METHODOLOGY

Open Access



Online distributed waveform-synchronization for acoustic sensor networks with dynamic topology

Aleksej Chinaev^{1*}, Niklas Knaepper¹ and Gerald Enzner¹

Abstract

Acoustic sensing by multiple devices connected in a wireless acoustic sensor network (WASN) creates new opportunities for multichannel signal processing. However, the autonomy of agents in such a network still necessitates the alignment of sensor signals to a common sampling rate. It has been demonstrated that waveform-based estimation of sampling rate offset (SRO) between any node pair can be retrieved from asynchronous signals already exchanged in the network, but connected online operation for network-wide distributed sampling-time synchronization still presents an open research task. This is especially true if the WASN experiences topology changes due to failure or appearance of nodes or connections. In this work, we rely on an online waveform-based closed-loop SRO estimation and compensation unit for nodes pairs. For WASNs hierarchically organized as a directed minimum spanning tree (MST), it is then shown how local synchronization propagates network-wide from the root node to the leaves. Moreover, we propose a network protocol for sustaining an existing network-wide synchronization in case of local topology changes. In doing so, the dynamic WASN maintains the MST topology after reorganization to support continued operation with minimum node distances. Experimental evaluation in a simulated apartment with several rooms proves the ability of our methods to reach and sustain accurate SRO estimation and compensation in dynamic WASNs.

Keywords Dynamic wireless acoustic sensor networks, Sampling rate offset estimation and compensation

1 Introduction

The availability of smart devices equipped with diverse sensors has stimulated ample research in wireless sensor networks (WSNs) [1–6]. Meanwhile, wireless acoustic sensor networks (WASNs) have emerged as a research area of its own [7–9]. Due to the autonomy of agents, methods for sampling-time synchronization are a crucial piece of network infrastructure to discipline all WASN nodes to a consistent sampling rate [10]. However,

considerable attention is still required for smooth and efficient network-wide treatment.

Importance of time synchronization for signal processing in WASN is evident from the fact that asynchronous signals even with sampling rate offset (SRO) values in the subhertz range cause a significant decrease of overall network performance, such as, in acoustic source separation that operates with a sampling rate of 16 kHz, an SRO of only 1 Hz leads to a drop of the signal-to-interferenceratio gain from 9 - 10 dB down to only 3 - 4 dB [11, 12]. For similar SRO values, the intelligibility of distributed beamforming-based noise reduction is reduced from 0.8 up to 0.5 in terms of extended short-term objective intelligibility values, if sensor nodes are equipped with one or two microphones [13]. The SRO quantity is often normalized to the sampling rate of a reference node and



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

^{*}Correspondence:

Aleksej Chinaev

aleksej.chinaev@uol.de

¹ Division of Speech Technology and Hearing Aids, Department of Medical Physics and Acoustics, Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany

measured in parts per million $(ppm)^1$, since in real-world WASN applications it is usually a rather small value within the range of ± 100 ppm [14, 15].

Two core tasks of time synchronization are estimation and compensation of all SRO values [16, 17]. SRO compensation can be implemented either in hardware by changing the oscillator frequency (requiring a direct access to respective circuitry of analog-to-digital converters) or in software by digital-to-digital conversion (i.e., resampling) of microphone signals [18]. In the scope of this publication, we rely on comprehensive options for software-based online-capable SRO compensation [19– 24]. Methods for SRO estimation are generally based either on time stamp exchange between network agents or on the acoustic waveforms already shared for joint signal processing [25].

Time stamp-based SRO estimation has traditionally received larger attention, especially for network-wide distributed synchronization of WSNs [26-30], which aims at shared responsibilities across the network and at scalability in terms of communication bandwidth and computational load in contrast to centralized network operation [10]. In such scenarios, the time stamps are exclusively exchanged either in one-way or in two-way communication procedure between neighboring nodes, which is referred to as a gossiping approach [31]. In the seminal work [32], a wide-spread timing-sync protocol for sensor networks (TPSN) has been proposed where network-wide clock synchronization is provided by two consecutive steps: organization of the network in a hierarchical topology and pair-wise synchronization of network agents along the topology edges. Furthermore, a reference node is set to whose timing all other nodes are to be aligned. Further control must be applied with the TPSN scheme to accommodate dynamic WSNs [33], meaning networks that may change their structure during operation as a reaction to failure or appearance of nodes or communication links. Similar techniques are hardly available for waveform-based network-wide synchronization of WASNs and a major goal of this paper is to fill this gap.

Waveform-based SRO estimation solely uses asynchronous acoustic signals without any time stamp information or protocol [34–45], which is particularly rational when the network already exchanges acoustic waveforms for joint acoustic signal processing over the network. Typical acoustic excitation here is a directional or diffuse sound field from single or multiple acoustic sources like speech, music, or even spatially correlated noise in non-reverberant and reverberant settings². With the exception of [45], waveform-based methods typically operate on pairs of sensor signals, i.e., one reference signal with nominal sampling and one non-reference signal with SRO. Apart from [34], the methods for pairwise waveform-based SRO estimation can be categorized into three groups. The first group makes explicit use of the complex-valued spectral coherence function, whose phase drift is directly connected to the underlying SRO [35, 39, 40, 44]. Methods of the second group rest upon statistical modeling of short-time Fourier transform (STFT) coefficients [36, 41]. A desired SRO value is estimated here via maximization of the likelihood function defined on STFT coefficients of asynchronous and pre-synchronized sensor signals. In the third group, different techniques for correlation or coherence processing are deployed either in the time domain or in the STFT domain [37, 38, 42, 43]. Note that with the exception of [38, 44], the majority of the waveform-based methods are designed for offline SRO estimation.

Considering a network-wide waveform-based synchro*nization*, small WASNs comprising more than two sensor nodes have been investigated in [38, 42, 44, 45] with no particular considerations regarding the network topology (it appears centralized). In [38, 42, 44] every sensor node is directly connected to the central reference node via a single-hop link. In larger networks, the centralized topology, however, leads to a computational overload of the central node and to an inefficient use of a communication bandwidth or can even be completely unfeasible [28]. In [45], all sensor nodes were linked with each other in a so-called fully connected topology that is even more demanding than a centralized topology. To avoid the drawbacks of the centralized method, a distributed SRO estimation for WASNs with arbitrary topology has been proposed very recently [47], however, only for offline signal processing based on a specific calibration signal and implemented only on fully connected or almost-fully connected topologies. From time stamp-based WSN synchronization [32], we know that networks can be more efficiently organized in hierarchical tree topologies and synchronized by distributed procedures where every node aligns its own signal to the sampling rate of the reference node. On the way to a distributed online-capable waveform-based synchronization, we have come up with a number of own developments that are briefly described as next.

1.1 Relation to own works

Before the synchronization of acoustic sensor networks received greater attention, a precursor of waveform-based

 $^{^1\,}$ An SRO of 1 Hz corresponds to 62.5 ppm for the sampling rate of 16 kHz.

² No calibration signal is explicitly required here in contrast to [11, 12, 46].



Fig. 1 Simulated SINS apartment (left); WASN representation (blue/red dots correspond to acoustic sources/nodes) by PaderWASN toolbox (right)

SRO estimation and compensation was described in the context of acoustic echo cancellation [48], where SRO was tracked by means of an LMS-type adaptive filter operating on two slightly asynchronous input signals. A related tracking theory for adaptive filters with asynchronous input and output signals was later reported in [49].

In the context of WASNs, as in Fig. 1, a double-crosscorrelation processor (DXCP) in the time domain with remarkable robustness to acoustic reverberation and noise has been proposed in [50] and restated as an FFTbased implementation with phase transform (PhaT) for online SRO estimation with outstanding accuracy [51]³. DXCP essentially refers to the concept of a secondary cross-correlation computed over a moving primary cross-correlation on signals with SRO. The secondary correlation then allows unbiased extraction of the underlying SRO. The DXCP-PhaT version has further evolved with a demonstration of robustness to packet loss in WASNs [52], with a closed-loop implementation to integrate sampling rate compensation [53], with extensions for tree-based distributed network-wide time synchronization [54], and very recently with robustness for long-term operation under nonpersistent acoustic activity [55].

The real-world utility of DXCP-based SRO estimation has been assessed with open-source developments of demonstrators in a larger research unit on acoustic sensor networks: (1) a first demo at WASPAA-2021 uses the

³ Similar to SRO estimators based on time stamp exchange, e.g. ,from [31], the waveform-based DXCP-PhaT achieves root-mean-square error (RMSE) of around 0.03 ppm without a need of an additional communication link.

MARVELO software on Raspberry Pi computers [56, 57] as a framework for our online SRO estimation between two sensor nodes; (2) a second demo at IWAENC-2022 uses Python notebooks to present the network-wide closed-loop WASN synchronization on various topologies and geometries created by means of the PaderWASN toolbox [44] applied to the Sound Interface to the Swarm (SINS) apartment [58] simulated as shown by [59] and depicted in Fig. 1⁴.

1.2 Proposals of this contribution

Based on our previous developments, a distributed online-capable network-wide waveform-synchronization will be proposed in this paper. Additionally, it will be extended for use in dynamic WASNs. *The specific novelty* of our contribution here is threefold.

- 1) All propagation of state and information in a network based on distributed local operation takes its time and effort [54]. To support the information flow for network consensus, we propose:
 - A buffer-based closed-loop (online) SRO estimation and compensation taking place from the outset and round-robin on all nodes of the network in Section 3.2,
 - A network topology according to a minimum spanning tree (MST) for better local connectivity in the network in Section 3.3.

⁴ Source code for demo (1) at https://github.com/CN-UPB/WASN and for demo (2) in "/distributed_synchro_demo" at https://github.com/fgnt/asnsig.



Fig. 2 Workflow of the paper

- 2) Real-world networks with continued operation will sooner or later experience radical modifications, such as the appearance of new nodes or failure of nodes and communication links between them. Section 4 therefore introduces a somewhat generic network protocol to handle these modifications with sustained synchronization of already synchronous network parts but with new MST configuration for continued operation.
- 3) An acoustic shoe-box room simulation might be an oversimplified enclosure regarding acoustic connectivity of the available network nodes. Thus, we simulate a sophisticated SINS apartment with several connected rooms in Sections 5.1 and 6.1 in order to meaningfully assess DXCP-based network-wide synchronization under the aforementioned organizational constraints.

The paper is otherwise organized as shown by Fig. 2, where sections with the specific novelty are marked by superscript asterisks. Methods for pairwise waveform-based synchronization are revisited in Section 2 to support our distributed network synchronization in Section 3 and our proposed synchronization protocol for dynamic WASN in Section 4. Experiments including a proof of concept and a large-scale quantitative assessment followed by some ablation studies are reported in Sections 5, 6, and 7.

2 Sampling rate offset and pairwise waveform-based signal synchronization

After introduction of SRO, its impact on the acoustic sensor signals in time and frequency domain is discussed. Furthermore, components of a waveform-based synchronization are considered including SRO compensation that consists of an integer-based time shift of asynchronous signal followed by signal resampling. Finally, a closed-loop architecture for pairwise signal synchronization [53] is explained more elaborately.

2.1 SRO parameter and its impact on a sensor signal

Considering a sensor node equipped with a single microphone, a noisy microphone signal can be represented by an additive signal model y(t) = x(t) + v(t), where t is the continuous time, x(t) a noise-free acoustic recording and v(t) a sensor self-noise. Assuming a perfect analog-to-digital converter (ADC) that is able to sample y(t) at a reference sampling rate f_r , a discrete-time noisy microphone signal is given by $y[n] = y(T_r \cdot n)$, where $n \ge 0$ is the discrete time and $T_r = 1/f_r$ the reference sampling time period. Due to oscillator imperfection, however, an imperfect ADC provides a time-scaled sampling $z[n] = y(T_{\varepsilon} \cdot n)$ with a slightly different sampling time period

$$T_{\varepsilon} = T_r \cdot (1 + \varepsilon), \qquad (1)$$

where the real-valued ε with magnitude $|\varepsilon| \ll 1$ is termed the SRO parameter⁵. Accordingly, the signals y[n] and z[n] are asynchronous and related via

$$z[n] = y(T_r \cdot (n + \tau_{\rm smp}[n])), \qquad (2)$$

where $\tau_{\text{smp}}[n] = \varepsilon \cdot n = \tau_{\text{smp}}[n-1] + \varepsilon$ is an accumulating time drift (ATD) induced by SRO ε .

In frame-based signal processing, an averaged SROinduced ATD is thus observed, i.e.,

$$\tau[\ell] = \varepsilon \cdot n_{\text{mid}}[\ell] = \tau[\ell-1] + \varepsilon \cdot N_s, \qquad (3)$$

where $\ell \ge 1$ is the frame index and $n_{\text{mid}}[\ell] = (N-1)/2 + N_s \cdot \ell$ are the time points on the dimensionless axis t/T_r corresponding to the midpoint of the ℓ -th data frame with frame size N and frame shift N_s .

A linear phase-drift (LPD) model [36, 37] in the STFT domain is then expressed as

$$Z[k,\ell] \approx Y[k,\ell] \cdot e^{j\frac{2\pi}{N}k \cdot \tau[\ell]},\tag{4}$$

where $Y[k, \ell]$ and $Z[k, \ell]$ are the STFT coefficients of y[n] and z[n], respectively, j is the imaginary unit, and $k \in \{0, ..., N-1\}$ denotes a discrete frequency index. According to Eqs. (2), (3), and (4), z[n] is a time-scaled waveform of y[n] corresponding to a time shift between y[n] and z[n] linearly growing with time for fixed SRO $\varepsilon \neq 0$. Note that this constitutes a common assumption, as in reality the SRO varies over time only very little⁶.

⁵ The SRO parameter $\varepsilon \neq 0$ here relates sampling frequencies according to $f_{\varepsilon} = f_{\varepsilon}/(1+\varepsilon) = (1-\varepsilon/(1+\varepsilon)) \cdot f_{\varepsilon} \approx (1-\varepsilon) \cdot f_{\varepsilon}$, if $|\varepsilon| \ll 1$.

⁶ A pairwise waveform-based SRO estimation introduced in this paper is online-capable and therefore can track time varying SRO as shown in [44].

2.2 Waveform-based SRO estimation and compensation

Considering any two acoustic nodes indexed by r and i, the node r is assumed to be the reference node with perfect ADC ($\varepsilon = 0$). In contrast, node i uses an imperfect ADC characterized by the SRO parameter $\varepsilon_{ri} \neq 0$. Waveform-based synchronization (WS) of $z_r[n]$ and $z_i[n]$ consists of SRO estimation and compensation. Using one of the methods for SRO estimation designed for frame-based processing [35–37, 39–42, 44, 45, 51], SRO estimates $\hat{\varepsilon}_{ri}[\ell]$ can be obtained from the observed asynchronous signals $z_r[n]$ and $z_i[n]$.

Next, $\hat{\varepsilon}_{ri}[\ell]$ should be appropriately removed from asynchronous signal $z_i[n]$, leading to an SRO-compensated, synchronized signal $z_{i,S}[n]$, aligned to the reference signal $z_r[n]$ in terms of sampling rate. For this, the realvalued time-variant ATD from (3) can be recursively estimated in every ℓ -th data frame by

$$\widehat{\tau}_{ri}[\ell] = \widehat{\tau}_{ri}[\ell-1] + \widehat{\varepsilon}_{ri}[\ell] \cdot N_s.$$
(5)

Note, Eq. (5) implies that both SRO estimation and compensation are executed at the same frame-rate $f_{\rm WS} = f_r / N_s$. Then, $\hat{\tau}_{ri}[\ell]$ can be compensated in every signal frame by execution of two processing steps: (a) correction of an integer-valued ATD

. . .

$$\widehat{\tau}_{ri}^{\text{int}}[\ell] = round(\widehat{\tau}_{ri}[\ell]) \tag{6}$$

that can be removed from $z_i[n]$ by sample-wise shift of the *i*-th sensor signal, leading to a roughly synchronized signal $z_i[n - \hat{\tau}_{ri}^{int}[\ell]]$ and (b) compensation of a fractional ATD

$$\hat{\tau}_{ri}^{\text{frc}}[\ell] = \hat{\tau}_{ri}[\ell] - \hat{\tau}_{ri}^{\text{int}}[\ell] \tag{7}$$

via resampling of the roughly synchronized signal; see Fig. 3. Various resampling methods can be applied for compensation of fractional ATD [19–23, 36]. Since the STFT resampling method from [36] proved to be a very computationally efficient and sufficiently accurate resampling method⁷, it seems to be an appropriate choice for frame-wise compensation of $\hat{\tau}_{ri}^{\text{frc}}[\ell]$. Thus, the STFT coefficients $Z_{i,S}[k, \ell]$ of a synchronized sensor signal $z_{i,S}[n]$ are obtained by

$$Z_{i,S}[k,\ell] = Z_i^{\text{int}}[k,\ell] \cdot e^{-j\frac{2\pi}{N}k \cdot \hat{\tau}_{ri}^{\text{trc}}[\ell]},\tag{8}$$

where $Z_i^{\text{int}}[k, \ell]$ are the STFT coefficients of the roughly synchronized signal $z_i[n - \hat{\tau}_{ri}^{\text{int}}[\ell]]$. Note that the LPD model (4) is used in (8). Further it should be mentioned



Fig. 3 Two-steps SRO compensation

that the FFT window size can be different for SRO estimation and compensation.

2.3 Closed-loop synchronization of sensor node pairs using internal model control

In order to accomplish a robust waveform-based time synchronization of large acoustic networks by using the subsystems for SRO estimation and compensation described in the previous section, a structural combination of both subsystems to obtain a feasible synchronization unit has to be discussed.

2.3.1 Open-loop synchronization

Retrieval of SRO from asynchronous signals $z_r[n]$ and $z_i[n]$ can lead to estimation with significant bias and uncertainty, where a subsequent SRO compensation can leave an unacceptable synchronization error [40]. In terms of control theory, such a consecutive implementation of the subsystems can be referred to as an open-loop control system depicted in Fig. 4a. A significant disadvantage of such architecture applied for online signal processing is that the SRO estimation is executed on the asynchronous signals with growing ATD between them. Consequently, the requirement of similar frame contents necessary for the LPD model (4) is only fulfilled if the condition $|\tau_{ri}[\ell]| \ll N$ is valid, i.e., as long as the average ATD between $z_r[n]$ and $z_i[n]$ is well within the frame size N [37]. Otherwise, SRO estimation (and also compensation) will collapse with time, making such architecture suitable only for short signal segments or small SROs [36].

2.3.2 Closed-loop synchronization

In offline signal processing, synchronization can be improved by applying the so-called multi-stage procedure with multiple closed-loop iterations of SRO estimation and compensation over the entire signal [40]. This mechanism can be converted into a continuous feedbackcontrol loop comprising a controlled subsystem for SRO compensation followed by an online implementation of SRO estimation as shown in Fig. 4b. Since the subsystem for SRO estimation operates on the synchronized signals, it estimates a current residual SRO $\Delta \hat{\epsilon}_{ri}[\ell]$ between $z_r[n]$ and $z_{i,S}[n]$ after SRO compensation. Thus, the requirement of similar frame content is always fulfilled here. Compared to the open-loop structure, however, such

⁷ For arbitrary sampling rate conversion of narrow-band, speech, and fullband signals, the STFT resampler has been proven to achieve accuracy of 50–60 dB in terms of signal-to-interpolation noise ratio at a very small computational effort in terms of the real-time factor of only 0.005 on average [24].



Fig. 4 Time-synchronization of two nodes in a open-loop and b closed-loop architectures

a closed-loop architecture requires an additional subsystem, a controller that accumulates the residual SRO estimates to the current SRO estimate $\hat{\varepsilon}_{ri}[\ell]$ between asynchronous signals $z_r[n]$ and $z_i[n]$. In the steady state, the system is meant to approach $\Delta \hat{\varepsilon}_{ri}[\ell] \rightarrow 0$ and $\hat{\varepsilon}_{ri}[\ell] \rightarrow \varepsilon_{ri}$. Therefore, since SRO estimation is more precise for smaller SRO values as shown in [50], the closed-loop structure naturally ensures operation of SRO estimation at the optimal working point. In contrast to multi-stage processing, the resulting control architecture merely applies a single treatment of each signal frame, while efficiently diminishing SRO bias and uncertainty with time.

2.3.3 Design of controller based on internal model control (IMC) theory

The controller has to be developed for the frame-based rate f_{WS} of the waveform-synchronization. As a discretetime system, it is designed in the domain of the bilateral *z*-transform, where an impulse response of the controller $g_{\mathbb{C}}[\ell]$ is represented by a system function $G_{\mathbb{C}}(z)$.

From various types of control strategies, we suggest to use a controller based on IMC theory [60, 61], while other designs are possible too. Therefore, an explicit model of the controlled system (plant) is required that consists of SRO compensation and estimation. Abstracting the underlying SRO from the audio signals, we can create a block diagram of the control loop as depicted in Fig. 5a. Here, the function of SRO compensation is described as a subtraction of the estimated SRO $\hat{\varepsilon}_{ri}[\ell]$ from the actual SRO $\varepsilon_{ri}[\ell]$. Furthermore, we suggest to use the DXCP-PhaT method [51] for residual SRO estimation, the dynamical behavior of which is characterized here with $G_{\text{DXCP}}(z)$. Aiming at perfect signal synchronization that would be observed as $\Delta \hat{\varepsilon}_{ri}[\ell] = 0$, the reference control signal $w[\ell]$ is defined as zero. The IMC control circuit implies a plant predictive model leg placed in parallel to the actual plant, where the SRO compensation simplifies to a "-1" multiplier and an approximation $\hat{G}_{\text{DXCP}}(z)$ is used instead of the actual $G_{\text{DXCP}}(z)$. The output difference $\Delta \hat{\varepsilon}_{ri}[\ell] - \Delta \tilde{\varepsilon}_{ri}[\ell]$ feeds back to an IMC filter $G_{\text{IMC}}(z)$. The latter is designed for quadratic minimization of the control error, i.e., the residual SRO signal $\Delta \varepsilon_{ri}[\ell] = \varepsilon_{ri}[\ell] - \hat{\varepsilon}_{ri}[\ell]$, resulting in an optimal IMC filter $G_{\text{IMC}}(z) = -1/G_{\text{DXCP}}(z)$ for ideal approximation $\hat{G}_{\text{DXCP}}(z) = G_{\text{DXCP}}(z)$ [53].

In order to deal properly with feasibility of the control circuit, the optimal solution is extended by a lag element of order n_f (PT_{*n_f*) [62] with filter function}

$$F_{\rm IMC}(z) = \left(\frac{1 - e^{-T_{\rm WS}/T_{\rm IMC}}}{z - e^{-T_{\rm WS}/T_{\rm IMC}}}\right)^{n_{\rm f}},\tag{9}$$

where $T_{\rm WS} = 1/f_{\rm WS}$ is the time shift between STFT frames, $T_{\rm IMC}$ a desired time-constant of $F_{\rm IMC}(z)$ and $n_{\rm IMC}$ the order of $F_{\rm IMC}(z)$. Overall, the IMC filter therefore becomes

$$G_{\rm IMC}(z) = F_{\rm IMC}(z) \cdot G_{\rm IMC}^{\rm opt}(z) = -\frac{F_{\rm IMC}(z)}{G_{\rm DXCP}(z)} .$$
(10)

A sophisticated DXCP-PhaT model $G_{\text{DXCP}}(z)$ as derived in [53] can be simplified regarding model order and complexity of the corresponding IMC controller to a minimum architecture

$$\widehat{G}_{\text{DXCP}}(z) = \frac{1 - \alpha_2}{z - \alpha_2} , \qquad (11)$$



Fig. 5 Equivalent block diagrams of the closed-loop synchronization architecture in the domain of control signals: **a** IMC filter, **b** IMC controller

parameterized by the dominant smoothing constant α_2 of DXCP-internal recursive averaging. The latter is used in DCXP for estimation of a secondary generalized cross-spectral density [51] and is responsible for its dominant time-constant $T_{\text{DXCP}} = T_{\text{WS}}/\ln(1/\alpha_2)$.

Now, the system function of the final IMC-based controller $G_{\rm C}(z)$ in Fig. 5b can be derived as

$$G_{\rm C}(z) = \frac{G_{\rm IMC}(z)}{G_{\rm IMC}(z) \cdot \hat{G}_{\rm DXCP}(z) + 1}$$
(12a)

$$= \frac{F_{\rm IMC}(z)}{F_{\rm IMC}(z) - 1} \cdot \frac{1}{\widehat{G}_{\rm DXCP}(z)} , \qquad (12b)$$

where the architecture in Fig. 5b is an equivalent reorganization of the block diagram in Fig. 5a and the IMC filter from (10) with approximation (11) is used in (12a) for obtaining (12b).

Given the closed-loop synchronization unit Fig. 4b with an embedded DXCP-PhaT method for SRO estimation and the derived IMC-based controller, a gossiping approach for distributed network-wide synchronization can be developed in the next section.

3 Online distributed network-wide synchronization using closed-loop unit

Based on the pairwise synchronization, our concept of a synchronization gossip from [54] is introduced first. A buffer-based implementation of the closed-loop synchronization unit is then described to prepare the appropriate flow of information in the gossip. Finally, a topological organization of WASN by means of a minimum spanning tree is introduced here to support the acoustic connectivity of involved node pairs.

3.1 Concept of synchronization gossip

We consider a WASN with N_{WASN} acoustic sensor nodes labeled with index $i \in \{0, ..., N_{WASN}-1\}$. Among these, a root node r is always defined/chosen to be the global reference node whose sampling rate is equal to the reference sampling rate f_r . In this kind of WASN, at least $N_{WASN}-1$ unknown SROs have to be estimated for a successful network-wide signal synchronization. From graph-theoretical point of view[63], the topology of a WASN can be described as a directed tree denoted as $\vec{T} = (\mathcal{V}, \mathcal{E})$, where the vertex set \mathcal{V} contains N_{WASN} nodes and the edge set \mathcal{E} consists of $N_{WASN}-1$ network links [10, 16, 64]. On such a tree, a network-wide time synchronization can be realized either in a centralized or in a distributed way.

3.1.1 Centralized synchronization

In contributions for waveform-based synchronization with more than two nodes, the centralized synchronization is considered implicitly [38, 42, 44, 45]. For this, all acquired signals are transmitted via a single-hop communication to the root node, where the entire synchronization takes place. The significant drawbacks here are a possible computational overload of the central node in a larger network and a simultaneous requirement of communication bandwidth [28].

3.1.2 Distributed synchronization

Here, on the contrary, the distributed scheme spreads the signal synchronization task over the network so that SRO of every non-reference node is estimated and compensated on the same node where the signal is acquired as it is proposed in publications with time stamp-based synchronization [29, 31]. Significant advantages of such a distributed scheme are the sharing of computational power required for synchronization and the scalability regarding communication bandwidth [10, 30].

3.1.3 Network topologies and their properties

Three particular types of topologies for distributed synchronization are distinguished here: a star tree, a path tree and a rooted tree. Every topology can further be considered with two different edge directions either as an intree (edges oriented to the root) or as an out-tree (edges oriented away the root). Examples of out-tree topologies for $N_{\text{WASN}} = 5$ placed in an isolated shoe-box room are depicted in Fig. 6: star-out-tree (SOT), path-out-tree (POT) and rooted-out-tree (ROT). The root node is highlighted with a bold circle. The direction of edges indicates a one-way out-flow of signals $z_i[n]$ from node *i* along the respective wireless links⁸. Accordingly, every WASN

 $^{^8}$ Note that the ${\cal N}_{WASN}-1$ links of these topologies are the necessary ones for the synchronization task, while other coherent processing of sensor signals may require additional communication links.



Fig. 6 Topologies for network-wide distributed synchronization with $N_{\text{WASN}} = 5$ and root r = 0



Fig. 7 Role-dependent closed-loop architecture unit for network-wide distributed synchronization

node has to be equipped with a digital receiver (RX) and transmitter (TX).

Sensor nodes organized in a certain topology can be characterized by the property of depth (or level). The depth of a node d_i is defined as the length of its path to the root node, which itself has zero depth ($d_r = 0$). The tree depth is given by the depth of its deepest node. In the case of SOT, this tree depth is always one. For some node locations, however, the SOT topology may trail off the acoustic connectivity to the root. In those cases, a multi-hop POT potentially improves upon this problem but does so at the expense of maximizing the tree depth. In many situations, the multi-hop ROT constitutes a compromise between SOT and POT with good acoustic connectivity and intermediate tree depth. Still, the optimal choice of topology generally depends on the actual node locations at hand.

3.1.4 Proposed scheme for distributed synchronization

For waveform-based network-wide synchronization on all distributed topologies, we consider a server-less peerto-peer operation on node pairs, i.e., one sending node providing the reference signal $z_i[n]$ and the receiving node owning the respective non-reference signal $z_i[n]$; see Fig. 7. Moreover, we aim at continuous processing of $z_i[n]$ and $z_i[n]$ on finite buffers, and, hence, their asynchronous generation of data needs to be continuously aligned with an asynchronous resampler in the loop. The closed-loop synchronization unit introduced in Section 2.3 can be efficiently used for such pairwise distributed synchronization. However, the synchronization unit must be configured on every non-reference node in a slightly different manner dependent on the role of the respective node. Specifically, the *i*-th non-reference node is to be configured either as a leaf node (switch position S = 0) or as an intermediate node (switch position S = 1) according to Fig. 7.

In other words, each node receives a local reference signal one-way, either directly from the reference node or from a parent node. Next, the node synchronizes its own microphone signal $z_i[n]$ and provides the synchronized signal $z_{i,S}[n]$ to its children according to the network

topology. By doing so, the signal synchronization is propagated network-wide and uses computational resources of the whole WASN. Naturally, the process of networkwide synchronization will accumulate more latency in deeper networks. The overall duration for the synchronization to propagate from the root node to the deepest leaf node is roughly composed of two contributions: the initialization phase of DXCP-PhaT and its time-constant T_{DXCP} (cf. Section 2.3) multiplied by the tree depth⁹. To accelerate network-wide synchronization, a synchronization gossip on rooted trees with moderate tree depth would thus be favorable.

The proposed network-wide distributed synchronization, however, was initially developed for use in a static WASN in [54], i.e., not considering any dynamic network changes usually occurring in real WASNs.

3.2 Buffer-based realization of closed-loop (online) synchronization unit

Our implementation of closed-loop time synchronization makes use of multiple buffers. A block diagram of the buffer-based time synchronization implemented on the *i*-th sensor node is depicted in Fig. 8a, where the node obtains the global reference signal from the root node rvia a single-hop link (j = r) and thus belongs to the first network level with node depth $d_i = 1$. From the estimated SRO values $\hat{\varepsilon}_{ii}[\ell]$ delivered by the IMC controller, a real-valued ATD estimate is obtained as in (5) under requirement of the same frame shift N_s in both SRO estimation and compensation. However, since both subsystems work on time-domain input signals, the former are allowed to use different frame sizes. The proposed bufferbased implementation of SRO compensation is designed for a frame size equal to the frame shift N_s . Hence, the size of required buffers is a simple multiple of N_s .

⁹ Furthermore, a communication latency between nodes practically needs to be considered, however, its analysis is beyond the scope of this paper.



Fig. 8 Block diagrams of buffer-based synchronization unit on the *i*-th sensor node: **a** single-hop linked to the reference node ($j = r, d_i = 1$); **b** connected only to a local reference ($j \neq r, d_i > 1$)

While the integer-valued ATD $\hat{\tau}_{ji}^{int}[\ell]$ from (6) is compensated using a sliding N_s -long window that is appropriately moved over the resampler buffer, the remaining fractional ATD $\hat{\tau}_{ji}^{frc}[\ell]$ from (7) is removed by applying the STFT resampling method [24, 36] that is also implemented for the frame size N_s . In order to provide for causal resampling, the resampler buffer must introduce at least one frame delay, such that the sliding window (SW) is able to move to the right. We choose a resampler buffer length of 3 frames, where the second frame corresponds to the reference position of the SW. In order to compensate for the resulting delay of one frame, an equivalent delay is applied to the received signal $z_i[n]$ via the delay buffer.

For sensor nodes with a bigger distance to the root node, i.e., $d_i > 1$, the individual delays of buffer-based SRO compensation in preceding levels accumulate and must be compensated using an additional microphone delay buffer (MDB) with a depth-dependent length $L_{\text{MDB}} = d_i$ frames as shown in Fig. 8b. Analogously to the delay buffer, the MDB appends to the sensor-own signal $z_i[n]$ a delay of $d_i - 1$ frames. In other words, the local microphone signal must be passed through the MDB for causal alignment with the delayed reference signal received along the network route.

3.3 Network organization using MST

For accurate waveform-based synchronization, acoustic connectivity between $z_i[n]$ and $z_i[n]$ is essential [55].

Since the connectivity is primarily governed by the distance between nodes, the network topology should generally be configured so as to keep geometric distances between nodes at a minimum.

3.3.1 Minimum spanning tree (MST) as topology

We therefore consider the graph-theoretical MST to maximize acoustic coupling and coherence between node pairs. The MST connects all vertices in a graph without loops and with the minimum possible total edge weight, in this context given by the distance between nodes [65]. As two prominent examples that make use of MST, [66, 67] utilize the concept of MST for route discovery to minimize a total Euclidean distance between nodes for energy-efficient multi-hop communication and networkwide signal enhancement, respectively. In contrast to our previous work in [54], we therefore adopt the MST topology to organize our network. Because algorithms for MST rooting are based on relative sensor positions, we assume that the coordinates of all involved nodes are known up to a certain estimation error¹⁰. In a realistic scenario, such estimates could be provided by dedicated methods for network self-calibration [68-71].

¹⁰ As shown in Section 7.3, the proposed buffer-based synchronization unit can be also successfully used for distributed synchronization of dynamic WASNs organized in simpler topologies without any knowledge of node positions.



Fig. 9 Network organization: a MST, b directed MST with highlighted depth of sensor nodes ($d_i = \{1, 2, 3, 4\}$ for black, blue, green, and red edges)

3.3.2 Optimal choice of the reference node

While the MST is considered as the optimal solution for connecting all nodes with regard to acoustic connectivity, a choice of the reference node r is further required to obtain an actual WASN topology. With the goal of keeping the tree depth as small as possible, we propose to assign this reference role to the node with the smallest average distance to its neighbors. Note that electing the reference node then determines the directions of all edges in the otherwise undirected MST.

An example of an MST in a network with 13 nodes is depicted in Fig. 9a. Identifying node 4 as the the optimal reference node in the aforementioned sense yields the WASN topology shown in Fig. 9b, where the depth of individual nodes is color-coded by means of the corresponding inward edges.

4 Maintaining waveform-based synchronization in dynamic WASNs

A dynamic WASN should be able to adapt to network changes, such as appearance or failure of network nodes or links, without the need to restart the synchronization procedure from scratch, thus avoiding another time-intensive convergence that can cause undesirable degradation in the performance of a network-wide signal processing. This can be achieved by appropriately adapting the network topology in response to observed changes, while *maintaining* an already achieved *waveform synchronization* state (MWS) of persistent nodes that was attained right before the topology change. In this section, we show how the network-wide distributed synchronization presented in Section 3 can be maintained in a dynamic WASN encountering four fundamental types of possible network modifications:

- (a) Appearance of new nodes,
- (b) Failure of communication links,
- (c) Failure of non-reference sensor nodes,
- (d) Failure of the reference sensor node.

For simplicity, we here restrict modifications to one node or communication link at a time and further assume that the WASN has reached a good synchronization state before any such change takes place. This assumption is deemed reasonable, as the initial convergence period takes relatively little time (as we shall see) when considering continued long-term WASN operation¹¹. Furthermore, the type and time-point T_c of a network change are required to be known for the respective treatment. Obtaining this knowledge is credited to the information basis as provided by an address resolution protocol (ARP) [72] or a network discovery protocol (NDP) [73] outside the scope of this paper.

In essence, our strategy for MWS then is to automatically generate an optimal MST network topology for any configuration of nodes and coordinates, respectively, and do so every time a network change occurs. This approach allows us to formulate a mostly universal MWS protocol for handling the various types of changes in the network, requiring only few case specific actions. A summary of the proposed algorithm for operating a dynamic WASN is provided in Algorithm 1 and described in the following line by line.

¹¹ How the proposed method works in the case of an early network change before reaching the initial good synchronization state is shown in Section 7.2.

1:	for frames $\ell = 1 : L$ do
2:	network-wide distributed synchronization by using the
	buffer-based closed-loop DXCP-PhaT proposed in Sec. 3.2;
3:	if any modification in WASN is observed then
4:	update set of nodes & node-coordinates;
5:	find the corresponding Minimum Spanning Tree;
6:	if former reference-node is missing then
7:	find node with the smallest average distance to its
8:	neighbourhood & appoint it as new reference node;
9:	infer directions of MST edges;
10:	for all nodes do
11:	if node changed its depth then
12:	resize & fill MDB appropriately;
13:	if node is a child of newly joined node then
14:	freeze SRO estimation for a freezing time T_f ;
15:	if node is a new reference node then
16:	freeze its SRO estimation permanently;
17:	continue resampling with frozen SRO;

Algorithm 1 MWS protocol for dynamic WASNs

4.1 Network-wide protocol steps (lines 4-9)

To find an optimal network topology, we generate a graph representation of all nodes, where edges are weighted with the Euclidean distance between nodes (line 4). This Euclidean graph is first corrected by removing those edges that correspond to unavailable communication links between nodes. Next, we find the minimum spanning tree via repeated execution of Prim's algorithm [74], considering every node as a possible starting point (line 5), while retaining the choice for global reference if the respective node is still available. If not, a node with the smallest average distance to its neighborhood has to be discovered and appointed as a new reference node (lines 6–8). Finally, based on the reference node, the direction of all edges in the MST are determined (line 9).

4.2 Node-specific protocol steps (lines 11-16)

Because the synchronized signal of each node is systematically delayed in proportion to the level it resides in the network, as discussed in Section 3.2, the MDB of nodes whose depth has changed needs to be resized accordingly (lines 11–12); see Fig. 8b. If a node moves closer to the reference, the MDB size is reduced by discarding the most recent frames. In contrast, the MDB of nodes that moved further away from the reference is increased in size by appending zeros on the right side. In both cases, this mechanism inevitably leads to a small time glitch in the synchronized microphone signals with respect to the local reference signal. This, however, does not negatively impact the SRO estimation process, as will be shown in Section 5.3 below¹². In addition to the adjustment of MDB size and content, the (a) and (d) types of network change require further attention that are detailed in the following.

Change type (a): A node newly integrated into the WASN can usually rely on an already synchronized signal of its topological parent node as a (local) reference and hence synchronize its own microphone signal to it. Until the synchronization is converged, however, its output signal is still asynchronous and should not be utilized as local reference by its topological children nodes. We therefore temporarily freeze the SRO estimation process of any node that directly receives reference from a newly integrated node for a freezing time T_f (lines 13–14). During T_f , the children of a newly integrated node discard a reference signal provided by it and hold their previous SRO estimate¹³.

Change type (d): As mentioned, failure of the reference sensor node requires appointing a new reference. Because this new reference node no longer receives a reference for itself, the previously explained method of freezing the SRO estimate is applied permanently (lines 15–16). By doing so, operation of the WASN can continue seamlessly and without the need to adjust to a significantly different reference sampling rate of the newly elected reference node.

5 Illustration of the proposed mechanism for dynamic WASN operation

In order to demonstrate the methods proposed in Sections 3 and 4 of this paper, we firstly create a synthetic dataset to simulate a WASN with an exemplary topology, which, after initial convergence, is subjected to one network modification of each type. Before examining the resulting effects on distributed SRO estimation for the initial and dynamic WASNs, we first discuss our procedure for generating the synthetic WASN data in a SINS apartment. A large-scale evaluation of the proposed methods is conducted in Section 6.

5.1 WASN simulation in a SINS apartment

With help of Paderbox and PaderWASN toolboxes [44], we simulate¹⁴ a WASN in an artificially generated SINS apartment [58]. In our setup, a total number of 13 nodes, each equipped with a single microphone, are distributed in the apartment. It consists of a living room, a hall, a bedroom, a bathroom, and a toilet. Furthermore, three static acoustic sources (music H4, female speaker N6, male speaker B0) are placed in the living

¹² Moreover, with knowledge of the incident time frame ℓ_c and the sensor depth d_i , if need be, the time glitches in the synchronized signals could be taken into account in further processing of the sensor signals beyond the waveform synchronization (which is not in the scope of this paper).

¹³ Although the SRO estimation process is frozen, the node continues synchronization of its own signal by resampling according to the estimated SRO.

¹⁴ For more details on the implementation of our simulation framework, please refer to https://github.com/STHALabUOL/MWSforDynWASN.

room, all of which are active for almost the total duration of simulated signals of 9 min. The locations of the acoustic sources¹⁵ and all nodes of the acoustic sensor network are depicted in Fig. 10a, where the SINS apartment from Fig. 1 is depicted as shaded background. The room impulse responses between sources and nodes in this simulated environment are provided by the authors of [75] with reverberation time $T_{60} \approx 700 \,\mathrm{ms}$. Node 9 participates in all WASN configurations to provide sufficient acoustic coupling between sensor nodes in the living room and outside¹⁶. The idea of this small-space WASN is a moderate set of proximity nodes with reasonable acoustic coherence and manageable wireless link for sustainable synchronization. Some critical nodes may temporarily leave the network and ideally return with continued synchronization to the momentary reference node (the time for network resynchronization may otherwise be in the order of 1-2 min as shown by Fig. 14 below) and new nodes shall gracefully integrate without disrupting the existing network.

All source signals exhibit a reference sampling rate of $f_r = 16$ kHz. While a music source is downloaded from the Freesound datasets [76], clean speech signals are taken from the LibriSpeech corpus [77]. The resulting microphone signals are superimposed by uncorrelated computer-generated sensor noise of constant power yielding a global signal-to-noise ratio (SNR) of around 33 dB averaged over all sensor nodes; see Fig. 10b. The SROs ε_i of individual nodes are simulated by using an overlap-save method (OSM) for signal resampling [22] with FFT size $N_{\text{OSM}} = 2^{13}$, a frame size of $N_{\text{OSM}}/2$, a frame shift $N_{\text{OSM}}/4$, and a Hann analysis window. The ε_i values are drawn from a uniform distribution on the interval [-100; 100] ppm except for ε_0 , which is set to zero.

The buffer-based closed-loop synchronization unit from Fig. 8 is implemented as described in Section 3.2. The parameters of the DXCP-PhaT, the IMC controller and the STFT resampler are given in Table 1.



Fig. 10 a Setup of synthetic data generation with an initial WASN: root r = 0 and $N_{WASN} = 5$; **b** individual SNRs observed on the sensor nodes; **c** activity pattern of the sources; and **d** SRO estimation in the initial static WASN

5.2 Synchronization in the initial WASN

From the generated WASN environment, an initial WASN with $N_{\text{WASN}} = 5$ nodes is drawn consisting of the nodes {0, 1, 6, 7, 9} as depicted in Fig. 10a. This initial WASN is used to demonstrate the behavior of our MWS protocol proposed in Section 4. While the node r = 0 is chosen as the reference node, the nodes {1, 9} and {6, 7} represent the first and the second rank of network depth, respectively, with SRO

¹⁵ Moving sources are not in the scope of the presented analysis. The typical experience of a moving source is a temporary perturbation of the waveform-based SRO estimation when a specific trajectory induces time-varying time delay (i.e., the equivalent of SRO-based time drift) at the microphones [37]. The precise analysis of the limitations is still an open research topic and the working assumption of spatially fixed acoustic sources is still very common in SRO estimation. Practically, the construction of realistic dynamic acoustic scenes for evaluation is already complicated by the computationally prohibitive simulation of time-varying room impulse responses, whereas the easier case of alternating sources does not impose a major problem [44]. In real-time systems with real signals, we have observed that the estimation will stabilize to a new steady state when the sources halt to a new position.

¹⁶ Further connections between nodes from the living room and other rooms are avoided in MST building to respect their potential acoustic decoupling.

Table 1 Parameters of pairwise synchronization unit

Module	Description	Symbol	Value
DXCP-	Frame length / FFT size	N	2 ¹³
PhaT	FFT shift	Ns	211
	Temporal distance	Lb	39
	Settling parameter	Lc	4
	Maximum lag	Λ	80
	Upsampling factor	р	4
	First smoothing constant	α_1	0.5
	Second smoothing constant	α_2	0.99
IMC-	Time constant	$T_{\rm IMC}$	8 s
controller	Filter order	n _{IMC}	2
STFT-	FFT size	N _{res}	2 ¹²
resampler	FFT shift	N _{res,s}	211

values of $\varepsilon_{ri} = \varepsilon_i - \varepsilon_r = \{20.89, 33.42, 13.44, 61.18\}$ ppm for $i = \{1, 6, 7, 9\}$, respectively.

Figure 10c provides an overview over the acoustic activity for each of the three sources over a limited timespan of 300 s. The first source H4 is playing music in order to provide for continuous acoustic excitation in the background, while the sources N6 and B0 correspond to female and male speakers, respectively, simulating a conversation in the living room.

For the initial WASN, Fig. 10d presents the convergence of SRO estimates after an initialization phase of DXCP-PhaT at the very beginning. The SRO trajectories nodes {1,9} with depth 1 converge rather fast to their target values, depicted by the dashed lines of the respective color. Note that the SRO estimations of nodes {6,7} initially take off in the wrong direction, which is however appropriate with respect to their local parent node 9 during the transitional time period before its settling. The wrong SRO estimations may even overshot according to the time constants of the DXPC-PhaT measurement and the IMC controller and are consistently pulled into the right direction of their target values upon settling of their parent node 9. Overall, the initial WASN then achieves good synchronization state within the first 100 s.

5.3 Dynamic WASN modifications

In order to apply a network modification of each type to the initial WASN of Fig. 10a, we choose the time point $T_c = 200$ s after settling. Specifically, consider

- (a) The appearance of a new sensor node 4
- (b) The failure of link between nodes 6 and 9
- (c) The failure of the non-reference node 7
- (d) The failure of the reference node 0

The modified topologies are depicted in Fig. 11 as a result of the network-wide processing steps of the proposed MWS protocol in Section 4. Taking a closer look at the modified topologies, it is plausible that all of them represent the desired MST under given constraints. Thus, the network topology remains optimal even after the network modification.

Figure 12 shows the SRO estimation of all involved nodes for each network modification type in subfigures using a freezing time $T_f = 100$ s. This value of T_f safely upper bounds the settling time of newly integrated nodes as will be shown in Section 6.2. Figure 12a firstly demonstrates the expected convergence of the newly integrated node 4 to its true SRO with respect to the reference, while the persistent nodes are obviously unaffected by the network modification. Figure 12b, c, and d show that all persistent nodes in case of these network modifications maintain their SRO estimation state, which is especially evident from Fig. 12b, where all nodes remain in the modified WASN. Naturally, in Fig. 12c and d the SRO trajectories of discontinued nodes 7 and 0 disappear for $t > T_c$. Most importantly, application of the proposed protocol avoids a timeconsuming reconvergence in (d).



(a)

(b)

Fig. 11 Modified topologies obtained after executing of the network-wide protocol steps of Algorithm 1 on the initial WASN from Fig. 10a: **a** appearance of a new node 4, **b** failure of communication link between node 6 and node 9, **c** failure of non-reference node 7, **d** failure of reference node 0



Fig. 12 SRO estimation of WASN undergoing modifications from Fig. 11 at $T_c = 200$ s: **a** appearance of a new node 4, **b** failure of communication link between node 6 and 9, **c** failure of non-reference node 7, **d** failure of reference node 0

6 Large-scale evaluation

For large-scale quantitative assessment, we describe the rendering of a richer database of dynamic WASN conditions. Our proposals from Sections 3 and 4 for network-wide SRO estimation and compensation are then evaluated on this data in terms of estimation precision, settling time, and synchronization accuracy.

6.1 Generation of database for dynamic WASN

Using the setup from Section 5.1, we now create random network modifications based on 50 random, unique initial WASN topologies. For the latter, we sample random numbers $N_{\text{WASN}} \in \{4, 5, 6\}$ from the entire set of 13 possible sensor nodes of the simulated WASN environment and construct the MST-optimal topology as described in Section 3.3. To avoid any ill-conditioned links through walls, every node outside the living room connects to node 9, which is included in every topology. The time point of a network change T_c is determined randomly from the interval $T_c \in [250, 290]$ s, such that sufficient simulation time is available for network-wide synchronization before and after the network modification. Network

modifications of each type are then drawn as follows. For modification (a), the new node is sampled from the set of nodes not part of the initial WASN. For modification (b), one of the existing communication links is randomly disabled, however, maintaining the previously described bottleneck-role of node 9. For modification (c), one non-reference node from the initial WASN is randomly selected to be removed. Finally, for modification (d), the global reference node is removed from each initial WASN.

6.2 Network-wide SRO estimation

In order to examine the immediate effect of topology changes including the application of Algorithm 1 on the SRO estimation error of persistent nodes, Fig. 13 specifically compares the root-mean-square error $RMSE_{\varepsilon}$ of SRO within the last 10 s "before" topology changes (left) with that of the first 10 s after topology changes (middle) by boxplots, where one data-point corresponds to one of the initial WASN topologies. We firstly observe that $RMSE_{\varepsilon}$ before T_{c} is very small with a median of only 0.04 ppm. This indicates that all topologies under investigation were given enough time for initial convergence. Moreover, regardless of the specific type of network modification (a)–(d) occurring at T_c , there is no significant increase in the $RMSE_{\varepsilon}$ values observed after T_c . A number of outliers can be noticed, all of which, however, rest safely below a threshold of 1 ppm. Apart from that, the average RMSE_{ε} in (d) appears to be slightly elevated compared to that of all other cases. This is due to the small SRO estimation error of the newly appointed reference node just before $t = T_c$ and it requires the duration of a network settling time T_s after T_c to propagate this slightly new reference sampling rate to all nodes. Overall, the MWS procedure in Algorithm 1 for handling the topology changes is successful in sustaining the SRO estimation accuracy of the persistent nodes.

Figure 13 (right) then shows an extra boxplot of the $RMSE_{\varepsilon}$ of only the newly integrated "joined" nodes based on the last 10 s of the entirely simulated signal. With its overall similar $RMSE_{\varepsilon}$ distribution as compared to the initial convergence "before" topology change, we can once more conclude the successful handling of the related network change.

Figure 14 (left) depicts the corresponding settling time T_s of the SRO estimation, which is here defined as the time period from initial synchronization startup until the temporal $\text{RMSE}_{\varepsilon}(t)$ falls below a threshold $\text{RMSE}_{\varepsilon}(t \ge T_s) \le 1$ ppm. In the diagram, the settling times of all initial WASNs are split by the depth of the involved nodes, which demonstrates a staggered nature of settling according to the synchronization gossip from the root to the leaves. Nodes located closest to the root naturally settle first, as they are directly connected to the



Fig. 13 RMSE_{*e*} values for persistent nodes within last 10 s before T_c (left) and within first 10 s after T_c (middle) and for newly joined nodes within last 10 signal seconds (right)



Fig. 14 Settling times T_s of SRO estimation of the initial WASN split by nodes depths $\in \{1, 2, 3\}$ (left) and of the newly joined nodes (right)

given reference, while deeper nodes still rely on the ongoing settling at intermediate node depths (as illustrated by Fig. 10d). After initial settling of the entire network, any newly "joined" node, irrespective of its corresponding node depth, exhibits the fast settling time with median of about 50 s (right) as found for initial settling at depth $d_i = 1$ (left) also. Of course, the actual settling times are also governed by the actual SRO of each node, which determines the spread of the boxplots. After 100 s, almost all of the newly "joined" nodes have attained synchronization, which determines our choice of the freezing-time parameter T_f in the MWS protocol of Algorithm 1.

6.3 Network-wide signal synchronization

After SRO estimation and evaluation across the network, the related time synchronization of waveforms is eventually assessed in terms of an averaged mean-squared coherence (AMSC) [78] and a signal-to-synchronization-noise ratio¹⁷

$$SSNR = 10 \cdot \log_{10} \frac{\operatorname{Var}(z_{i,r})}{\operatorname{Var}(z_{i,AS} - z_{i,r})},$$
(13)

where $Var(\cdot)$ is an operator for signal variance and the waveform $z_{i,r}[n]$ refers to a synchronous representation of the actual node signal $z_i[n]$ at the sampling rate of the respective reference node r. The signal $z_{i,AS}[n]$ in the *i*-th node is determined by the resampled signal $z_{i,S}[n]$ from Fig. 8, but compensated for a residual time offset

 $\tau_{ri}^{\text{res}}[n] = \sum_{m=1}^{n} (\widehat{\varepsilon}_{ri}[n] - \widehat{\varepsilon}_{ri}[m])$ that accumulates in the closed-loop synchronization unit due to transitional SRO estimation.

Firstly analyzing the initial WASN before a topology change, the resulting AMSC and SSNR values obtained within last 10 s before T_c are presented in Fig. 15 (left). The results confirm poor signal synchronization of the raw asynchronous "async" signals, indicated by a median AMSC of only 0.15 and a median SSNR of about $-3 \, dB$. Outliers at AMSC = 1 do belong to the initial WASNs with node 0 in the role of a non-reference node with $\varepsilon_0 = 0$ ppm, while similar outliers are not visible in the SSNR due to axis limitations. For synchronized "sync" signals, however, the AMSC values appear to be very close to the maximum possible value of 1 and the SSNR assumes a reasonable median of about 12 dB with some variance. The moderate SSNR here is explained by the well-known sensitivity of the SSNR metric with respect to remaining small SRO and timing errors of signals. In summary, these results indicate good WASN synchronization just before the time point of network change T_c .

Then, with dynamic network conditions (a) to (d) according to Section 6.1 and with the application of the MWS protocol of Algorithm 1, the distribution of resulting AMSC and SSNR values obtained on the persistent nodes within first 10 s after T_c are shown in Fig. 15 (middle). As a result of our coordinated treatment of the dynamic conditions, the signal synchronization attained before topology changes well sustains into the phase after the modification for the subset of persistent nodes with a median of 12 to 14 dB SSNR. As shown in Fig. 15 (right),



Fig. 15 Synchronization performance in terms of **a** AMSC and **b** SSNR for persistent nodes within last 10 s before T_c (left) and first 10 s after T_c (middle) and for newly integrated nodes within last 10 signal seconds (right)

¹⁷ Higher values of both AMSC and SSNR values mean better performance.

the remaining subset of newly "joined" nodes evaluated within last 10 signal seconds of the simulation indicates a synchronization comparable to that of persistent nodes, i.e., with very good AMSC values and only a slight loss of SSNR once more being attributed to the sensitivity of this metric to small residual timing errors.

7 Ablation studies

Due to the absence of a reference approach that would operate precisely under the same dynamic network conditions as the proposed methods, this section investigates the requirement of certain processing steps and the robustness to assumptions made. Specifically, the Algorithm 1 for maintaining waveform synchronization is evaluated against several ablated versions of itself in Section 7.1. Then, the former assumption of topology changes after network convergence is abandoned for early network changes taking place in Section 7.2. Eventually, a fallback network configuration to operate without the knowledge of the sensor coordinates and consequently without MST is described in Section 7.3.

7.1 Ablation studies of the proposed method

We investigate the effects on SRO estimation when omitting parts of Algorithm 1, specifically

- (i) When not temporarily freezing children of newly integrated nodes (line 14)
- (ii) When not resizing the MDB of nodes whose depth has changed (line 12) and
- (iii) When not freezing the SRO estimation of newly elected reference nodes (line 16).

In doing so, we rely on the previously introduced network modifications (a) and (d) as shown by Figs. 11 and 12 for which the former simulations are here repeated with ablations but otherwise under the identical conditions as before. Only for a considerable effect of ablation (ii) we have to reduce the DXCP-PhaT frame length to $N = 2^{11}$ to effectively increase the WASN's sensitivity to MDB size mismatches under the limited WASN size and considered time span.

Figure 16 depicts the resulting SRO estimation over time after network modifications at $T_c = 200s$, as before, where ablation (i) is applied with network modification (a), while ablations (ii) and (iii) are applied with network modification (d).

Figure 16 (i) shows the contrast with the former Fig. 12a that the SRO estimation of node 9, as a child of the newly integrated node 4, degrades shortly after T_c and only recovers upon convergence of node 4. Of course, the grandchildren of node 4 (i.e., nodes 6 and 7) are affected, too, although with a delay according to their depth within the MST. As known from Fig. 12a, temporarily freezing



Fig. 16 Ablation studies for Algorithm 1

direct children of newly integrated nodes would alleviate this problem.

Figure 16 (ii) refers to a dynamic modification with a new reference node 9. Since the depth relationship of nodes 6, 7, and 9 remains unchanged in this very example, their SRO estimation is apparently stable with time. However, node 1 severely degrades at about 75 s after the change time T_c due to a modified depth relationship between node 1 and node 9 (formerly relayed by node 0) and with the corresponding mismatch of MDBs not resized properly, which eventually violates the assumption of similar content of the input signals for waveform-based SRO estimation.

Figure 16 (iii) finally depicts a contrast with former Fig. 12d when resampling of the newly elected reference node 9 is somewhat naively discontinued, which corresponds to resetting its SRO estimation to zero (instead of continued resampling with frozen SRO estimation according to Algorithm 1). Hence, all descendants of the new reference node (the entire WASN) are required to adjust to the new reference condition by reconvergence, which temporarily and unnecessarily presents an undetermined state of the sensor network.

7.2 Dynamic WASN with early network change

For clarity of the arrangement, a steady-state synchronization was assumed in Section 5.3 before any network change takes place and is being coordinated by the proposed MWS protocol in Algorithm 1. The steady-state assumption there was inherently reasonable, since it is less interesting to maintain the state of a WASN if its nodes have not yet converged. However, in practice an early network change may arise before convergence and the intention now is to show that convergence is not a strict requirement for the employment of the proposed methods. Figure 17 therefore considers a network modification (a), a newly integrated node 4 before convergence of the initial network. It turns out that the early network change does not cause any permanent complications when Algorithm 1 is applied. The acoustic sensor network only needs more time to settle (here around 2.5 min after the change) compared to the idealized case from Fig. 12a (only 1 min for new settling after the change).

7.3 Distributed buffer-based synchronization

without knowledge of positions of the sensor nodes In this section, the performance of the buffer-based online realization of distributed WASN synchronization from Section 3.2 is investigated for the case if no knowledge of the node coordinates is available. In such a scenario, it is neither possible to build the geometric MST topology nor to optimally choose the reference node as described in Section 3.3. Instead, we may fall back to a centralized SOT topology mentioned in Section 3.1 with a fixed or randomly chosen reference node. For the analysis, we rely on sampled sets of nodes as described in Section 6.1 and evaluate WASN performance attained in the steady state between time 190 and 200 secs (the same time span as used for previous Figs. 13 (left) and 15 (left) with MST).

Figure 18 summarizes the outcomes, where "MST" stands as an anchor for the previous results, "SOT" refers to star-out topology with node 9 always the reference, and "rSOT" instead uses a random reference node (newly sampled without special treatment of node 9). With the metrics at hand we do observe similar network performances for all configurations, with maybe marginally reduced RMSE of the SRO estimation and slightly advanced synchronization SSNR for the SOT topologies. This can be attributed to the minimum network depth of the SOT and thus an earlier and slightly better network convergence in the available simulation time, while the larger geometric distances between nodes connected



Fig. 17 Early network change for the network modification (a) from Fig. 11a



with and without knowledge of positions of the sensor nodes: **a** RMSE values of SRO estimation, **b** SSNR values of SRO compensation

along topology edges do not significantly impair the acoustic coupling in our small-space SINS environment.

In light of this ablation, the MST topology indeed has not proved superior in our simulated context, but we do see the reason in our relatively small-scale configuration and in the simulation of low-noise microphones. Conversely, we still see the necessity of local operation organized in an MST configuration (rather than SOT) when considering larger scenarios or use cases with increased requirements as of

- A lower acoustic coherence between distant nodes,
- An increased noise floor of low-cost microphones,
- A limited wireless connectivity of distant nodes,
- Larger number of sensor nodes in the network,
- And a necessary decentralization for network robustness or distribution of computational load.

These requirements may appear in crowded indoor networks with numerous sensors or in large-scale outdoor networks, for instance, biosphere monitoring. It turns out that such immense diversity of WASNs has not been represented in our analysis of small-scale configurations yet. Still, it was our intention to demonstrate the utility of proposed methods, including the closed-loop synchronization unit and the dynamic MWS protocol, under several circumstances.

8 Conclusions

An online distributed waveform-based sampling-time synchronization for dynamic wireless acoustic sensor networks (WASNs) has been described in this paper and applied to a simulated smart home environment for evaluation. The essential system component is a buffer-based implementation of a closed-loop synchronization unit (with resampling and sampling-rate offset estimation in a loop) for any two nodes of the network. Our specific unit makes use of a double-cross-correlation processor for waveform-based estimation of sampling rate offset (SRO) and of a buffer-based SRO compensation by an STFT-based resampling method. This estimation and compensation in the closed-loop architecture are here coupled by an internal-model-control unit. The suggested pairwise node synchronization unit is then employed for distributed synchronization of WASNs organized in a rooted-tree topology with minimum spanning tree. Our paper has demonstrated how the synchronization gossip in this case propagates from the root to the leaves of the network. Eventually, a protocol for maintaining waveform-based synchronization has been proposed for scenarios with random modifications of the original WASN taking place. Our experimental evaluation in the environment of a simulated apartment with several connected rooms proved efficiency and robustness of the proposed system (for instance, against unknown sensor coordinates, early modification, and some of the ablations studied) for sustainable network-wide SRO estimation and signal synchronization in dynamic WASNs.

Abbreviations

ADC	Analog-to-digital converter
AMSC	Averaged mean-squared coherence
ATD	Accumulating time drift
DXCP-PhaT	Double cross-correlation processor with phase transform
IMC	Internal model control
MDB	Microphone delay buffer
MST	Minimum spanning tree
MWS	Maintaining of waveform-based synchronization
SINS	Sound Interface to the Swarm
SRO	Sampling rate offset
SSNR	Signal-to-synchronization noise ratio
WASN	Wireless acoustic sensor network
WS	Waveform-based synchronization

Acknowledgements

The authors would like to thank the research unit DFG FOR 2457 "Acoustic Sensor Networks" (https://asn.uni-paderborn.de/) for diverse collaboration.

Authors' contributions

A.C. conceptualized the publication and coordinated the implementation. N.K. implemented the proposed system in Python and performed the experimental evaluation. A.C. and N.K. wrote the original draft of the manuscript. G.E. contributed the key ideas, supervised the work, and revised the article. All authors reviewed and approved the final manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL. This work was partially supported by German Research Foundation (DFG) - Project 282835863.

Availability of data and materials

Download data links and source code examples of our simulation framework are available at https://github.com/STHALabUOL/MWSforDynWASN.

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 29 March 2023 Accepted: 6 October 2023 Published online: 18 December 2023

References

- K. Sohraby, D. Minoli, T. Znati, Wireless sensor networks: technology, protocols, and applications (John Wiley & Sons, New Jersey, USA, 2007)
- V.C. Güngör, G.P. Hancke, Industrial wireless sensor networks: applications, protocols, and standards (CRC Press of Taylor & Francis Group, Boca Raton, 2013)
- S. Khan, A.-S.K. Pathan, N.A. Alrajeh, Wireless sensor networks: current status and future trends (CRC Press of Taylor & Francis Group, Boca Raton, 2016)
- M. Elhoseny, A.E. Hassanien, Dynamic wireless sensor networks, vol. 165 (Springer, Cham, 2019)
- 5. X. Chen, Randomly deployed wireless sensor networks (Elsevier, Amsterdam, 2020)
- 6. H.M. Ammari, *Theory and practice of wireless sensor networks*, vol. 214 (Springer, Cham, 2022)
- I.F. Akyildiz, T. Melodia, K.R. Chowdhury, A survey on wireless multimedia sensor networks. Comput. Netw. 51(4), 921–960 (2007)
- A. Bertrand, S. Doclo, S. Gannot, N. Ono, T. van Waterschoot, Special issue on wireless acoustic sensor networks and ad hoc microphone arrays. Signal Process. Elsevier 107-C, 1–3 (2015)
- G. Ciccarelli, J. Barber, A. Nair, I. Cohen, T. Zhang, Challenges and opportunities in multi-device speech processing. arXiv:2206.15432. 1–5 (2022)
- A. Bertrand, in *Proc. IEEE Symp. Commun. Veh. Technol.* Applications and trends in wireless acoustic sensor networks: a signal processing perspective (IEEE, Ghent, 2011), pp. 1–6
- R. Lienhart, I. Kozintsev, S. Wehr, M. Yeung, in *Proc. IEEE Int. Conf. Acoust.,* Speech, Signal Process. On the importance of exact synchronization for distributed audio signal processing, vol. 4 (IEEE, Hong Kong, 2003), pp. 840–843
- S. Wehr, I. Kozintsev, R. Lienhart, W. Kellermann, in *IEEE Int. Symp. Multimedia* Softw. Eng. Synchronization of acoustic sensors for distributed ad-hoc audio networks and its use for blind source separation (IEEE, Miami, 2004), pp. 18–25
- P. Didier, T. Van Waterschoot, S. Doclo, M. Moonen, Sampling rate offset estimation and compensation for distributed adaptive node-specific signal estimation in wireless acoustic sensor networks. IEEE Open J. Signal Process. 4, 71–79 (2023)
- M. Guggenberger, M. Lux, L. Böszörmenyi, in *Proc. Int. Conf. on Multimedia Modeling*. An analysis of time drift in hand-held recording devices (Springer International Publishing, Sydney, 2015), pp. 203–213
- J. Schmalenstroeer, T. Gburrek, R. Haeb-Umbach, LibriWASN: a data set for meeting separation, diarization, and recognition with asynchronous recording devices. arXiv:2308.10682. 1–5 (2023)
- R. Olfati-Saber, R.M. Murray, Consensus problems in networks of agents with switching topology and time-delays. IEEE Trans. Autom. Control. 49(9), 1520–1533 (2004)
- Y. Zeng, R.C. Hendriks, N.D. Gaubitch, in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.* On clock synchronization for multi-microphone speech processing in wireless acoustic sensor networks (IEEE, Brisbane, 2015), pp. 231–235
- I. Stojmenovic, Handbook of Sensor Networks: Algorithms and Architectures, vol. 49 (John Wiley & Sons, New Jersey, 2005)
- R.E. Crochiere, L.R. Rabiner, *Multirate* digital signal processing (Prentice Hall, New Jersey, 1983)
- 20. J.G. Proakis, D.G. Manolakis, *Digital signal processing: princ*iples (algorithms and applications. Prentice-Hall Int. Corp, New Jersey, 1996)
- A.V. Oppenheim, R.W. Schafer, *Discrete-time signal processing* (Prentice Hall, New Jersey, 1999)
- 22. J. Schmalenstroeer, R. Haeb-Umbach, in *Proc. Eur. Signal Process. Conf.* Efficient sampling rate offset compensation - an Overlap-Save based approach (EURASIP, Rome, 2018), pp. 499–503
- A. Chinaev, P. Thuene, G. Enzner, in *Proc. Eur. Signal Process. Conf.* Low-rate Farrow structure with discrete-lowpass and polynomial support for audio resampling (EURASIP, Rome, 2018), pp. 475–479
- A. Chinaev, G. Enzner, J. Schmalenstroeer, in *Proc. ITG Conf. Speech Com*mun. Fast and accurate audio resampling for acoustic sensor networks by polyphase-Farrow filters with FFT realization (VDE VERLAG GmbH, Berlin/ Offenbach, 2018), pp. 96–100
- H. Karl, A. Willig, Protocols and architectures for wireless sensor networks (John Wiley & Sons, West Sussex, 2007)
- J. Elson, L. Girod, D. Estrin, Fine-grained network time synchronization using reference broadcasts. ACM SIGOPS Oper. Syst. Rev. 36(SI), 147–163 (2002)
- Y.W. Hong, A. Scaglione, A scalable synchronization protocol for large scale sensor networks and its applications. IEEE J. Sel. Areas Commun. 23(5), 1085–1099 (2005)

- L. Schenato, F. Fiorentin, Average TimeSynch: A consensus-based protocol for clock synchronization in wireless sensor networks. Automatica 47(9), 1878–1886 (2011)
- J. Du, Y.C. Wu, Distributed clock skew and offset estimation in wireless sensor networks: asynchronous algorithm and convergence analysis. IEEE Trans. Wirel. Commun. **12**(11), 5908–5917 (2013)
- Y. Qiao, W. Yang, M. Fu, in *Proc. Chinese Control Conf.* A new power-efficient distributed method for clock synchronization in sensor networks (IEEE, Chengdu, 2016), pp. 7572–7577
- J. Schmalenstroeer, P. Jebramcik, R. Haeb-Umbach, A combined hardware-software approach for acoustic sensor network synchronization. Signal Process. 107-C, 171–184 (2015)
- S. Ganeriwal, R. Kumar, M.B. Srivastava, in *Proc. Int. Conf. on Embedded Networked Sensor Systems*. Timing-sync protocol for sensor networks (ACM, New York, 2003), pp. 138–149
- W. Su, I.F. Akyildiz, Time-diffusion synchronization protocol for wireless sensor networks. IEEE/ACM Trans. Netw. 13(2), 384–397 (2005)
- Z. Liu, in Proc. Int. Workshop on Acoustic Echo and Noise Control. Sound source separation with distributed microphone arrays in the presence of clock synchronization errors (Inderscience Enterprises Ltd, Geneva, 2008), pp. 1–4
- S. Markovich-Golan, S. Gannot, I. Cohen, in *Proc. Int. Workshop Acoust. Signal Enhancement*. Blind sampling rate offset estimation and compensation in wireless acoustic sensor networks with application to beamforming (VDE, Aachen, 2012), pp. 1–4
- S. Miyabe, N. Ono, S. Makino, Blind compensation of interchannel sampling frequency mismatch for ad hoc microphone array based on maximum likelihood estimation. Signal Process. Elsevier 107-C, 185–196 (2015)
- L. Wang, S. Doclo, Correlation maximization-based sampling rate offset estimation for distributed microphone arrays. IEEE/ACM Trans. Audio Speech Lang. Process. 24(3), 571–582 (2016)
- D. Cherkassky, S. Gannot, Blind synchronization in wireless acoustic sensor networks. IEEE/ACM Trans. Audio Speech Lang. Process. 25(3), 651–661 (2017)
- M.H. Bahari, A. Bertrand, M. Moonen, Blind sampling rate offset estimation for wireless acoustic sensor networks through weighted least-squares coherence drift estimation. IEEE/ACM Trans. Audio Speech Lang. Process. 25(3), 674–686 (2017)
- J. Schmalenstroeer, J. Heymann, L. Drude, C. Boeddecker, R. Haeb-Umbach, in *Proc. IEEE Int. Workshop Multimedia Signal Process.* Multi-stage coherence drift based sampling rate synchronization for acoustic beamforming (IEEE, London, 2017), pp. 1–6
- S. Araki, N. Ono, K. Kinoshita, M. Delcroix, in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.* Estimation of sampling frequency mismatch between distributed asynchronous microphones under existence of source movements with stationary time periods detection (IEEE, Brighton, 2019), pp. 785–789
- K. Itoyama, K. Nakadai, in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Syst.* Synchronization of microphones based on rank minimization of warped spectrum for asynchronous distributed recording (IEEE, Las Vegas, 2020), pp. 4842–4847
- K. Yamaoka, N. Ono, Y. Wakabayashi, in *Proc. Eur. Signal Process. Conf.* Sampling frequency mismatch estimation by auxiliary-function-based iterative maximization of double-cross-correlation (EURASIP, Dublin, 2021), pp. 1125–1129
- T. Gburrek, J. Schmalenstroeer, R. Haeb-Umbach, in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process.* On synchronization of wireless acoustic sensor networks in the presence of time-varying sampling rate offsets and speaker changes (IEEE, Singapore, 2022), pp. 916–920
- Y. Masuyama, K. Yamaoka, N. Ono, Joint optimization of sampling rate offsets based on entire signal relationship among distributed microphones. arXiv:2206.13014. (2022)
- R. Wang, Z. Chen, F. Yin, Active sampling rate calibration method for acoustic sensor networks. IEEE/ACM Trans. Audio Speech Lang. Process. 28, 3095–3107 (2020)
- D. Hu, H. Zhang, F. Bao, R. Wang, Distributed sampling rate offset estimation over acoustic sensor networks based on asynchronous network newton optimization. IEEE/ACM Trans. Audio Speech Lang. Process. 31, 301–312 (2023)
- M. Pawig, G. Enzner, P. Vary, Adaptive sampling rate correction for acoustic echo control in voice-over-ip. IEEE Trans. Signal Process. 58(1), 189–199 (2010)
- P. Thüne, G. Enzner, in *Proc. of Eur. Signal Process. Conf.* Tracking theory of adaptive filters with input-output sampling rate offset (EURASIP, Corunna, 2019), pp. 1–5

- A. Chinaev, P. Thüne, G. Enzner, in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.* A double-cross-correlation processor for blind sampling rate offset estimation in acoustic sensor networks (IEEE, Brighton, 2019), pp. 641–645
- A. Chinaev, P. Thüne, G. Enzner, Double-cross-correlation processing for blind sampling-rate and time-offset estimation. IEEE/ACM Trans. Audio Speech Lang. Proces. 29, 1881–1896 (2021)
- A. Chinaev, G. Enzner, T. Gburrek, J. Schmalenstroeer, in *Proc. Eur. Signal* Process. Conf. Online estimation of sampling rate offsets in wireless acoustic sensor networks with packet loss (EURASIP, Dublin, 2021), pp. 1110–1114
- A. Chinaev, S. Wienand, G. Enzner, in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.* Control architecture of the double-cross-correlation processor for sampling-rate-offset estimation in acoustic sensor networks (IEEE, Toronto, 2021), pp. 801–805
- A. Chinaev, G. Enzner, in Proc. Int. Workshop Acoust. Signal Enhancement. Distributed synchronization for ad-hoc acoustic sensor networks using closed-loop double-cross-correlation processing (IEEE, Bamberg, 2022), pp. 1–5
- A. Chinaev, N. Knaepper, G. Enzner, in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.* Long-term synchronization of wireless acoustic sensor networks with nonpersistent acoustic activity using coherence state (IEEE, Rhodes, 2023), pp. 1–5
- H. Afifi, S. Auroux, H. Karl, in *Proc. IEEE Wireless Commun. and Networking* Conf. MARVELO: wireless virtual network embedding for overlay graphs with loops (Barcelona, 2018)
- H. Afifi, J. Schmalenstroeer, J. Ullmann, R. Haeb-Umbach, H. Karl, in *Proc. ITG Symp. on Speech Commun.* MARVELO - a framework for signal processing in wireless acoustic sensor networks (VDE VERLAG GmbH, Berlin/ Offenbach, 2018), pp. 311–315
- G. Dekkers, S. Lauwereins, B. Thoen, M.W. Adhana, H. Brouckxon, B. Van den Bergh, T. van Waterschoot, B. Vanrumste, M. Verhelst, P. Karsmakers, in *Proc. Workshop on Detection and Classification of Acoustic Scenes and Events*. The SINS database for detection of daily activities in a home environment using an acoustic sensor network (Tampere University of Technology, Tampere, 2017), pp. 1–5
- A. Nelus, R. Glitza, R. Martin, in *Proc. Eur. Signal Process. Conf.* Unsupervised clustered federated learning in complex multi-source acoustic environments (EURASIP, Dublin, 2021), pp. 1115–1119
- 60. B. Francis, W. Wonham, The internal model principle of control theory. Automatica **12**, 457–465 (1976)
- 61. M. Morari, E. Zafiriou, *Robust process control* (Prentice Hall, New Jersey, 1989)
- 62. J. Lunze, Regelungstechnik 1: System theoretische Grundlagen, Analyse und Entwurf Einschleifiger Regelungen, 10th edn. (Springer Vieweg, Berlin, 2014)
- D.B. West, Introduction to Graph Theory, vol. 2, 2nd edn. (Pearson Education, Inc., Delhi, 2001)
- J. Lunze, Networked control of multi-agent systems: consensus and synchronisation, communication structure design, self-organisation in networked systems (Event-triggered Control. De Gruyter, Berlin, 2019)
- R.L. Graham, P. Hell, On the history of the minimum spanning tree problem. Ann. Hist. Comput. 7(1), 43–57 (1985)
- 66. M. Saravanan, M. Madheswaran, A hybrid optimized weighted minimum spanning tree for the shortest intrapath selection in wireless sensor network. Hindawi Math. Probl. Eng. 2014, 1–8 (2014) https://www.hinda wi.com/journals/mpe/2014/713427/
- J. Szurley, A. Bertrand, M. Moonen, Topology-independent distributed adaptive node-specific signal estimation in wireless sensor networks. IEEE Trans. Signal Inform. Process. Over Netw. 3(1), 130–144 (2016)
- V.C. Raykar, I.V. Kozintsev, R. Lienhart, Position calibration of microphones and loudspeakers in distributed computing platforms. IEEE Trans. Speech Audio Process. 13(1), 70–83 (2004)
- M. Parviainen, P. Pertilä, M.S. Hämäläinen, in *Proc. IEEE Joint Workshop on* Hands-free Speech Comm. and Microphone Arrays. Self-localization of wireless acoustic sensors in meeting rooms (IEEE, Nancy 2014), pp. 152–156
- L. Wang, T. Hon, J.D. Reiss, A. Cavallaro, Self-localization of ad-hoc arrays using time difference of arrivals. IEEE Trans. Signal Process. 64(4), 1018–1033 (2016)
- T. Gburrek, J. Schmalenstroeer, R. Haeb-Umbach, Geometry calibration in wireless acoustic sensor networks utilizing DoA and distance information. EURASIP J. Audio Speech Music Process. 2021(1), 25 (2021)
- D. Plummer, An ethernet address resolution protocol: or converting network protocol addresses to 48. bit ethernet address for transmission on ethernet hardware. Technical report (1982)

- T. Narten, E. Nordmark, W. Simpson, H. Soliman, Neighbor discovery for IP version 6 (IPv6). Technical report (2007)
- R.C. Prim, Shortest connection networks and some generalizations. Bell Syst. Tech. J. 36(6), 1389–1401 (1957)
- R. Glitza, L. Becker, R. Martin, in *Proc. Europ. Signal Process. Conf.* Database of simulated room impulse responses for acoustic sensor networks deployed in complex multi-source acoustic environments (EURASIP, Helsinki, 2023)
- E. Fonseca, J. Pons Puig, X. Favory, F. Font Corbera, D. Bogdanov, A. Ferraro, S. Oramas, A. Porter, X. Serra, in *Proc. Int. Soc., Music Inform. Retrieval Conf.* Freesound datasets: a platform for the creation of open audio datasets (TISMIR, Suzhou, 2017), pp. 486–493
- V. Panayotov, G. Chen, D. Povey, S. Khudanpur, in *Proc. IEEE Int. Conf. on Acoust., Speech, Signal, Process.* Librispeech: an ASR corpus based on public domain audio books (IEEE, Brisbane, 2015), pp. 5206–5210
- M. Jeub, C. Nelke, C. Beaugeant, P. Vary, in *Proc. Eur. Signal Process. Conf.* Blind estimation of the coherent-to-diffuse energy ratio from noisy speech signals (EURASIP, Barcelona, 2011), pp. 1347–1351

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[™] journal and benefit from:

- Convenient online submission
- ► Rigorous peer review
- Open access: articles freely available online
- ► High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com