimation algorithm (LPDE) proposed in [20]. To facilitate the comparison, we utilized the above specified simulative setup with $M=2$, $\sigma_i^2=0$ and set the nominal sampling rate to 16 kHz. The signal at the second microphone was resampled to introduce an artificial SRO. Since CM-2 and LPDE do not address DISS estimation, d_2 was set to zero.

In the first study, we evaluated the influence of signal length T and γ on the estimation performance of CM-2, LPDE, and RBI, while ρ was set to 40 dB. The expected SRO in real

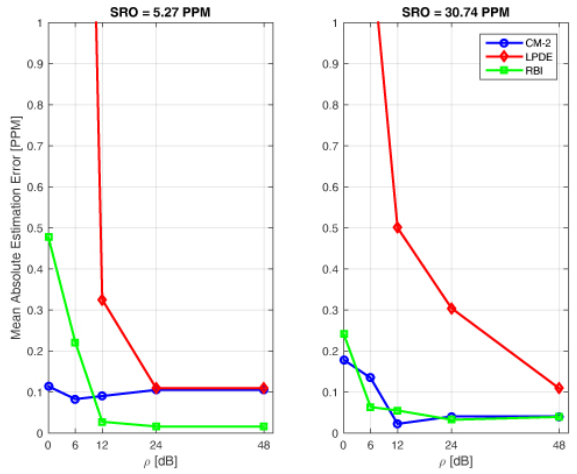

Fig. 4: Comparison of the robustness of the CM-2, linear-phase drift estimation, and recursive band-limited interpolation algorithms to an incoherent noise.

WASN is in the range of $0-50$ PPM, as shown in Appendix A. Accordingly, we evaluated the performance of the algorithms for SROs of this order of magnitude and present hereafter two representative test cases. In Fig. 2, the performance of the algorithms are depicted for a test case where the SRO was set to 5.27 PPM, while in Fig. 3 we present the performance for a test case where the SRO was set to 30.74 PPM. We implemented 100 independent realizations for both of the considered test cases; the average absolute SRO estimation error is presented in Fig. 2 and Fig. 3.

The LPDE algorithm considers SRO estimation in a stationary environment only. Accordingly, it demonstrates a relatively poor performance for $\gamma \geq 0$ dB. It should be noted, however, that LPDE results in a reasonable estimation error for $\gamma \leq -10$ dB. In contrast to LPDE, assuming that sufficiently long signals are available, e.g., T $\geq$ 20 Sec, the value of γ has only a minor effect on the CM-2 and RBI performance.

In terms of the required signal length, CM-2 seems to be the most sensitive to the available signal length, in particular when a small SRO is considered. On the other hand, RBI demonstrated a relatively stable estimation performance for the considered lengths of the signals.

In the second study, we evaluated the robustness of CM-2, LDPE, and RBI to the sensor noise. For this purpose, we set $T = 20$ sec, $\gamma = 10$ dB; otherwise, the setup was identical to that previously described. The average absolute estimation error over 100 independent realizations as a function of ρ is depicted in Fig. 4. As can be readily seen, CM-2 demonstrates the most robust performance against incoherent noise, while relatively poor robustness is demonstrated by LPDE. The robustness demonstrated by proposed RBI algorithm is comparable to that demonstrated by CM-2.

C. Beamforming in a simulated environment

In this simulation study, we explored the SRO's effect on the beamforming performance and tested the effectiveness of the proposed RBI algorithm in resynchronizing the WASN.

TABLE I: Introduced artificial sampling rate offset and the respective estimates using the recursive band-limited interpolation estimator

	True SRO [PPM]	Estimated SRO [PPM]
Mic 1	0	0
Mic 2	100	99.559
Mic 3	-100	-99.772
Mic 4	50	49.577
Mic 5	-50	-49.787
Mic 6	150	149

To facilitate this experiment, we utilized the setup described in Section IV-A with $M = 6$, nominal sampling frequency $f_s = 16$ kHz, SIR $= 0$ dB, SNR $= 8$ dB, and $\rho = 40$ dB. The network was desynchronized by introducing an artificial SROs w.r.t. f_s to each microphone signal individually, while $d_m = 0, \forall m = 1, ..., 6$. The artificial SROs were generated by resampling the microphone signals in the time domain.

The first microphone was selected as a reference, and the SROs at the remaining microphones were blindly estimated by the RBI (17) w.r.t. the first microphone. The introduced artificial SROs and the respective estimates are presented in Table I. It should be noted that the presented estimates were obtained only after the RBI algorithm had converged. The estimated SROs were utilized to resynchronize the network by applying (20) on each signal individually.

The LCMV performances in a synchronized, unsynchronized, and re-synchronized network are presented in Fig. 5. The behavior of the SIR gain, SNR gain and the distortion over time, averaged across a time window of 1.0 sec, is presented. As can be readily observed, the beamforming performance in an unsynchronized network is considerably lower than in a synchronized network. It is also seen that the performance in an unsynchronized network degrades over time, until a stationary state is reached after approximately 30 sec. By resynchronizing the network, we considerably improved the beamforming performance. However, because of the SRO estimation errors, which are presented in Table I, a perfect synchronization was not achieved. By carefully observing the graphs, one can notice that the gap between the solid and the dotted curves becomes wider with time. This variation in time demonstrates that even a modest SRO (or in this case SRO estimation error) reduces the beamforming performance, after sufficient time has elapsed.

D. Tracking a time-varying sampling rate offset

In the simulation study described in the following, we examined the ability of the proposed RBI algorithm to track a time-varying SRO. The setup in this study was identical to that utilized in the previous one, except for a single change: only two microphones were considered. The first microphone was selected as a reference, while a time-varying SRO was introduced to the second microphone by re-sampling its signal. The SRO of the second microphone w.r.t. the first one was then blindly estimated by applying the RBI algorithm (17). The introduced artificial SROs and the resulting estimates produced

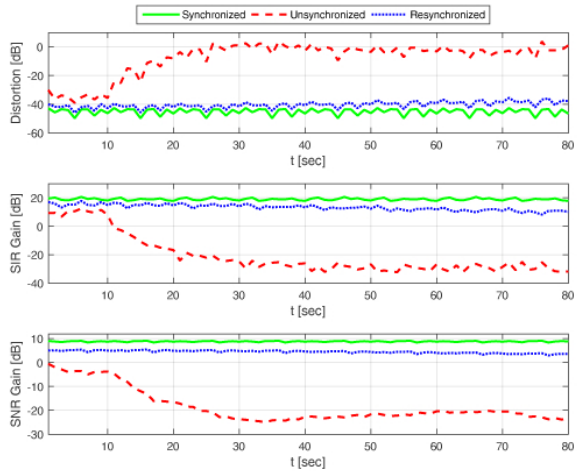

Fig. 5: Summary of linearly constrained minimum variance beamformer performance.

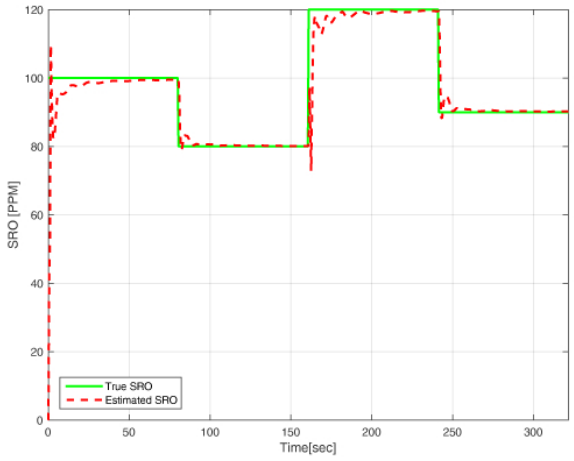

Fig. 6: Summary of linearly constrained minimum variance beamformer performance.

by the RBI are presented in Fig. 6. As demonstrated in this "toy example," the RBI continuously tracks the actual SRO between the microphones. In a real WASN, we might expect high frequency and small amplitude SRO variations because of the fast and small variations in the local clock oscillators. However, the analysis of the RBI tracking capabilities in a real WASN is beyond the scope of the current contribution.

E. Beamforming in a real environment

In this experiment, we evaluated the beamforming performance in a real WASN. To facilitate this experiment, the Acoustic Laboratory at Bar-Ilan University was arranged according to the setup described in Section IV-A with $M = 4$ and $T_{60} \approx 200$ msec. Each of the microphones was connected to an independent recording device. The following recording devices were tested: iPhone 5, Sony Ericsson Xperia Z, iPad Air 2, and Samsung Gallaxy S4 (the devices are similar to those analyzed in Appendix A). The nominal sampling frequency of all the devices was configured to 44.1 kHz, the

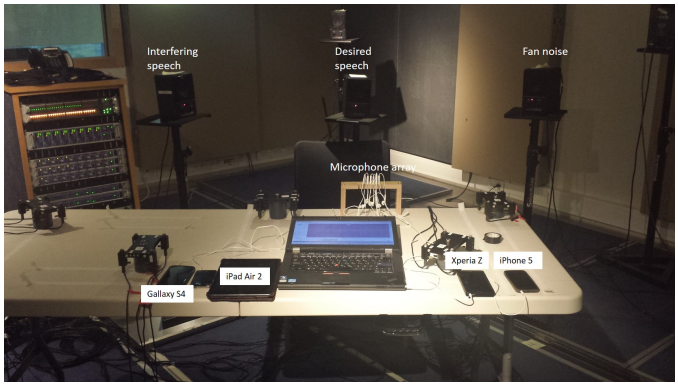

Fig. 7: Laboratory setup. Reverberation level in the room was adjusted to $T_{60} \approx 200$ msec.

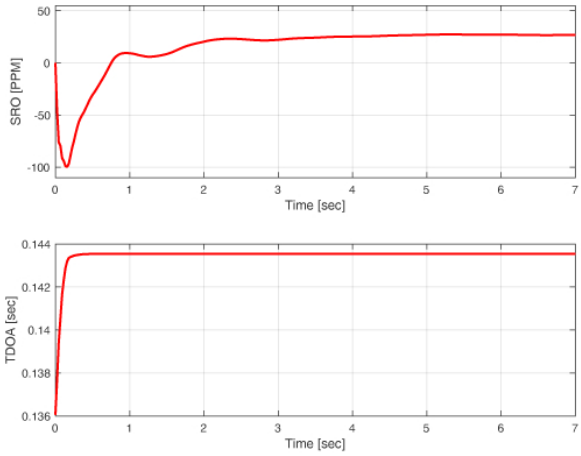

Fig. 8: Convergence processes of $\hat{\boldsymbol{\theta}}_4$. The upper panel presents the convergence of $\hat{\boldsymbol{\alpha}}_4$, and in the lower panel the convergence of $\hat{\boldsymbol{\tau}}_4$ is shown.

signal to interference ratio (SIR) was set to 0 dB, and the signal to noise ratio (SNR) was set to 10 dB. A picture of the laboratory setup is shown in Fig. 7.

The SRO and the TDOA were estimated by applying the RBI to the received signals. The first signal, acquired by the iPhone 5, was set to be the reference and $\{\boldsymbol{\theta}_m\}_{m=2}^{4}$ were estimated w.r.t. it. As an example, we depict the convergence processes for $\hat{\alpha}_4$ and $\hat{\tau}_4$ in Fig. 8. The SROs estimation procedure resulted in the following SROs estimates: $\hat{\alpha}_2 = -13.6, \hat{\alpha}_3 = 9.91, \hat{\alpha}_4 = 30.49$ PPM, upon convergence. These values can be compared to the SROs evaluated for the respective devices in Appendix A. As easily deduced, the SRO estimates here are different from those presented in Table II. The cause for this inconsistency is the fact that, in general, by optimizing (23) only the absolute value of the SRO can be recovered. Additionally, in our specific case, the accuracy of the optimization in Appendix A was set to 2.5 PPM.

Resynchronization of the network was achieved by applying (20) with $\boldsymbol{\theta}_2, \boldsymbol{\theta}_3$, and $\boldsymbol{\theta}_4$ to the second, third, and fourth signals, respectively. The LCMV performance in an asynchronous versus a re-synchronized network is shown in Fig. 9. As can be readily observed, the re-synchronized LCMV

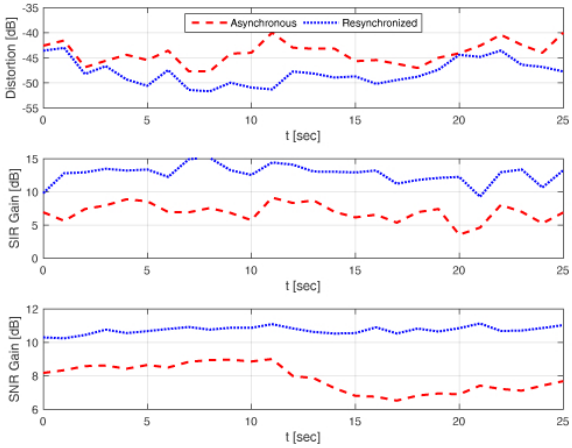

Fig. 9: Linearly constrained minimum variance beamformer in a real wireless acoustic sensor network and a re-synchronized wireless acoustic sensor network - performance summary.

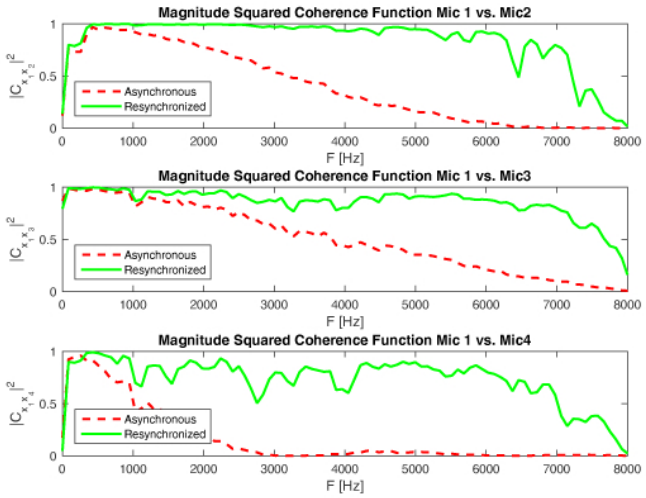

Fig. 10: Coherence level between the signals before and after resynchronization. Mic 1 is connected to iPhone, Mic 2 is connected to Sony Ericsson Xperia Z, Mic 2 is connected to iPad Air 2, and Mic 4 is connected to Samsung Gallaxy S4.

outperforms the asynchronous one in terms of all parameters. The superior performance of the re-synchronized LCMV can be understood by observing the coherence function between the microphones signals before and after resynchronization, as depicted in Fig. 10. As shown, the resynchronization dramatically increases the coherence level in a frequency range up to 6000 Hz, while above this frequency we still observe a drop in the coherence level. This may be attributed to a fast and small variation of the SROs between the signals that we failed to compensate. An additional reason for the drop in the coherence at a higher frequency band may be a lower level of $s_d(t)$, $s_i(t)$, and $v(t)$ on that frequency band.

V. Summary

The challenge of resynchronizing the data acquisition processes in a WASN was addressed. The SRO was precisely modeled as a time scaling and its equivalence to the Doppler effect was shown. The applicability of a wideband correlation processor for estimating the SRO, even in a reverberant and multiple source environment, was shown. An explicit expression of the ambiguity function, which in our case involves time scaling of the received signals, was derived by applying truncated band-limited interpolation. We then proposed the RBI algorithm for recursive SRO estimation.

The performance of the RBI algorithm was compared to that of alternative SRO estimation algorithms. The alternative SRO estimation algorithms operate in the STFT domain and assume a constant SRO during the observation time. The RBI algorithm estimates the SRO in the time domain in a sequential manner. This circumvents the assumption of a constant SRO during the observation period, and enables time-varying SRO tracking.

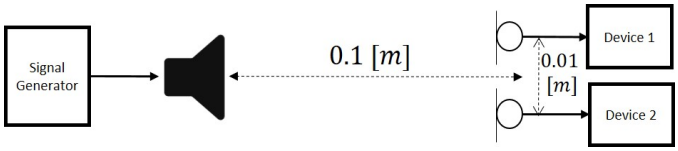

Fig. 11: Apparatuses for comparative sampling rate offset estimation between two devices.

Appendix A
Sampling Rate Offset Magnitude in Real Wireless Acoustic Sensor Networks

In this section, we describe the evaluation of the expected SRO magnitude in a practical WASN, when the sampling of the microphone signal in each node was implemented by using a typical data acquisition device, such as a laptop computer, smart-phone, or tablet.

To facilitate the evaluation, we established the following experimental setup, as depicted in Fig. 11. Two microphones were placed in close proximity to each other and 10 cm away from a simple PC loudspeaker. The setup was arranged on a desk in a normal office room. A narrow-band pilot signal was played through the loudspeaker and captured by the microphones. Each microphone was connected to an independent recording device. The devices were configured to a nominal sampling frequency of $f_s^n = 44.1$ kHz and the signal generator was set to generate a single tone with a frequency of $f_0 = 4$ KHz.

Let $x_1[n]$, and $x_2[n]$ be the discrete-time signals recorded by the first and the second device, respectively. We expect a frequency shift between $x_1[n]$ and $x_2[n]$ because of lack of synchronization between the devices. In order to estimate the frequency shift between the signals, we maximized the well-known, narrow-band cross ambiguity function (CAF), defined by [34]:

$$\mathrm{CAF}(\tau,k) = \sum_{n=1}^{N_\mathrm{T}} x_1[n] x_2^*[n+\tau] e^{-i2\pi \frac{kn}{N}}, \qquad (23)$$

where N_T is the length of the signals in samples. We were interested in $\hat{\tau}, \hat{k}$ such that

$$\hat{\tau}, \hat{k} = \underset{\tau,k}{\mathrm{argmax}} \left\{ CAF(\tau,k) \right\}. \qquad (24)$$

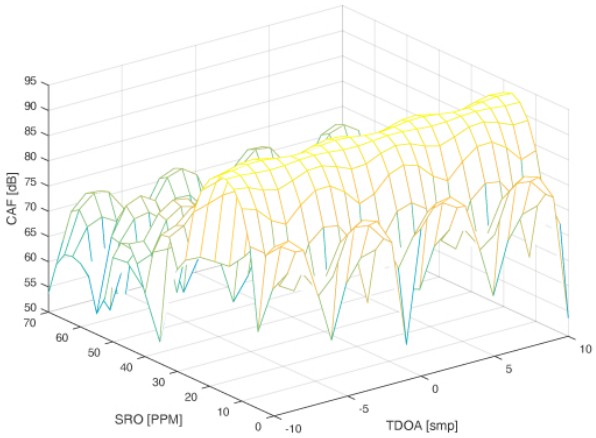

Fig. 12: Instance of the cross ambiguity function. The peak is obtained for a sampling rate offset of 17.5 PPM.

Let us define the frequency shift between the acquired signals as $f_{\text{dif}} \triangleq \hat{k} f_s^n / N$. It is well known that, in free-field applications, $\hat{\tau}$ and f_{dif} are the maximum likelihood estimators of the TDOA and the frequency shift between $x_1[n]$ and $x_2[n]$, respectively [35]. As demonstrated in Sec. II-B, $\hat{\tau}$ cannot serve as an estimator for the DISS between the first and the second device. Moreover, the selected pilot signal may result in an ambiguous TDOA estimator because of its periodicity. In our case, τ is a nuisance parameter, as we were not interested in estimating the DISS in this experiment. However, $\hat{k}$ can be considered an unbiased SRO estimate between the devices. In the scope of our study, the CAF was calculated for every pair of devices as suggested by (23), while (24) was solved by a straightforward exhaustive grid search. The resolution of the search along the k axis was set such that the respective resolution of f_{dif} was 0.01 Hz, which is equivalent to 2.5 PPM in this experiment, where PPM $= 10^{-6}$. The search across the τ axis was limited to a single period of the pilot signal, with a resolution of a single sample.

A single instance of the CAF is depicted in Fig. 12, while the resulting SROs for all the considered devices are presented in Table II. In the table, f_{dif} is estimated by using the above described procedure. The absolute value of the SRO between each pair of devices is calculated based on the respective f_{dif}, using the following relations: $|\text{SRO}| = |f_s^n - f_s^n(f_0 + f_{\text{dif}})/f_0|$ Hz, $|\text{SRO}| = 1 \times 10^6 \, |1 - f_0/(f_0 + f_{\text{dif}})|$ PPM, and the time-scaling factor $\alpha = f_0/(f_0 + f_{\text{dif}})$.

The results are presented in Table II. We deduce that the SRO in practical WASNs is indeed inevitable in most cases, while in a typical case one should expect an SRO of several tens of PPMs between each pair of independent devices. It is worth noting that an SRO of several tens of PPMs will significantly degrade the performance of any coherent array processing technique. For example, the effect of an SRO of 10 PPM on the minimum variance distortionless response (MVDR) beamformer was demonstrated in [24]. It was shown that in such scenarios the response of the beamformer is entirely smeared within less than 5 sec.

TABLE II: Estimated absolute value of the sampling rate offset between the considered devices

Devices	f_{dif} [Hz]	|SRO| [Hz]	|SRO| [PPM]	α
iPhone 5 - Galaxy S4	0.12	1.32	30	0.999970
iPhone 5 - iPhone 6	0	0	0	1
Lenovo T60 - Galaxy S4	0.03	0.33	7.5	0.999993
Lenovo T60 - Lenovo T420	0.05	0.55	12.5	0.999998
iPhone 5 - Lenovo T420	0.12	1.32	30	0.999970
iPhone 5 - LG Optimus	0.20	2.20	50	0.999950
iPhone 5 - Nexus 4	0.19	2.09	47.5	0.999953
iPhone 5 - Phone 4	0	0	0	1
iPhone 5 - iPad Air	0.03	0.33	7.5	0.999993
iPhone 5 - Sony Experia	0.07	0.77	17.5	0.999983

ACKNOWLEDGMENT

The authors would like to thank L. Wang and S. Doclo for providing the CM-2 algorithm program and S. Markovich-Golan for providing the LPDE algorithm program.

Dani Cherkassky (S'13) was born in Latvia in 1979. He received the B.S. degree from the Tel Aviv Engineering Collage (summa cum laude) and the M.Sc. degree from Tel-Aviv University, both in Electrical Engineering, in 2005 and 2012, respectively. He is now beginning his PhD degree at the School of Electrical Engineering at Bar-Ilan University. Since 2005, he has been working in the high-tech industry as a DSP and Algorithm Engineer. His areas of interest are in statistical signal processing and estimation theory.

Sharon Gannot (S'92-M'01-SM'06) received his B.Sc. degree (summa cum laude) from the Technion Israel Institute of Technology, Haifa, Israel in 1986 and the M.Sc. (cum laude) and Ph.D. degrees from Tel-Aviv University, Israel in 1995 and 2000 respectively, all in Electrical Engineering. In 2001 he held a post-doctoral position at the department of Electrical Engineering (ESAT-SISTA) at K.U.Leuven, Belgium. From 2002 to 2003 he held a research and teaching position at the Faculty of Electrical Engineering, Technion-Israel Institute of Technology, Haifa, Israel. Currently, he is a Full Professor at the Faculty of Engineering, Bar-Ilan University, Israel, where he is heading the Speech and Signal Processing laboratory and the Signal Processing Track.

Prof. Gannot is the recipient of Bar-Ilan University outstanding lecturer award for 2010 and 2014. He is also a co-recipient of seven best paper awards.

Prof. Gannot has served as an Associate Editor of the EURASIP Journal of Advances in Signal Processing in 2003-2012, and as an Editor of several special issues on Multi-microphone Speech Processing of the same journal. He has also served as a guest editor of ELSEVIER Speech Communication and Signal Processing journals. Prof. Gannot has served as an Associate Editor of IEEE Transactions on Speech, Audio and Language Processing in 2009-2013. Currently, he is a Senior Area Chair of the same journal. He also serves as a reviewer of many IEEE journals and conferences.

Prof. Gannot is a member of the Audio and Acoustic Signal Processing (AASP) technical committee of the IEEE since Jan., 2010. Since Jan. 2017, he serves as the committee chair. He is also a member of the Technical and Steering committee of the International Workshop on Acoustic Signal Enhancement (IWAENC) since 2005 and was the general co-chair of IWAENC held at Tel-Aviv, Israel in August 2010. Prof. Gannot has served as the general co-chair of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA) in October 2013. Prof. Gannot was selected (with colleagues) to present a tutorial sessions in ICASSP 2012, EUSIPCO 2012, ICASSP 2013 and EUSIPCO 2013. Prof. Gannot research interests include multi-microphone speech processing and specifically distributed algorithms for ad hoc microphone arrays for noise reduction and speaker separation; dereverberation; single microphone speech enhancement and speaker localization and tracking.

REFERENCES

[1] S. Gannot, D. Burshtein, and E. Weinstein, "Signal enhancement using beamforming and nonstationarity with applications to speech," *IEEE Transactions on Signal Processing*, vol. 49, no. 8, pp. 1614–1626, 2001.

[2] Herbert Buchner, Robert Aichner, and Walter Kellermann, "Trinicon: A versatile framework for multichannel blind signal processing," in *Proceedings of IEEE International Conference on, Acoustics, Speech and Signal Processing (ICASSP)*, 2004, vol. 3, pp. iii–889.

[3] S. Doclo and M. Moonen, "GSVD-based optimal filtering for single and multimicrophone speech enhancement," *IEEE Transactions on Signal Processing*, vol. 50, no. 9, pp. 2230–2244, 2002.

[4] S. Makino, T. Lee, and H. Sawada, *Blind speech separation*, Springer, 2007.

[5] Y. Dorfan and S. Gannot, "Tree-based recursive expectation-maximization algorithm for localization of acoustic sources," *IEEE/ACM Transactions on Audio, Speech, and Language Processing,*, vol. 23, no. 10, pp. 1692–1703, 2015.

[6] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Optimum Array Processing*, John Wiley & Sons, 2004.

[7] Jacob Benesty, Jingdong Chen, and Yiteng Huang, *Microphone array signal processing*, vol. 1, Springer Science & Business Media, 2008.

[8] S. Markovich-Golan, A. Bertrand, M. Moonen, and S. Gannot, "Optimal distributed minimum-variance beamforming approaches for speech enhancement in wireless acoustic sensor networks," *Signal Processing*, vol. 107, pp. 4–20, 2015.

[9] IEEE Standards Association, "Standard for a precision clock synchronization protocol for networked measurement and control systems," *IEEE 1588*, 2002.

[10] Y. Wu, Q. Chaudhari, and E. Serpedin, "Clock synchronization of wireless sensor networks," *IEEE Signal Processing Magazine*, vol. 28, no. 1, pp. 124–138, 2011.

[11] Luca Schenato and Federico Fiorentin, "Average timesync: A consensus-based protocol for time synchronization in wireless sensor networks," *IFAC Proceedings Volumes*, vol. 42, no. 20, pp. 30–35, 2009.

[12] Q. M. Chaudhari, E. Serpedin, and K. Qaraqe, "On maximum likelihood estimation of clock offset and skew in networks with exponential delays," *IEEE Transactions on Signal Processing,*, vol. 56, no. 4, pp. 1685–1697, 2008.

[13] W. Su and I. F. Akyildiz, "Time-diffusion synchronization protocol for wireless sensor networks," *IEEE/ACM Transactions on Networking,*, vol. 13, no. 2, pp. 384–397, 2005.

[14] S. Wehr, I. Kozintsev, R. Lienhart, and W. Kellermann, "Synchronization of acoustic sensors for distributed ad-hoc audio networks and its use for blind source separation," in *the 6th IEEE International Symposium on Multimedia Software Engineering*, 2004, pp. 18–25.

[15] J. Schmalenstroeer, P. Jebramcik, and R. Haeb-Umbach, "A combined hardware–software approach for acoustic sensor network synchronization," *Signal Processing*, vol. 107, pp. 171–184, 2015.

[16] M. Pawig, G. Enzner, and P. Vary, "Adaptive sampling rate correction for acoustic echo control in voice-over-IP," *IEEE Transactions on Signal Processing*, vol. 58, no. 1, pp. 189–199, 2010.

[17] G. Evangelista, "Design of digital systems for arbitrary sampling rate conversion," *Signal processing*, vol. 83, no. 2, pp. 377–387, 2003.

[18] L. Erup, F. M Gardner, and R. A. Harris, "Interpolation in digital modems. ii. implementation and performance," *IEEE Transactions on Communications*, vol. 41, no. 6, pp. 998–1008, 1993.

[19] L.R. Rabiner, *Multirate digital signal processing*, Prentice Hall PTR, 1996.

[20] S. Markovich-Golan, S. Gannot, and I. Cohen, "Blind sampling rate offset estimation and compensation in wireless acoustic sensor networks with application to beamforming," in *International Workshop on Acoustic Signal Processing (IWAENC)*, 2012.

[21] N. Miyabe, S. Ono and S. Makino, "Blind compensation of inter-channel sampling frequency mismatch with maximum likelihood estimation in stft domain," in *Proceedings of IEEE International Conference on, Acoustics, Speech and Signal Processing (ICASSP)*, 2013, pp. 674–678.

[22] Y. Zeng, R. C. Hendriks, and N. D. Gaubitch, "On clock synchronization for multi-microphone speech processing in wireless acoustic sensor networks," in *Proceedings of IEEE International Conference on, Acoustics, Speech and Signal Processing (ICASSP)*, 2015, pp. 231–235.

[23] M. H. Bahari, A. Bertrand, and M. Moonen, "Blind sampling rate offset estimation based on coherence drd7t in wireless acoustic sensor networks," in *Proceedings of the 23rd EuropeanSignal Processing Conference (EUSIPCO), 2015*, 2015, pp. 2281–2285.

[24] D. Cherkassky, S. Markovich-Golan, and S. Gannot, "Performance analysis of MVDR beamformer in WASN with sampling rate offsets and blind synchronization," in *Proceedings of the 23rd European Signal Processing Conference (EUSIPCO), Nice, France*, 2015, pp. 245–249.

[25] L. Wang and S. Doclo, "Correlation maximization based sampling rate offset estimation for distributed microphone arrays," *IEEE/ACM Transactions on Audio, Speech, and Language Processing,*, vol. 24, no. 2, pp. 571–582, 2016.

[26] D. Cherkassky and S. Gannot, "Blind synchronization in wireless sensor networks with application to speech enhancement," in *Proceedings of the 14th International Workshop on Acoustic Signal Enhancement (IWAENC), 2014*, 2014, pp. 183–187.

[27] L. G. Weiss, R.K. Young, and L.H. Sibul, "Wideband processing of acoustic signals using wavelet transforms. part i. theory," *The Journal of the Acoustical Society of America*, vol. 96, no. 2, pp. 850–856, 1994.

[28] A. V. Oppenheim, R. W. Schafer, J. R. Buck, et al., *Discrete-time signal processing*, vol. 2, Prentice hall Englewood Cliffs, NJ, 1989.

[29] P. M. Schultheiss and E. Weinstein, "Estimation of differential doppler shifts," *The Journal of the Acoustical Society of America*, vol. 66, no. 5, pp. 1412–1419, 1979.

[30] S. S Rao, *Engineering optimization: theory and practice*, John Wiley & Sons, 2009.

[31] E.A.P. Habets, "Room impulse response generator," *Tech. Rep, Technische Universiteit Eindhoven*, 2006.

[32] H. L. Van Trees, *Detection, estimation, and modulation theory, optimum array processing*, Wiley, New York, 2004.

[33] S. Markovich-Golan, S. Gannot, and I. Cohen, "Multichannel eigenspace beamforming in a reverberant noisy environment with multiple interfering speech signals," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 6, pp. 1071–1086, 2009.

[34] S. Stein, "Algorithms for ambiguity function processing," *IEEE Transactions on acoustics, Speech and Signal Processing,*, vol. 29, no. 3, pp. 588–599, 1981.

[35] N. Levanon and E. Mozeson, *Radar signals*, John Wiley & Sons, 2004.