

ON CLOCK SYNCHRONIZATION FOR MULTI-MICROPHONE SPEECH PROCESSING IN WIRELESS ACOUSTIC SENSOR NETWORKS

Yuan Zeng, Richard C. Hendriks and Nikolay D. Gaubitch

Delft University of Technology, The Netherlands
{Y.Zeng, R.C.Hendriks, N.D.Gaubitch}@tudelft.nl

ABSTRACT

In wireless acoustic sensor networks (WASNs), clock synchronization is crucial for multi-microphone signal processing, since clock differences between capturing devices will cause signal drift. This in turn severely degrades the performance of multi-microphone signal processing. After a theoretical analysis of the effect of clock synchronization, we evaluate the use of three different clock synchronization algorithms in the context of multi-microphone noise reduction. Our experimental study shows that the achieved precision of clock synchronization enables sufficient accuracy of clock synchronization for the MVDR beamformer in ideal scenarios. However, in practical scenarios with measurement noise on the parameters of interest, time-stamp based clock synchronization algorithms get degraded, while signal based algorithms are still accurate enough for the MVDR beamformer, albeit at a much higher transmission cost.

Index Terms— Clock synchronization, speech enhancement, wireless acoustic sensor networks

1. INTRODUCTION

Wireless acoustic sensor networks (WASNs) have been proposed for speech enhancement by means of multi-microphone noise reduction [1, 2, 3, 4]. Multi-microphone noise reduction algorithms such as beamforming, heavily depend on timing information as they usually employ the delay that is experienced when an acoustic signal is observed at different positions. However, in WASNs, each node usually has its own processor with an independent internal clock. Employing multi-microphone noise reduction, in practice, requires that these clocks are synchronized. Most of the multi-microphone signal processing algorithms for WASNs (e.g., [1, 2, 3, 4]) are based on the often implicit assumption that the internal clocks are synchronized. An unsynchronized clock can cause drift of time differences of the observed signal at the different nodes, and, as a consequence degrade the performance of multi-microphone noise reduction.

Since clock synchronization is an important aspect for signal processing in wireless sensor networks (WSNs), several algorithms addressing this issue (not specifically for speech enhancement) have been presented [5, 6, 7, 8, 9, 10]. However, most studies on beamforming/speech enhancement in WASNs neglect the clock synchronization problem and simply assume the clocks to be synchronized. Studies on the required precision and the applicability of such algorithms in terms of data transmissions and robustness in practical scenarios is generally lacking. In this paper we therefore present an in-depth comparison of several clock synchronization algorithms for distributed speech enhancement. The consequence of an unsynchronized clock is a clock skew and a clock offset. To focus this paper, we will only consider the clock skew, which causes signal

drift and results in a poor performance of speech enhancement algorithms. Many algorithms for clock skew compensation employ a series of time message transmissions [6] [7]. In [6], a joint ranging and clock synchronization (JCS) algorithm is proposed to estimate relative clock skews, clock offsets and pairwise distances in a WSN using a single clock reference. This algorithm requires that the node with the reference clock serves as a central processor, connected with all other nodes in the network. The gossip-based clock synchronization (GbCS) algorithm in [7] is an algorithm based on time stamps and the randomized gossip algorithm [11]. Unlike the JCS algorithm where the clocks of all nodes are synchronized with respect to a reference node, the GbCS algorithm synchronizes them with a virtual clock. Thus, the GbCS algorithm synchronizes the clocks in a distributed way without the requirements of a reference clock or a specific network topology. The accuracy of these time-stamp based algorithms is, generally, proportional to the number of timing message transmissions. Another class of clock synchronization algorithms is based on the observed signal, such as the blind sampling-rate offset estimation (BSrOE) algorithm in [8] and the blind synchronization algorithm in [12]. Assuming that there is a reference node in the WASN, the BSrOE estimates relative clock skews using the phase drift in the coherence between the observed signals of two communicating nodes. Similar to the JCS, the BSrOE requires that the node with the reference clock serves as a central processor.

In the remainder of this paper, we present a study of the effect of clock synchronization on multi-microphone signal processing where each node has an individual clock. We perform theoretical and experimental investigations of the effects of clock synchronization on the delay-and-sum (DSB) and the minimum variance distortionless response (MVDR) beamforming using three state-of-the-art algorithms (i.e., the JCS, the GbCS and the BSrOE). In particular we analyze communication cost of the three algorithms and investigate their robustness to noise on the parameters used to synchronize the clocks.

2. PROBLEM STATEMENT AND NOTATION

Consider a WASN comprising N nodes randomly distributed in a noisy environment, where each node is driven by its own processor with an internal clock and contains one microphone. Let $y_i(t)$ denote a continuous-time signal observed at node i . Assuming that the signal $y_i(t)$ consists of a target source signal $x_i(n)$ degraded by additive noise $v_i(n)$, a common data model of $y_i(t)$ is given by $y_i(t) = x_i(t) + v_i(t)$. The challenge for noise reduction algorithms is to estimate the target signal from the noisy observations. With a conventional microphone array, the speech signal can be estimated using beamforming methods, such as the DSB or the MVDR beamformer, since all microphones have the same clock and sampling

rate. However, in a WASN, each node is equipped with an independent clock oscillator. Clock differences are therefore inevitable. Let t_i denote the local clock reading at node i , given by

$$t_i = \alpha_i t + \beta_i, \quad (1)$$

where t is the global time or the local time of a reference node, α_i is the clock skew and β_i is the clock offset. We assume the clock offset parameter to be known and we concentrate on the clock skew. The clock model t_i in (1) can then be simplified to

$$t_i = \alpha_i t. \quad (2)$$

Without loss of generality, we assume that the first node is the reference node (i.e., $t_1 = t$). Based on the time model in (2), the sampling rate at node i is given by $f_{s_i} = \alpha_i f_s$, where f_s is the sampling rate at a reference node (i.e., $f_s = f_{s_1}$). Let $y_i[n]$ denote the discrete-time observed signal at time-sampling index n . The discrete-time signal $y_i[n]$ can be obtained by sampling the continuous-time signal $y_i(t)$ at time $\frac{n}{f_{s_i}}$, i.e.,

$$y_i[n] = y_i\left(\frac{n}{f_{s_i}}\right), \quad t = n/f_{s_i} \text{ and } -\infty < n < +\infty. \quad (3)$$

Equation (3) indicates that different sampling rates cause drift of time difference between the observed digital signals. This problem can be solved by synchronizing sampling rates of all nodes, which can be realized by synchronizing clock skews of all nodes.

3. ANALYSIS OF THE CLOCK SYNCHRONIZATION PROBLEM FOR BEAMFORMING TECHNOLOGIES

In this section, we analyze the effect and importance of clock synchronization on beamforming technologies. To facilitate a simple and clear insight into the problem, we use the DSB and a synthetic signal.

As beamforming algorithms are usually conducted in the short-time discrete Fourier transform (DFT) domain, signals are windowed and transformed into the frequency domain by applying a DFT. Let $Y_i(f, m)$, $X_i(f, m)$ and $V_i(f, m)$ denote the observed signal, the desired signal and the noise DFT coefficient at frequency-bin index f and discrete-time frame index m , respectively. The speech DFT coefficient $X_i(f, m)$ of the target source is given by $X_i(f, m) = d_i(f, m)S_i(f, m)$, where $d_i(f, m)$ is the acoustic transfer function (ATF), and $S_i(f, m)$ is the clean signal at the target location, both with sampling rate f_{s_i} . To estimate the clean signal, $Y_i(f, m)$ can be stacked into a vector, say $\mathbf{Y}(f, m) = [Y_1(f, m), \dots, Y_N(f, m)]^T$, with $[\cdot]^T$ the transposition of a vector or a matrix, followed by filtering with $\mathbf{W}(f, m)$ (i.e., $\hat{S}_1(f, m) = \mathbf{W}^H(f, m)\mathbf{Y}(f, m)$). However, since all nodes have different sampling rates, the beamformer performance will be degraded depending on the differences between sampling rates.

Consider a WASN with two nodes, each with one microphone. The sampling-rate of node 1 is $f_{s_1} = f_s = 16$ kHz (i.e., the reference node) and the sampling-rate of node 2 is $\alpha_2 f_s$. To assess the performance of the DSB versus α_2 , we use the spatial directivity pattern $Q(\omega)$ of the DSB, that is

$$Q(\omega) = \mathbf{d}^H \tilde{\mathbf{d}} \left(\mathbf{d}^H \mathbf{d} \right)^{-1}, \quad (4)$$

with $(\cdot)^H$ Hermitian transposition, $\tilde{\mathbf{d}}$ the ATF under sampling-rate mismatch, and \mathbf{d} the ATF without sampling-rate mismatch. If there is a sampling-rate mismatch and $\mathbf{d} \neq \tilde{\mathbf{d}}$ and $Q(\omega)$ measure the amount of mismatch, which reflects the distortion that

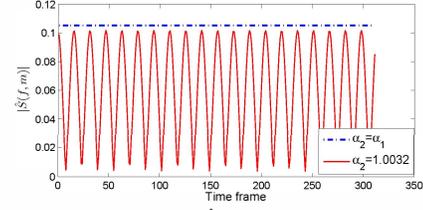


Fig. 1. The distortion $|\hat{S}(f, m)|$ versus time frames

the clean signal will undergo after processing by the beamformer. In a free-field without damping (i.e., $|d_i| = 1$), we then have $\mathbf{d} = \left[e^{-j\omega \frac{l_1}{c} f_s}, e^{-j\omega \frac{l_2}{c} f_s} \right]^T$ and $\tilde{\mathbf{d}} = \left[e^{-j\omega \frac{l_1}{c} f_s}, e^{-j\omega \frac{l_2}{c} f_{s,2}} \right]^T$, where l_i is the distance from the source to the i th node, and $c = 340$ m/s is the speed of sound. Then, $Q(\omega)$ can be expressed as

$$Q(\omega) = \cos\left(\omega \frac{l_2}{2c} (1 - \alpha_2) f_s\right) e^{j\omega \frac{l_2}{2c} (1 - \alpha_2) f_s}. \quad (5)$$

Notice that damping can easily be included, but is left out here for clarity of presentation. Equation (5) shows that $|Q(\omega)|$ is periodic as a function of the clock skew. Ideally, for the case when there is no clock skew (i.e., $\alpha_i = 1, \forall i$), $Q(\omega) = 1$. Obviously, when there is clock skew, $|Q(\omega)|$ deviates from 1 and distortions are introduced in the spatial directivity pattern. That is, the beamformer response in the target direction may be suppressed depending on α_i .

To further assess the distortions introduced in the estimated clean DFT coefficients, we investigate the output of the DSB when applied to the clean input only, i.e., $V_i = 0, \forall i$. In this case, we set the clean target signal to a sinusoidal signal $s(t) = \cos(2\pi\nu t)$ with $\nu = 1250$ Hz. Sampling this signal with the sampling frequency of node 1 and node 2 (i.e., $f_{s_1} = f_s$ and $f_{s_2} = \alpha_2 f_s$), leads to $s_1[n] = \cos(2\pi\nu n/f_s)$ and $s_2[n] = \cos(2\pi \frac{\nu}{\alpha_2 f_s} n)$ with the two different frequencies ν and ν/α , respectively. Let $S_1(f, m)$ and $S_2(f, m)$ denote the DFT of a windowed frame of $s_1[n]$ and $s_2[n]$, respectively. Stacking these DFT coefficients in a vector, and including the delays τ_1 and τ_2 due to the signal propagation over distances l_1 and l_2 , respectively, we get $\mathbf{X}(f, m) = [S_1(f, m)e^{-j\omega_f \tau_1}, S_2(f, m)e^{-j\omega_f \tau_2}]^T$, with $\tau_1 = l_1 f_{s_1}/c$ and $\tau_2 = l_2 f_{s_2}/c$.

Let $\hat{S}(f, m)$ denote the DSB output when applying the beamformer to the clean signals only (i.e., to $\mathbf{X}(f, m)$). When there is no clock skew, the output equals the clean signal, $\hat{S}(f, m) = \frac{1}{2}(S_1(f, m) + S_2(f, m))$. However, the DSB output under clock skew is given by

$$\hat{S}(f, m) = \frac{1}{2} \left(S_1(f, m) + S_2(f, m) e^{j2\pi f (1 - \alpha_2) l_2 / c} \right). \quad (6)$$

Two effects become apparent. What should be the average of the DFT coefficient $\hat{S}(f, m)$ of two windowed sinusoids with similar frequency and compensated delay such that they constructively add, becomes the sum of the DFT coefficients of two windowed sinusoids with a) two different frequencies, and b) a delay with respect to each other that is not correctly compensated.

As an example, Fig. 1 shows the value $|\hat{S}(f, m)|$ with and without clock skew for a fixed frequency bin $f = 20$ (chosen as the bin with center frequency closest to $\nu = 1250$) across time frames. The blue dashed line shows the estimated $|\hat{S}(f, m)|$ when there is no clock skew ($\alpha_1 = \alpha_2$) and the red solid line shows for a fixed

$\alpha_2 = 1.0032$. The distortion $|\hat{S}(f, m)|$ varies periodically across time and the distortion in $\hat{S}(f, m)$ is upper and lower bounded.

The above analysis and simulations show that asynchronous clocks in WASNs can severely degrade the performance of beamformers. First, the delay compensations by \mathbf{d} are incorrect, leading to an undesired beamformer response. Secondly, at nodes with clock skew $\alpha_i \neq 1$, the signal gets translated to another frequency. Depending on α_i this is audibly perceived as the sum of two speech signals that are not aligned with respect to each other and have different sampling frequencies. A solution is to perform clock synchronization and/or sampling-rate offset compensation.

4. COMMUNICATION COST ANALYSIS

In this section, we make a communication cost analysis of the three clock synchronization algorithms that we compare. These are the JCS [6], the GbCS [7] and the BSrOE [8], which were briefly explained in the introduction. Although each algorithm has its own requirements (e.g., the number of transmissions, the network topology, centralized or decentralized processing, etc.), they can all be used to solve the sampling-rate synchronization problem in WASNs albeit at different costs and requirements.

For the analysis, we define one data transmission as the sending of a scalar value from one node to another. In both the JCS and the GbCS, clocks are synchronized by exchanging time information, which is a scalar value of the time-stamp. In a fully connected network with N nodes, the number of data transmissions of the JCS is given by

$$T_J = 2K(N - 1), \quad (7)$$

since all $N - 1$ pairs of neighboring nodes (each pair includes the reference node and one other node) communicate K times with 2 transmissions each time.

The number of data transmissions of the GbCS algorithm is

$$T_G = 4C, \quad (8)$$

with C number of iterations. At each iteration, two neighboring nodes communicate a time message and clock skew compensation parameter. This means twice the transmission of two variables per iteration.

In the BSrOE, all $N - 1$ nodes (all nodes except the reference node) send the DFT coefficients of their observed signals to the reference node, which serves as the central processor. Thus, the data transmission of the BSrOE can be computed as

$$T_B = (N - 1)P f_{\max}, \quad (9)$$

where there are P segments of microphone signals and f_{\max} frequency bins per segment. Notice that the required number of data transmissions of the JCS and the BSrOE given in (7) and (9), respectively, consider only a fully connected network. More data transmissions are required for clock synchronization when used in non-fully connected networks as the time message information or microphone signals need to be sent to the central processor using relay nodes. Only the data transmissions of the GbCS given in (8) is directly applicable for randomly connected networks.

5. EXPERIMENTAL STUDY

In this section, we study the performance of the three clock synchronization algorithms and evaluate their effect on the performance of

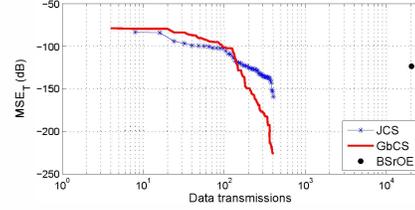


Fig. 2. The MSE versus number of transmissions.

the MVDR beamformer in terms of instrumental speech quality and speech intelligibility metrics.

We simulate a WASN with five fully connected nodes, and consider a free-field scenario. Thus, the steering vector \mathbf{d} is determined by gain and delay values. The speech source consists of a 30 seconds speech signal sampled at 16 kHz originating from the Timit [13] database, and the noise source is a babble noise signal. All nodes in the WASN first synchronize their clocks using one of the three algorithms, and then process the signals in the frequency domain using a frame-based MVDR beamformer, with a frame length of 32 ms and a 50%-overlapping Hann window. The following parameters are used in the BSrOE. The Welch method is used with a DFT size of $F = 4096$ and 75% overlap. Each segment consists of $L = 16000$ samples and $P = 32$ segments with 50% overlap are used to estimate the sampling-rate offset, which is bounded by $e_{\max} = 800$ ppm with $\text{ppm} = 10^{-6}$. The frequency bins per segment f_{\max} can be obtained as $f_{\max} = \frac{F}{2 \times L \times e_{\max}}$. The clock skew of the five nodes are set to $\alpha = [1, 1.0001, 1.0002, 1.0003, 1.0004]^T$. All simulations in this section are based on this scenario.

To assess the estimation accuracy of the clock skew, we define the mean square error (MSE) between the estimated clock skews $\hat{\alpha}_i$ of all nodes and the reference clock skew α_{ref} as $\text{MSE}_T = \frac{1}{N} \sum_{i=1}^N |\hat{\alpha}_i - \alpha_{\text{ref}}|^2$. Furthermore, we use the segmental signal-to-noise ratio (SNR_{seg}) [4] and the short-time objective intelligibility measure (STOI) [14] to assess the speech quality and speech intelligibility of the MVDR beamformer, respectively. As reference signal in STOI and SNR_{seg} we use the clean signal sampled by the reference clock at the reference node. For notational convenience, we denote the MVDR with perfect clock synchronization by C-MVDR, the MVDR beamformer without clock synchronization by E-MVDR, the MVDR beamformer with the JCS by J-MVDR, the MVDR beamformer with the GbCS by G-MVDR, the MVDR beamformer with the BSrOE by B-MVDR.

5.1. Clock synchronization without measurement noise

We begin with the assumption that there is no measurement noise on the time-stamp in the JCS and GbCS, and the observed signal used in the BSrOE is a babble noise-only. This follows the ideal circumstances described in the original papers.

Figure 2 shows that all three algorithms can reach the same accuracy of clock synchronization in terms of MSE_T with enough data transmissions. The estimation accuracy of the clock skew in the JCS and GbCS is increased with increasing number of data transmissions. Further, the BSrOE needs more data transmissions to reach a performance similar to that of the JCS and GbCS.

Figure 3 shows the effect of clock synchronization on the MVDR beamformer. In Fig. 3(a), we see that the SNR_{seg} of the E-MVDR output is even lower than those of the input noisy signal for global input SNRs larger than 2 dB. In Fig. 3(b), it can be seen

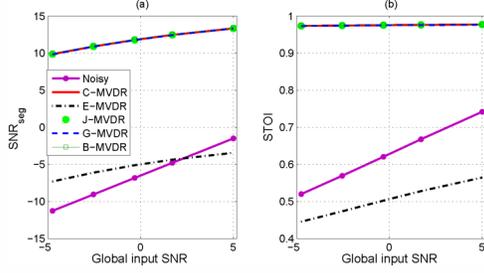


Fig. 3. (a) The SNR_{seg} of node 1 versus the global input SNR. (b) The STOI of node 1 versus the global input SNR.

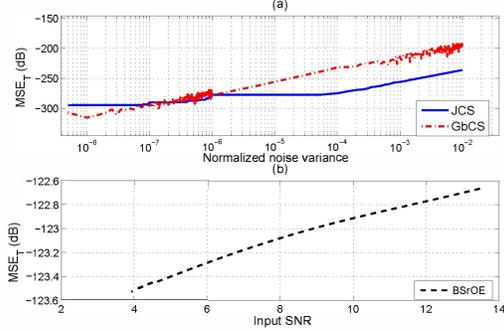


Fig. 4. (a) The MSE_T versus the noise variance on time stamp information. (b) The MSE_T versus the global input SNR.

that the STOI values of the E-MVDR output are smaller than those of the noisy input signal. This indicates that the predicted speech quality and intelligibility of the MVDR is severely degraded without clock synchronization. Note that to obtain absolute intelligibility scores, the STOI output needs to be mapped using for example a logistic function. Moreover, these results also indicate that noise reduction performance of the MVDR with clock synchronization (i.e., J-MVDR, G-MVDR and B-MVDR) can reach the same performance as the C-MVDR, where clocks of all microphones are perfectly synchronized with the reference clock.

5.2. Clock synchronization with noisy parameters

Next, we investigate the performance of the clock synchronization algorithms in a realistic setup where the measurements are subject to imperfections. For the JCS and GbCS, this means that we add white Gaussian noise to the time-stamps with a variance that is normalized by the precision of the internal clock. The system clock in modern PCs runs at 66 MHz. The minimum difference between two time stamps is thus $1/(66 \times 10^6)$. The variance on the measurement noise is then given by $66 \times 10^6 \times \sigma^2$, with σ^2 the variance of the white Gaussian noise process. Since the performance of the JCS and GbCS depends on the number of transmissions, we use for both algorithms the same amount of 400 data transmissions. For the BSrOE we use a noisy speech signal instead of the pure noise. The noisy signal consists of a speech signal degraded by additive babble noise at several input SNRs. Note that the BSrOE requires much more data transmissions than the other two reference algorithms, namely 20480.

In Fig. 4(a), the MSE_T for both the JCS and the GbCS is seen to increase with increasing measurement-noise variance. In addition,

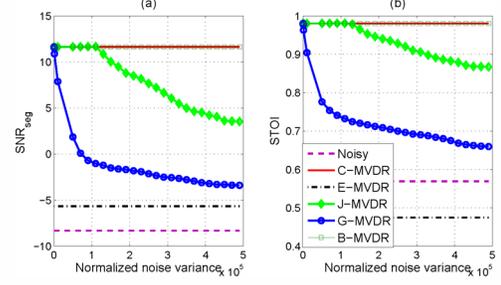


Fig. 5. (a) The SNR_{seg} of node 1 versus the noise variance. (b) The STOI of node 1 versus the noise variance.

the results show that the MSE_T of the JCS increases slower than those of the GbCS. This is reasonable, since the time model in the JCS can take measurement noise of time-stamps into account, and the clock parameters in the JCS are estimated by minimizing the least squares norm of the measurement noise of time-stamps, while the time model in the GbCS assumes that there is no measurement noise on the time-stamps. The JCS shows a better estimation accuracy of the clock skew than the gossip based algorithm. In Fig. 4(b), the MSE_T of the BSrOE is slightly increased with increasing global input SNR, which indicates that the effect of the SNR of the observed signal on the estimation accuracy of the BSrOE in terms of MSE is small, around 1 dB in this SNR range.

To illustrate the performance of the MVDR beamformer in the situation with measurement noise on the time-stamp, we investigate the SNR_{seg} and the STOI of the MVDR output of reference node 1 versus the normalized noise variance. The global input SNR of the signal at node 1 is -2.5 dB. In Figs. 5(a) and 5(b), both the SNR_{seg} and the STOI of the J-MVDR and the G-MVDR are decreased with increasing noise variance on the time stamps. Although the B-MVDR uses the noisy speech signal for clock synchronization, it reaches the same performance as the C-MVDR, since the B-MVDR is a signal-based algorithm, not sensitive for time-stamp noise. The performance of the J-MVDR decreases slower than that of the G-MVDR. This is consistent with the simulation results in Fig. 4. Moreover, for small noise variances, the J-MVDR reaches the same performance as the C-MVDR, while this is at a much lower transmission cost than the B-MVDR.

6. CONCLUSIONS

In this paper, we first analyzed effects of clock synchronization on the DSB with a synthetic signal. Then, we analyzed communication cost of three different clock synchronization algorithms. From this, it follows that the BSrOE requires a significantly larger amount of transmissions than the JCS and the GbCS approaches due to the fact that it is signal based. To which extent this high data-transmissions is a problem for distributed signal processing, depends on the processing in the subsequent steps. The experimental study has shown that the accuracy of clock synchronization of the three algorithms is sufficient for the MVDR beamformer under ideal circumstances. In scenarios with measurement uncertainty or noise, the output of the MVDR with the JCS and the GbCS degrades, but the MVDR with the BSrOE reaches the same performance as the centralized MVDR beamformer. For small amounts of measurement noise, the JCS gives similar performance as the BSrOE, but, at a significantly lower amount of data transmissions.

7. REFERENCES

- [1] A. Bertrand and M. Moonen, "Distributed adaptive node-specific signal estimation in fully connected sensor networks – part I: Sequential node updating," *IEEE Trans. Signal Processing*, vol. 58, no. 10, pp. 5277–5291, oct. 2010.
- [2] R. Heusdens, G. Zhang, R. C. Hendriks, Y. Zeng, and W. B. Kleijn, "Distributed MVDR beamforming for (wireless) microphone networks using message passing," in *Int. Workshop on Acoustic Echo and Noise Control*, 2012.
- [3] S. Markovich-Golan, S. Gannot, and I. Cohen, "Distributed multiple constraints generalized sidelobe canceler for fully connected wireless acoustic sensor networks," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, pp. 343–356, Oct. 2012.
- [4] Y. Zeng and R. C. Hendriks, "Distributed delay and sum beamformer for speech enhancement via randomized gossip," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 22, pp. 260 – 273, Jan. 2014.
- [5] Y. C. Wu, Q. Chaudhari, and E. Serpedin, "Clock synchronization of wireless sensor networks," *IEEE Signal Processing Magazine*, vol. 21, pp. 260 – 273, Nov. 2013.
- [6] R. T. Rajan and A. Veen, "Joint ranging and clock synchronization for a wireless network," in *IEEE Int. Workshop on Computational Advances in Multi-Sensor Adaptive Processing*, 2011, pp. 297–300.
- [7] L. Schenato and F. Fiorentin, "Average timesynch: a consensus-based protocol for time synchronization in wireless sensor networks," *Automatica*, vol. 47, no. 9, pp. 1878–1886, 2011.
- [8] 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 *Int. Workshop on Acoustic Echo and Noise Control*, 2012.
- [9] J. Schmalenstroerer and R. Haeb-Umbach, "Sampling rate synchronisation in acoustic sensor networks with a pre-trained clock skew error model," in *Proc. European Signal Proc. Conf. Eusipco*, Marrakesh, Morocco, Sep 2013.
- [10] J. Schmalenstroerer, P. Jebramcik, and R. Haeb-Umbach, "A gossiping approach to wireless acoustic sensor network sampling clock synchronization," in *IEEE Int. Conf. Acoust. Speech, Signal Process. (ICASSP)*, Florence, May 2014, pp. 4037–4040.
- [11] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2508 – 2530, june 2006.
- [12] D. Cherkassky and S. Gannot, "Blind synchronization in wireless sensor networks with application to speech enhancement," in *Int. Workshop on Acoustic Echo and Noise Control*, 2014.
- [13] J. S. Garofolo, "DARPA TIMIT acoustic-phonetic speech database," *National Institute of Standards and Technology (NIST)*, 1988.
- [14] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "An algorithm for intelligibility prediction of time-frequency weighted noisy speech," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 7, pp. 2125–2136, Sep. 2011.