SAMPLING RATE SYNCHRONISATION IN ACOUSTIC SENSOR NETWORKS WITH A PRE-TRAINED CLOCK SKEW ERROR MODEL

Joerg Schmalenstroeer, Reinhold Haeb-Umbach

Department of Communications Engineering, University of Paderborn, Germany

{schmalen, haeb}@nt.uni-paderborn.de

ABSTRACT

In this paper we present a combined hardware/software approach for synchronizing the sampling clocks of an acoustic sensor network. A first clock frequency offset estimate is obtained by a time stamp exchange protocol with a low data rate and computational requirements. The estimate is then postprocessed by a Kalman filter which exploits the specific properties of the statistics of the frequency offset estimation error. In long term experiments the deviation between the sampling oscillators of two sensor nodes never exceeded half a sample with a wired and with a wireless link between the nodes. The achieved precision enables the estimation of time difference of arrival values across different hardware devices without sharing a common sampling hardware.

Index Terms: synchronization, acoustic sensor network

1. INTRODUCTION

Wireless sensor networks have been an active field of research for many years [1]. The sensor networks have found numerous application areas including environmental, medical, military, transportation and entertainment systems, as well as smart spaces. Recently, acoustic sensor networks have gained increased interest, where a network of distributed microphones is employed for, e.g., surveillance tasks [2]. A particular focus of research is collaborative signal processing, where the signal processing task is distributed over the sensor nodes, each node communicating only with its neighborhood, while still maximizing a common objective function. Compared to centralized processing this has a couple of advantages, such as reduced communications bandwidth and the avoidance of a single point of failure. An example is the distributed beamforming algorithm described in [3].

Many cooperative acoustic signal processing tasks, however, require that the sampling clocks of the distributed microphones are synchronized. Consider for example the use of the GCC-PHAT algorithm [4] for speaker localization, which requires the estimation of the time difference of arrival between microphone signals. Let us assume we are given two microphones at a distance of $0.2\,\mathrm{m}$, which are sampled at $16\,\mathrm{kHz}$ sampling rate. If the input signal comes from the endfire position, it takes $0.2\,\mathrm{m}/(340\,\mathrm{m/s}) * 16.000\,\mathrm{Hz} = 9.4$ sampling intervals to travel from the first to the second microphone. Now assume that the two micropones are attached to different hardware devices, whose sampling frequencies have a clock frequency offset of $50\,\mathrm{ppm}$. Thus, if the first oscillator has a frequency of $16\,\mathrm{kHz}$, the second may have one of $16\,000.8\,\mathrm{Hz}$. Then it takes $9.4/(0.8\,\mathrm{Hz}) = 11.75$ seconds until the two oscillators are 9.4 sampling intervals apart. Thus, if the signal is arriving from the endfire position, after $11.75\,\mathrm{s}$ the signal is perceived to be arriving from broadside! Due to the oscillator offset it thus seems that the signal source is moving at an angular velocity of 90° per $11.7\,\mathrm{s}$, i.e. at 7.66° per second!

Thus there is a clear need to compensate for sampling rate offsets in acoustic signal processing tasks. Pawig, Enzner and Vary showed that even small deviations between D/A and A/D clocks result in a time-variant acoustic impulse response as seen by the echo canceler [5]. In [6] a blind sampling rate offset estimator is derived in the context of acoustic beamforming. It is based on the phase drift of the coherence of the two microphone signals. After offset estimation the signal is resampled for clock rate adjustment.

In this paper we take a different avenue. We assume that the acoustic sensors are connected to a wireless or wired communications network – which is a reasonable assumption, as the sensor data need to be forwarded to other nodes or a central processing device. We then take advantage of the large literature on synchronisation of communication networks [7] to align the A/D oscillators via the network.

Several algorithms have been proposed which rely on the exchange of time stamps between the nodes. For example, the algorithm of [8] achieves good synchronisation properties at low computational effort. To further improve precision, clock skew estimates can be further processed. The authors of [9] for example distribute a Kalman filter across all sensor nodes and each node use the state estimates of other nodes to synchronize with a virtual master clock. Advantageous is the lack of a single point of failure and the robustness against packet losses within the IEEE 802.15.4 (ZigBee-PHY) wireless network. In [10] a IGMKPF called approach was proposed which utilizes a computationally demanding combination of particle and Kalman filter which iteratively estimates a Gaussian mixture model for the observation error.

Collaborative acoustic signal processing requires particularly precise clock synchronisation. In the approach presented here, we employ a simplified version of the time stamp exchange method of [8], to save computational effort at little loss in precision. The obtained initial clock frequency offset estimates are further processed in a dedicated Kalman Filter. The filter exploits the specific properties of the clock skew estimates when the nodes are connected via a wireless sensor network. Subsequently, the Kalman filter state estimate is used to readjust the sampling frequency of one node with the purpose of bringing the sampling frequencies of both nodes in line.

The paper is organized as follows: In the next section we briefly discuss the clock frequency offset estimation via the time stamp exchange protocol of [8], and we present our modifications to it. In Section 3 we show some experimental results regarding the observation error which were the motivation for the specific design of the Kalman filter presented in Section 4. The experimental results in Section 5 starts with a description of our hardware platform, followed by measurement results for a wired and wireless link between two sensor nodes. The paper finishes off with conclusions drawn in Section 6.

2. CLOCK FREQUENCY OFFSET ESTIMATION

The clock frequency offset estimation is based upon the approach from [8] which we shortly describe in the following. Let us assume that the sensor network has two nodes A and B which periodically exchange time stamps. The node B is selected as the master defining its clock to be the reference clock.

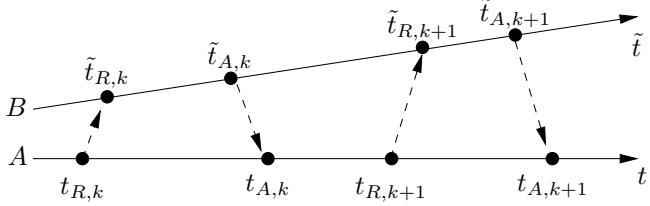

Fig. 1. Development of local time at nodes A and B with phase and frequency offset during time stamp exchange

In Fig. 1 the k-th and the $(k+1)$-th time stamp exchange is depicted. The slope of the time line $\tilde{t}$ of node B indicates its clock frequency offset compared to node A.

Node A triggers the time stamp exchange by sending a request at time instance $t_{R,k}$ which is received at the local time instance $\tilde{t}_{R,k}$ at node B. The answers to the request from B at time instance $\tilde{t}_{A,k}$ is received at $t_{A,k}$ at node A. It follows that

$$\tilde{t}_{R,k} = (t_{R,k} + \xi_{R,k}) \cdot \omega + \varphi \tag{1}$$

$$\tilde{t}_{A,k} = (t_{A,k} - \xi_{A,k}) \cdot \omega + \varphi \tag{2}$$

where $\xi_{R,k}$ and $\xi_{A,k}$ are the transmission times of the time stamp packets, ω is the clock frequency offset, also called clock skew, between the two oscillators and φ is the phase offset.

According to [8] the clock frequency offset can be estimated from the k-th and l-th time stamp exchange by computing:

$$\frac{\tilde{t}_{R,k} - \tilde{t}_{A,l}}{t_{R,k} - t_{A,l}} = \left(1 + \frac{\xi_{R,k} + \xi_{A,l}}{t_{R,k} - t_{A,l}}\right) \omega. \tag{3}$$

From eq. (3) it is obvious that the influence of the transmission times $\xi_{R,k}$ and $\xi_{A,l}$ vanishes with an increasing temporal distance between $t_{R,k}$ and $t_{A,l}$.

Depending on the chosen indices k and l the term $(\xi_{R,k} + \xi_{A,l})/(t_{R,k} - t_{A,l})$ can be positive or negative. In [8] the author proposes to use a minimum and a maximum search across all combinations of time stamp transmissions to find those indices, for which the term assumes either the smallest positive or the largest negative value. In informal experiments we found that the influence of the temporal distance between the time stamp exchanges is very large such that the min/max search usually selects the time stamps with the largest temporal distance. We therefore decided to transmit time stamp pairs only occasionally, where the choice of temporal distance between two successive exchanges k and $k+1$ must strike a balance between the contradicting requirement of having a large distance to have the right-hand side of (3) as close to ω as possible, and the requirement to track the time variant clock skew. By this both the computational effort and the data rate for time stamp exchange could be significantly reduced compared to [8] with only marginal impact on the quality of the estimate.

A clock frequency offset estimate is obtained from two consecutive time stamp exchanges by first computing

$$\frac{\Delta\tilde{t}^{+}}{\Delta t^{+}} = \frac{\tilde{t}_{R,k+1} - \tilde{t}_{A,k}}{t_{R,k+1} - t_{A,k}} = \left(1 + \frac{\xi_{R,k+1} + \xi_{A,k}}{t_{R,k+1} - t_{A,k}}\right) \omega \tag{4}$$

$$\frac{\Delta\tilde{t}^{-}}{\Delta t^{-}} = \frac{\tilde{t}_{R,k} - \tilde{t}_{A,k+1}}{t_{R,k} - t_{A,k+1}} = \left(1 - \frac{\xi_{R,k} + \xi_{A,k+1}}{|t_{R,k} - t_{A,k+1}|}\right) \omega. \tag{5}$$

An estimate of the clock frequency offset is now obtained by

$$\widehat{\omega} = \frac{\Delta\tilde{t}^{+} - \Delta\tilde{t}^{-}}{\Delta t^{+} - \Delta t^{-}} \tag{6}$$

$$= \left(1 + \frac{(\xi_{R,k+1} - \xi_{R,k}) + (\xi_{A,k} - \xi_{A,k+1})}{(t_{R,k+1} - t_{R,k}) + (t_{A,k+1} - t_{A,k})}\right) \omega. \tag{7}$$

Note that the first summand in the numerator of the fraction in eq. (7) is always positive, while the second is negative. Thus they partly cancel each other. If further the denominator is fairly large, a good estimate for ω can be obtained.

3. ERROR MODEL

In Fig. 2 a measured histogram of frequency offset errors (blue curve) form a wireless 802.15.4 link is depicted. For this experiment we connected the same oscillator to two sensor nodes, such that both run on the same sampling frequency,

i.e., with zero frequency offset. Then we used the time stamp exchange algorithm on the wireless link between the nodes to estimate the frequency offset. What we then measure is the estimation error of the time stamp exchange algorithm.

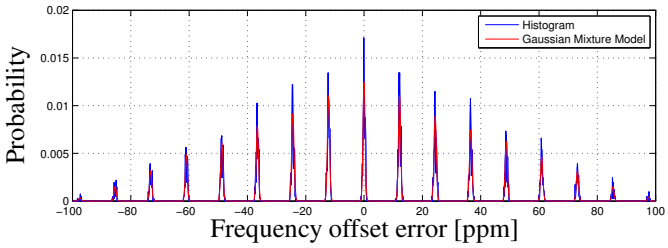

Fig. 2. Observation error distribution from experimental results (true value was 0, approximately 6 hours of data)

The characteristic structure of the frequency offset error can be explained as follows. The time stamp exchange over a wireless interface introduces at times additional non-symmetric latencies during the transmission. Reasons can be packet losses, protocol dependent wait states or the requirement to wait for the medium access control to begin the transmission. These effects introduce large frequency offset errors which are much larger than errors introduced by, e.g., the temporal drift of the crystal oscillators. These "large scale" errors are essentially discrete in nature: the transmission protocol may introduce an integer number of wait states, and they can be either positive or negative. The large scale errors are superposed by "small scale" errors, which are caused by the aforementioned temporal drift or by the spread of nominal frequency between two crystal oscillator devices.

Fig. 2 suggests to approximate the histogram by a Gaussian mixture model (GMM)

$$p(v_o) = \sum_{n=0}^{N-1} \gamma_n \mathcal{N}(v_o; \mu_{n,o}, \sigma^2_{n,o}) \tag{8}$$

where $\mathcal{N}(v_o; \mu_{n,o}, \sigma^2_{n,o})$ denotes a Normal distribution, $\mu_{n,o}$ are the means, $\sigma^2_{n,o}$ are the variances and γ_n are the weights of the GMM. The GMM is depicted in red in Fig. 2.

4. KALMAN FILTER

A simple kinematic model is employed in a Kalman filter to model oscillator frequency drifts. The two dimensional state vector $\boldsymbol{x} = [\omega, \Delta\omega]^T$ consists of the current clock frequency offset ω and its first derivative $\Delta\omega$. The state equation is given by

$$\boldsymbol{x}(n+1) = \underbrace{\begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix}}_{\boldsymbol{F}} \boldsymbol{x}(n) + \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{\boldsymbol{G}} \cdot \underbrace{\begin{bmatrix} 0 \\ v_s(n) \end{bmatrix}}_{\boldsymbol{v}_s} \tag{9}$$

with $v_s(n)$ being zero mean white Gaussian noise of variance σ_s^2, T being the time interval at which the time stamp exchange algorithm delivers estimates. The covariance matrix of the system noise is given by $\boldsymbol{Q}_s = E[\boldsymbol{v}_s \boldsymbol{v}_s^T]$.

The observation equation is given by

$$z(n) := \widehat{\omega}(n) = \underbrace{\begin{bmatrix} 1 & 0 \end{bmatrix}}_{\boldsymbol{H}^T} \boldsymbol{x}(n) + v_o(n) \tag{10}$$

with v_o being the observation noise, whose probability density function (PDF) is given by eq. (8) and can be estimated in advance from the experiment described above.

The minimum mean square error (MMSE) estimate of the systen state is then given by

$$\begin{aligned} \widehat{\boldsymbol{x}}(n|n) =& E[\boldsymbol{x}(n)|z(1), \ldots, z(n)] && (11) \\ =& \widehat{\boldsymbol{x}}(n|n-1) + \boldsymbol{K}(n) \cdot \\ & \cdot \left(z(n) - \boldsymbol{H}^T \widehat{\boldsymbol{x}}(n|n-1) - E[v_o(n)]\right) && (12) \end{aligned}$$

where we assumed that the observation error can have a nonzero mean. Here, $\boldsymbol{K}(n)$ is the Kalman gain and $\widehat{\boldsymbol{x}}(n|n-1)$ is the shorthand notation for the state estimate at time instance n given all observations up to time instance $n-1$.

We now assume that the prediction of the Kalman filter $E[z(n)|z(1), \ldots, z(n-1)] = \boldsymbol{H}^T \widehat{\boldsymbol{x}}(n|n-1)$ is so close to the true value ω that the contributions of large scale effects to the observation error can be uniquely detected. In other words, we assume that

$$|E[z(n)|z(1), \ldots, z(n-1)] - \omega| \ll \delta, \tag{13}$$

where δ is the minimal distance between two mixture component means:

$$\delta = \min_{k,l} |\mu_{k,o} - \mu_{l,o}| . \tag{14}$$

Then the large scale error can be identified by finding that mixture component that is closest to the prediction

$$\hat{k} = \operatorname*{argmin}_k |E[z(n)|z(1), \ldots, z(n-1)] - \mu_{k,o}| \tag{15}$$

and removed by subtracting the mean of the identified mixture component, $\mu_{\hat{k},o}$, from the observation

$$\begin{aligned} \widehat{\boldsymbol{x}}(n|n) =& \widehat{\boldsymbol{x}}(n|n-1) \\ & + \boldsymbol{K}(n) \left(z(n) - \boldsymbol{H}^T \widehat{\boldsymbol{x}}(n|n-1) - \mu_{\hat{k},o}]\right) . \end{aligned} \tag{16}$$

The experimental results presented in Section 5 indeed confirm that the assumption on the preciseness of the prediction is justified. The purpose of the Kalman filter is to account for the small scale errors. This results in a rather narrow filter bandwidth, which in turn allows to detect large scale errors, since they are much larger than the prediction error of the Kalman filter. Essentially, we exploit the specific nature of the observation noise, as depicted in Fig. 2.

According the the Kalman filter equations, the Kalman gain $\boldsymbol{K}(n)$ can be calculated by

$$\boldsymbol{K}(n) = \frac{\boldsymbol{\Sigma}(n|n-1)\boldsymbol{H}^T}{\boldsymbol{H}^T \boldsymbol{\Sigma}(n|n-1)\boldsymbol{H} + \widehat{\sigma}^2_{\hat{k},o}(n)} \tag{17}$$

with $\boldsymbol{\Sigma}(n|n)$ being the covariance matrix of the a posteriori state distribution and $\boldsymbol{\Sigma}(n|n-1)$ being its prediction from the

last time instance:

$$\boldsymbol{\Sigma}(n|n-1) = \boldsymbol{F}\boldsymbol{\Sigma}(n-1|n-1)\boldsymbol{F}^T + \boldsymbol{G}\boldsymbol{Q}_s\boldsymbol{G}^T \quad (18)$$

$$\boldsymbol{\Sigma}(n|n) = \boldsymbol{\Sigma}(n|n-1) - \boldsymbol{K}(n)\boldsymbol{H}^T\boldsymbol{\Sigma}(n|n-1). \quad (19)$$

5. EXPERIMENTAL RESULTS

5.1. Hardware platform

The hardware platform is a self development and it consists of a set of network connected multi-channel acoustic sensor nodes (see Fig. 3). Each node supports 16 channels, which are synchronously sampled. We use a sigma-delta analog-to-digital converter (ADC) with an oversampling factor of 512 to generate a 16 kHz sampling rate. The sampling rate is generated by a direct digital sequence (DDS) circuit driven by a 20 MHz oscillator. The DDS internally increases the oscillator frequency by a factor of 6 to an internal frequency of 120 MHz. The DDS enables the generation of arbitrary frequencies with sub-herz resolution (quantization $120\,\mathrm{MHz}/(2^{32}) = 0.0279\,\mathrm{Hz}$, which equals $0.003\,41\,\mathrm{ppm}$ at 16 kHz with 512 oversampling).

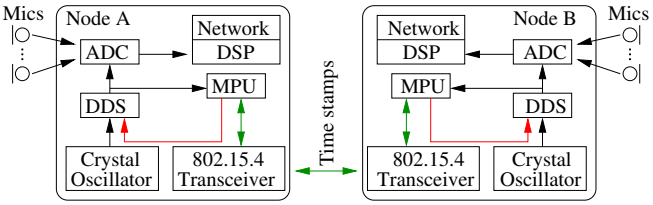

Fig. 3. Acoustic sensor network hardware components

A built-in hardware counter on the microprocessor unit (MPU) counts the oscillations of the DDS and uses this information as a time stamp. This enables a precise time stamp generation with low latency and jitter.

The MPUs of two sensor nodes exchange the time stamps via a wireless link using an IEEE 802.15.4 compatible transceiver (MAC & physical layer). Based upon the exchanged information the MPUs can estimate the clock frequency and phase offset and subsequently readjust the DDS such that the slave DDS generates the same sampling frequency as the master DDS. Although both nodes have different crystal oscillators the two DDS will generate approximately the same sampling frequencies.

5.2. Wired connection

In our first experiment we employed two sensor nodes which were connected via a wired USART connection. This connection introduces an observation error which can be approximately modeled by a single Gaussian distribution with zero mean. The above discussed algorithm to estimate the clock frequency offset and a Kalman filter (implemented on the MPUs) were used to adjust the oscillator of the slave node. An additional hardware device counted the oscillations of the DDS output signal of both nodes ($\approx 8.192\,\mathrm{MHz}$) and computed the difference between them.

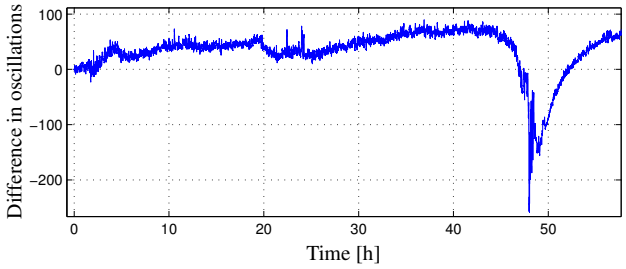

Fig. 4. Experiment using wired USART connection between two sensor nodes for time stamp exchange

In Fig. 4 the results of the experiment are shown. Within the observation time of 60 hours the maximum difference was kept below 250 oscillations, which equals a maximum sampling error of half a sample (note the oversampling factor of 512!). Thus, the proposed idea of using a time stamp exchange for synchronizing two sensor nodes and the DDS/MPU curcuit seem to deliver suitable results, if the frequency offset error estimate is reliable enough.

5.3. Wireless connection

In the second experiment the two nodes are connected via an IEEE 802.15.4 link which results in an observation error distribution as depicted in Fig. 2.

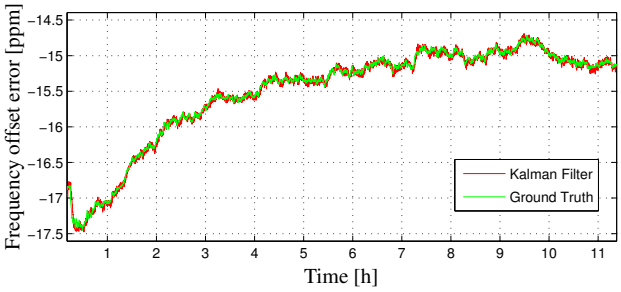

Fig. 5. Kalman filter results for frequency offset error estimates using a wireless 802.15.4 connection (no sampling frequency readjustment)

We estimated the GMM from training data in the manner described in Section 3 and used it within the proposed Kalman filter. From Fig. 5 it is obvious that the new approach is well suitable to track the frequency offset error, although the observation error is several orders of magnitude larger than the tracked signal. Note, that the ground truth value was measured via an extra hardware device counting the oscillations of both nodes and that in this experiment we did not adjust the DDS. The mean square error was measured to be $0.0019\,\mathrm{ppm}$.

In Fig. 6 the results of a third experiment are shown. Here, the Kalman filter state estimates are used to adjust the DDS of the slave node such that it aligns with the sampling frequency

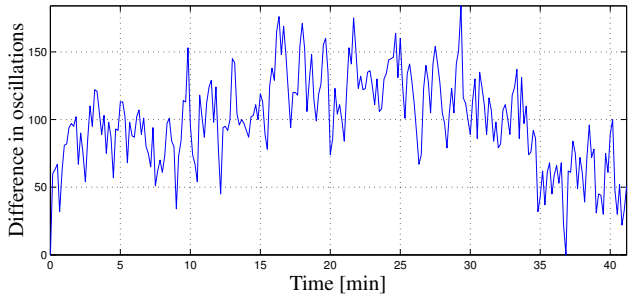

Fig. 6. Difference in oscillations using a Kalman filter on a wireless 802.15.4 connection (active readjustment of sampling frequency)

of the master node. As can be seen from the figure, the feedback control loop is able to keep the absolute difference between the two data streams below $180/512 = 0.35$ samples.

The exchange of the 64-Bit time stamps with an inter time stamp delay of 10 s generates a moderate data rate of $4 \cdot 64\,\mathrm{Bit}/10\,\mathrm{s} = 25.6\,\mathrm{Bit/s}$. Thus, the proposed approach seems to be applicable to larger networks with multiple nodes.

5.4. Impact on Localization

The presented experimental results have the following impact on a localization task. From Fig. 4 we can infer that the difference between the two data streams remains below a limit of 250 oscillations which, when divided by the oversampling factor of 512, gives a maximum difference of $250/512 \approx 0.49$ samples. With an inter microphone distance of 0.2 m and the acoustic source signal impinging from the endfire position it takes 9.4 sampling intervals for the signal to travel from the first to the second microphone, as stated in the introduction. Hence, the maximum error of 0.49 samples corresponds to a maximum angle error of $4.69°$. In the first 40 hours of the wired experiment the difference remained even in an area of 50 oscillations equaling an error of less than $1°$.

In Fig. 6 the results from the wireless connection are depicted, where the difference of oscillations between two nodes is kept below a value of 180 oscillations. This equals an angle error of less than $180/512 \cdot 90°/9.41 = 3.36°$.

The results of the experiments reveal that the chosen approach is able to keep the data streams well aligned. However, a clock phase compensation may further improve the results by first detecting the difference in oscillations and subsequently reducing it.

6. CONCLUSIONS

We have proposed a Kalman filter approach for post filtering the clock frequency offset estimates between two nodes of an acoustic sensor network. The filter exploits the characteristic properties of the estimation error in clock frequency offset estimation in wireless sensor networks. The synchronisation algorithm has been implemented on microprocessor units which communicate via an IEEE 802.15.4 wireless network at a low data rate. The estimated clock frequency offset is used to readjust the sampling frequency generator of one node such that the nodes sample the signals at approximately the same rate. In long term experiments the maximum difference between two data streams were kept below a maximum of half a sample.

7. ACKNOWLEDGMENTS

The authors wish to thank Jörg Ullmann and Karsten Böddeker for their work on the hardware of the sensor network nodes and their software code contributions.

8. REFERENCES

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," *Computer Networks*, vol. 38, pp. 393–422, 2002.

[2] Keewook Na, Yungeun Kim, and Hojung Cha, "Acoustic sensor network-based parking lot surveillance system," in *Proceedings of the 6th European Conference on Wireless Sensor Networks*, Berlin, Heidelberg, 2009, EWSN '09, pp. 247–262, Springer-Verlag.

[3] R. Heusdens, G. Zhang, R.C. Hendriks, Y. Zeng, and W.B. Kleijn, "Distributed MVDR beamforming for (wireless) microphone networks using message passing," in *Proc. International Workshop on Acoustic Signal Enhancement (IWAENC)*, Aachen, 2012.

[4] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, no. 4, pp. 320–327, Aug. 1976.

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

[6] 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 *Proc. International Workshop on Acoustic Signal Enhancement (IWAENC)*, Aachen, 2012.

[7] Jeremy Elson and Kay Römer, "Wireless sensor networks: a new regime for time synchronization," *SIGCOMM Comput. Commun. Rev.*, vol. 33, no. 1, pp. 149–154, Jan. 2003.

[8] Q.M. Chaudhari, "A simple and robust clock synchronization scheme," *Communications, IEEE Transactions on*, vol. 60, no. 2, pp. 328–332, February 2012.

[9] Fabian Kirsch and Martin Vossiek, "Distributed kalman filter for precise and robust clock synchronization in wireless networks," in *Proceedings of the 4th international conference on Radio and wireless symposium*, Piscataway, NJ, USA, 2009, RWS'09, pp. 455–458, IEEE Press.

[10] Jang-Sub Kim, Jaehan Lee, E. Serpedin, and K. Qaraqe, "A robust clock synchronization algorithm for wireless sensor networks," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, May 2011, pp. 3512–3515.