Online Observation Error Model Estimation for Acoustic Sensor Network Synchronization

Joerg Schmalenstroeer, Weile Zhao, Reinhold Haeb-Umbach

Department of Communications Engineering, University of Paderborn
Email: {schmalen, haeb}@nt.uni-paderborn.de
Web: nt.uni-paderborn.de

Abstract

Acoustic sensor network clock synchronization via time stamp exchange between the sensor nodes is not accurate enough for many acoustic signal processing tasks, such as speaker localization. To improve synchronization accuracy it has therefore been proposed to employ a Kalman Filter to obtain improved frequency deviation and phase offset estimates. The estimation requires a statistical model of the errors of the measurements obtained from the time stamp exchange algorithm. These errors are caused by random transmission delays and hardware effects and are thus network specific. In this contribution we develop an algorithm to estimate the parameters of the measurement error model alongside the Kalman filter based sampling clock synchronization, employing the Expectation Maximization algorithm. Simulation results demonstrate that the online estimation of the error model parameters leads only to a small degradation of the synchronization performance compared to a perfectly known observation error model.

1 Introduction

While wireless sensor networks have been an active area of research for many years, acoustic sensor networks have only recently received increased attention. With distributed microphones improved acoustic source tracking and speech enhancement can be achieved, as it is likely that at least one sensor is close to a source. However, if each sensor node is equipped with its own sampling oscillator, there is a need for synchronizing the distributed clocks to a precision sufficient for beamforming and localization purposes. Without clock synchronization, the deviation of the distributed crystal oscillators from their nominal frequency is so large, on the order of 50 ppm for standard low-cost devices, that the sampled acoustic streams can no longer be used for joint multi-channel processing. Even computer network synchronization methods, like the Network Time Protocol [1] or the time stamp exchange protocol by Chaudhari [2] do not achieve a precision that is high enough for acoustic signal processing.

A way to raise the precision of clock synchronization is to postprocess the phase and frequency deviation estimates obtained from an analysis of the properties of the sampled signals [3, 4] or by time stamp exchange. In [5] we devised a Kalman Filter to obtain refined estimates of the clock frequency deviation and the phase offset between a master and a slave sensor node from a measurement of the deviation that was computed using the two-way time stamp exchange protocol of [2]. Later this approach was extended to synchronize a whole sensor network to a common time base by devising a gossiping approach which requires only local exchange of timing information between neighboring nodes [6].

An important component of the employed Kalman filter is the so-called measurement or observation model which relates the measurements to the state variables of the filter in a probabilistic way. The relationship between the measured and the true clock frequency deviation and phase offset depends on the measurement approach used – in our case the time stamp exchange protocol of Chaudhari [2] – and the specific properties of the employed hardware and communication protocol. In [5] the observation model was estimated in advance in a special measurement setup, prior to its use in the clock synchronizing Kalman filter.

The goal of this paper is to develop an approach to jointly estimate the observation model parameters and track the frequency deviation and phase offset in a Kalman filter. Note that the first requires an estimate of the second and vice versa. This chicken-and-egg problem is solved by employing an iterative algorithm that alternates between observation model parameter estimation and Kalman filtering. The algorithm is derived as an instance of the Expectation Maximization (EM) algorithm. With the online estimation of the observation model the hope is to have a universal clock synchronization approach that is applicable to different networks and without a prior measurement phase: the system adapts to the characteristics of the network while it synchronizes the nodes to a common clock and while the nodes carry out signal processing tasks.

2 Clock Synchronization

In the following we shortly summarize the two-way message exchange algorithm by Chaudari [2], which is one of the basic components of the proposed synchronization approach.

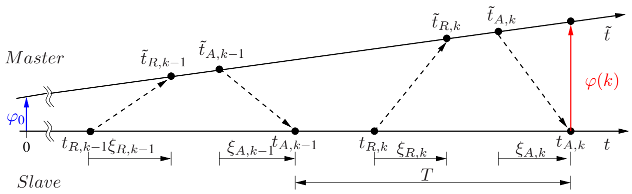

Figure 1: Two-way message exchange between master and slave node

Figure 1 shows the $(k-1)$-st and the k-th message exchange round, which have a temporal distance of T seconds. During the k-th message exchange the slave node sends a packet to the master at time $t_{R,k}$ (measured in units of the slave clock) which is received by the masters at time $\tilde{t}_{R,k}$ (master time). The unknown transmission time is given by $\xi_{R,k}$ (slave time). Subsequently the slave replies to the request with a packet at time $\tilde{t}_{A,k}$ which the master receives at time $t_{A,k}$ after a transmission time $\xi_{A,k}$.

Let us assume that the oscillator of the master node has a frequency of f_M with $f_M = (1+\epsilon)\cdot f_S$, where f_S is the oscillator frequency of the slave node. In the following,

$\epsilon = \epsilon(k)$ will be called the (relative) frequency deviation at the k-th time stamp exchange. According to [6] it can be estimated from the exchanged time stamps, see Fig. 1, via

$$\tilde{\epsilon}(k) = \frac{(\tilde{t}_{R,k} - \tilde{t}_{A,k-1}) - (\tilde{t}_{R,k-1} - \tilde{t}_{A,k})}{(t_{R,k} - t_{A,k-1}) - (t_{R,k-1} - t_{A,k})} - 1, \quad (1)$$

which can be regarded as the sum of the true frequency deviation $\epsilon(k)$ and an additive error term $v_\epsilon(k)$ [5]:

$$\tilde{\epsilon}(k) \approx \epsilon(k) + \underbrace{\frac{(\xi_{R,k} - \xi_{R,k-1}) + (\xi_{A,k-1} - \xi_{A,k})}{(t_{R,k} - t_{R,k-1}) + (t_{A,k} - t_{A,k-1})}}_{v_\epsilon(k)}. \quad (2)$$

Similary, we find for the phase offset estimate

$$\tilde{\varphi}(k) = \frac{1}{2}\left((\tilde{t}_{R,k} + \tilde{t}_{A,k}) - (t_{R,k} + t_{A,k})\right) \quad (3)$$

$$\approx \varphi(k) + \frac{1}{2}(\xi_{R,k} - \xi_{A,k}) = \varphi(k) + v_\varphi(k), \quad (4)$$

again being the sum of the true phase difference $\varphi(k)$ between master and slave node and an additive error $v_\varphi(k)$. In the following discussion $\tilde{\epsilon}(k)$ and $\tilde{\varphi}(k)$ will be called observations, as they constitute the measurements at the input of the Kalman filter.

2.1 Observation Error

We found in long term experiments on hardware devices (see [5]) that the distribution of v_ϵ has the shape as given by Fig. 2.

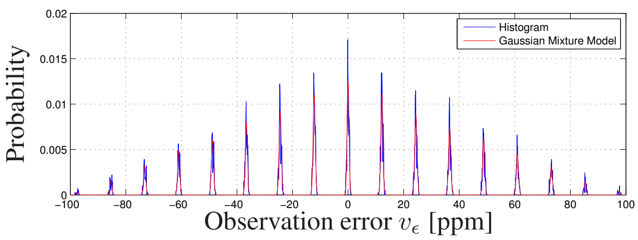

Figure 2: Histogram of observation error v_ϵ (obtained from 6 hours of data) and GMM approximation

This structure can be explained by the assumption that the unknown transmission times ξ consist of a constant minimum delay T_c, an integer multiplicity of delays T_d (e.g., from MAC wait times and timeouts) and an exponentially distributed random component T_e with $\xi = T_c + n \cdot T_d + T_e, n \in \mathbb{Z}$. Approximating the denominator of eq. (2) by $2T$, where T is the interval between two time stamp exchange rounds, see Fig. 1, the additive error term can be written as

$$v_\epsilon(k) \approx \frac{T_r(k)}{2T} + h(k) \cdot \frac{T_d}{2T}, \quad (5)$$

where $h(k)T_d$ summarizes the positive and negative multiples of T_d, and $T_r(k)$ comprises the exponentially distributed contributions from the four transmission times ξ. In the following, we refer to the first term in (5) as the "small scale error" and to the second as the "large scale error".

According to Fig. 2 the histogram can be well approximated by a Gaussian Mixture Model (GMM):

$$p(v_\epsilon) = \sum_{h=-M}^{M} \pi_{\epsilon,h} \cdot \mathcal{N}(v_\epsilon; h \cdot \mu_\epsilon, \sigma_\epsilon^2), \quad (6)$$

with weights $\pi_{\epsilon,h}$, where the variances of all component Gaussians are equal and denoted by σ_ϵ^2. Further the means of the Gaussians are on an equispaced grid and can thus be expressed by: $h \cdot \mu_\epsilon = h \cdot T_d/(2T)$.

Using this model, eq. (2) can be approximated by

$$\tilde{\epsilon}(k) = \epsilon(k) + c_\epsilon(k)\mu_\epsilon + n_\epsilon(k), \quad (7)$$

where $c_\epsilon(k) \in \{-M, \ldots M\}$ is the index of the mixture component from which the noise $v_\epsilon(k)$ is drawn, and $n_\epsilon(k)$ is Gaussian noise of variance σ_ϵ^2.

The phase offset observation error $v_\varphi(k)$ from eq. (3) is also approximated by a GMM. The means of the component Gaussians are multiples of $\mu_\varphi := T_d/2$. In the following we will derive the parameter estimation formulas for the frequency deviation related GMM only, since the phase offset GMM parameters can be either handled in the same manner or inferred from the frequency deviation parameters by $\mu_\varphi = T \cdot \mu_\epsilon$ and $\sigma_\varphi^2 = \sigma_\epsilon^2/2$.

Note, that the distribution in Fig. 2 depends on the hardware, the wireless network and the parameters of the two-way message exchange algorithm and thus will vary if other hardware components or parameter sets are used.

3 EM Parameter Estimation

We employ the EM algorithm to estimate the GMM parameters $\Theta_\epsilon = [\pi_{\epsilon,-M}, \ldots, \pi_{\epsilon,+M}, \mu_\epsilon, \sigma_\epsilon^2]$. Given N message exchanges, the observations are $\tilde{\boldsymbol{\epsilon}} = [\tilde{\epsilon}(1), \ldots \tilde{\epsilon}(N)]$. The hidden variables are chosen to be the true frequency deviations $\boldsymbol{\epsilon} = [\epsilon(1), \ldots \epsilon(N)]$, and the mixture component labels $\boldsymbol{c}_\epsilon = [c_\epsilon(1), \ldots c_\epsilon(N)]$.

The expected value of the loglikelihood of the complete data given the observations is

$$Q(\Theta_\epsilon; \Theta_\epsilon^{(\kappa)}) = E\left[\ln\{p(\boldsymbol{\epsilon}, \tilde{\boldsymbol{\epsilon}}, \boldsymbol{c}_\epsilon; \Theta_\epsilon)\} \,|\, \tilde{\boldsymbol{\epsilon}}; \Theta_\epsilon^{(\kappa)}\right] \quad (8)$$

$$= \sum_{\boldsymbol{c}_\epsilon} \int \ln\{p(\tilde{\boldsymbol{\epsilon}}|\boldsymbol{\epsilon}, \boldsymbol{c}_\epsilon) p(\boldsymbol{\epsilon}|\boldsymbol{c}_\epsilon) p(\boldsymbol{c}_\epsilon)\} \, p(\boldsymbol{\epsilon}, \boldsymbol{c}_\epsilon|\tilde{\boldsymbol{\epsilon}}) d\boldsymbol{\epsilon}$$

where κ is the iteration counter and where the summation is over all sequences $\boldsymbol{c}_\epsilon$ of length N. For the terms inside the logarithm in (8) we have

$$p(\tilde{\boldsymbol{\epsilon}}|\boldsymbol{\epsilon}, \boldsymbol{c}_\epsilon) = \prod_{k=1}^{N} \mathcal{N}(\tilde{\epsilon}(k); \epsilon(k) + c_\epsilon(k)\mu_\epsilon, \sigma_\epsilon^2) \quad (9)$$

$$p(\boldsymbol{c}_\epsilon) = \prod_{k=1}^{N} \pi_{\epsilon, c_\epsilon(k)}, \quad (10)$$

assuming that the labels $c_\epsilon(k)$ are i.i.d. Note that $p(\boldsymbol{\epsilon}|\boldsymbol{c}_\epsilon) = p(\boldsymbol{\epsilon})$ is independent of Θ_ϵ.

Expectation step

In the E-step the posterior distribution of the hidden variables is calculated. We assume that the frequency deviation $\epsilon(k)$ is Gaussian and follows a Markov process. Note, however, that the observation noise $v_\epsilon(k)$ is not Gaussian! Equation (7) can be interpreted as a switching observation model [7], where, given the value of 'switch variable' $c_\epsilon(k)$, the observation noise is Gaussian, see eq. (9). It is well-known that exact inference in the presence of a switching observation model is computationally intractable, since the number of hypotheses (i.e., switching variable sequences) to be considered grows exponentially with time [7]. In light of an intended implementation on low-cost hardware, the exponential growth is avoided by considering only the hypothesis with the highest probability.

The posterior probability of the label indicating the mixture component can be computed as

$$\begin{aligned}\gamma_h(k) &:= P(c_\epsilon(k)=h|\tilde{\epsilon}(k),\hat{\epsilon}(k|k-1);\Theta_\epsilon^{(\kappa)}) \qquad (11)\\ &= \frac{\mathcal{N}(\tilde{\epsilon}(k);\hat{\epsilon}(k|k-1)+h\mu_\epsilon^{(\kappa)},(\sigma_\epsilon^2)^{(\kappa)})\pi_{\epsilon,h}^{(\kappa)}}{\sum_{m=-M}^{M}\mathcal{N}(\tilde{\epsilon}(k);\hat{\epsilon}(k|k-1)+m\mu_\epsilon^{(\kappa)},(\sigma_\epsilon^2)^{(\kappa)})\pi_{\epsilon,m}^{(\kappa)}},\end{aligned}$$

where $\hat{\epsilon}(k|k-1)$ is the state prediction of the Kalman filter and

$$\hat{c}_\epsilon(k) = \underset{h}{\operatorname{argmax}}\{\gamma_h(k)\} \qquad (12)$$

is the class label of the most likely hypothesis. The standard Kalman filter is fed with the observations

$$z(k) = \tilde{\epsilon}(k) - \hat{c}_\epsilon(k)\mu_\epsilon. \qquad (13)$$

Note that this Kalman filter only deals with the small scale errors, while the large scale errors are removed prior to the Kalman filter by eq. (13).

With these approximations the auxiliary function of eq. (8) can be written as

$$Q(\Theta_\epsilon;\Theta_\epsilon^{(\kappa)}) \propto \sum_{k=1}^{N} Q_k(\Theta_\epsilon;\Theta_\epsilon^{(\kappa)}), \qquad (14)$$

where

$$\begin{aligned}Q_k(\Theta_\epsilon;\Theta_\epsilon^{(\kappa)}) = \sum_{h=-M}^{M} \ln\Big\{&\pi_{\epsilon,h}\mathcal{N}(\tilde{\epsilon}(k);\hat{\epsilon}(k|k)\\ &+\hat{c}_\epsilon(k)\mu_\epsilon,\sigma_\epsilon^2(k))\Big\}\gamma_h(k). \qquad (15)\end{aligned}$$

Maximization step

For the M-Step we take the derivatives of eq. (14) with respect to the GMM parameters. It holds for the weights

$$\pi_{\epsilon,h}^{(\kappa+1)} = \frac{1}{N}\sum_{k=1}^{N}\gamma_h(k), \qquad (16)$$

the mean is updated via

$$\mu_\epsilon^{(\kappa+1)} = \frac{\sum_{k=1}^{N}\sum_{h=-M}^{+M}\gamma_h(k)\cdot h\cdot(\tilde{\epsilon}(k)-\hat{\epsilon}(k|k))}{\sum_{k=1}^{N}\sum_{h=-M}^{+M}\gamma_h(k)\cdot h^2} \qquad (17)$$

and the variance with

$$(\sigma_\epsilon^2)^{(\kappa+1)} = \frac{\sum_{k=1}^{N}\sum_{h=-M}^{+M}\gamma_h(k)\cdot(\tilde{\epsilon}(k)-\hat{\epsilon}(k|k)-h\mu_{\epsilon,\kappa})^2}{\sum_{k=1}^{N}\sum_{h=-M}^{+M}\gamma_h(k)}. \qquad (18)$$

3.1 Initialization of parameters

At the beginning the identification of the large scale error via eq. (12) cannot be performed, because the Kalman filter state estimate $\hat{\epsilon}(k|k)$ is not available. However, we can assume that the deviation of the clock does not change abruptly over time and that the ground truth during the k-th and $(k-1)$-st message exchange are roughly equal. We subtract subsequent observations to get observation error estimates $\Delta\tilde{v}_\epsilon(k)$ with

$$\Delta v_\epsilon(k) = \tilde{\epsilon}(k) - \tilde{\epsilon}(k-1) \approx v_\epsilon(k) - v_\epsilon(k-1). \qquad (19)$$

The distribution of $\Delta v_\epsilon(k)$ is a GMM with double the number of mixture components as the distribution of v_ϵ, and whose component means are also equispaced at multiples of μ_ϵ, while the variances of the component densities are doubled. However, initial values for the mean μ_ϵ and the variance σ_ϵ^2 can be easily found by clustering a set of L values $\Delta\tilde{v}_\epsilon(k), k=1,\dots L$.

3.2 Online parameter estimation

To arrive at an online parameter estimation the iterative block EM algorithm is replaced by a recursive online EM. The latter can be derived employing a gradient ascent approach:

$$\begin{aligned}\Theta_\epsilon(k) &= \Theta_\epsilon(k-1)\\ &+\tau\cdot\nabla_{\Theta_\epsilon}\{Q_k(\Theta_\epsilon;\Theta_\epsilon(k-1))\}\Big|_{\Theta_\epsilon=\Theta_\epsilon(k-1)} \qquad (20)\end{aligned}$$

where the iteration index κ is replaced by the observation index k and τ is the step size. We find for the mean

$$\begin{aligned}\mu_{\epsilon,k} &= \mu_{\epsilon,k-1}\\ &+\tau_\epsilon\sum_{h=-M}^{+M}\gamma_h(k)h\left[\tilde{\epsilon}(k)-\hat{\epsilon}(k|k)-h\mu_{\epsilon,k-1}\right] \qquad (21)\end{aligned}$$

and for the variance

$$\begin{aligned}\sigma_{\epsilon,k}^2 = \sigma_{\epsilon,k-1}^2 - \tau_\sigma\sum_{h=-M}^{+M}\gamma_h(k)\Big[&\sigma_{\epsilon,k-1}^2\\ &-\Big(\tilde{\epsilon}(k)-\hat{\epsilon}(k|k)-h\mu_{\epsilon,\kappa}\Big)^2\Big]. \qquad (22)\end{aligned}$$

Finally, the weight recursion is given by

$$\pi_{\epsilon,h,k} = \pi_{\epsilon,h,k-1} + \tau_\pi\left[\gamma_h(k)-\pi_{\epsilon,h,k-1}\right], \qquad (23)$$

while by renormalization it is assured that the sum of all weights equals one.

4 Simulation results

We simulated networks with random topologies to study the influence of the online error model estimation on the synchronization process. For each network size we conducted 300 simulations of 2 h length. The average number of neighboring nodes a node has is approximately 3.

The clock synchronization between a master and a slave node that was discussed so far in this paper was extended to synchronize the clocks of many sensor nodes to a common time base by employing a gossiping algorithm,

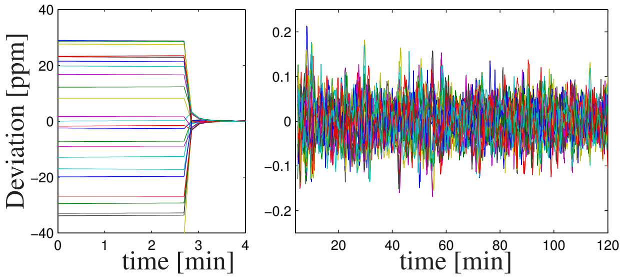

Figure 3: Deviation of sampling frequencies between virtual master node and slave nodes (network with 25 nodes)

which exchanges timing information only among neighboring nodes. For a description of the gossiping based sensor network clock synchronization the reader is referred to [6].

In Fig. 3 a sample result for a network of 25 nodes is depicted (left part shows the first 4 min, right part shows 4 min till 120 min. Note the different resolutions of the y-axis.). In the first 3 min the initialization phase takes place as discussed in Sec. 3.1. Given $T = 10$ s this amounts 18 observations. If the number of observations is further reduced, there is a risk that initialization fails. Afterwards, the network remains in a synchronized state keeping the frequency deviations of all nodes below 0.2 ppm (see Fig. 3, right).

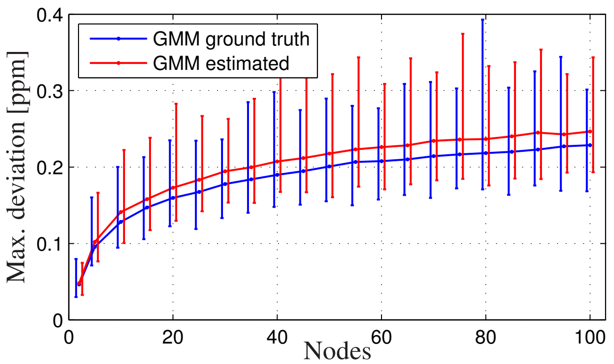

Figure 4: Maximum frequency deviation between nodes in the network after initialization phase

In Fig. 4 the results for the maximum frequency deviation between nodes, i.e. the largest observed frequency deviation between all combinations of nodes after the initialization phase, are depicted. For each network size the span width between the best and the worst experiment as well as the average value of all experiments are shown. Here, the proposed approach for estimating the GMM (“GMM estimated”) is compared against our previous approach where the ground truth GMM (“GMM ground truth”) is estimated from training data in advance.

A deviation of 0.2 ppm for the duration of $T = 10\,\mathrm{s}$ causes a phase error of $0.2\,\mathrm{ppm} \cdot 10\,\mathrm{s} \cdot 8.192\,\mathrm{MHz} = 16.384\,\mathrm{Oscillations}$. Since our system works at an oversampling factor of 512 (oscillator frequency: 8.192 MHz, sampling frequency: 16 kHz) the phase error between the audio streams increases by $16.384\,\mathrm{Oscillations}/512 = 0.032\,\mathrm{Sample} \widehat{=} 3.2\,\%$.

In Fig. 5 the RMS phase errors are shown, where the span for the best and the worst results of the 300 simulations and the average RMS are marked. Since the phase offset is the accumulation of the frequency deviation over time it is not surprising that the phase offset error increases. However, the results for the frequency deviation and the phase offset are promising that the proposed approach is capable to estimate the error models precise enough, that the overall performance of the network synchronization is only slightly degraded.

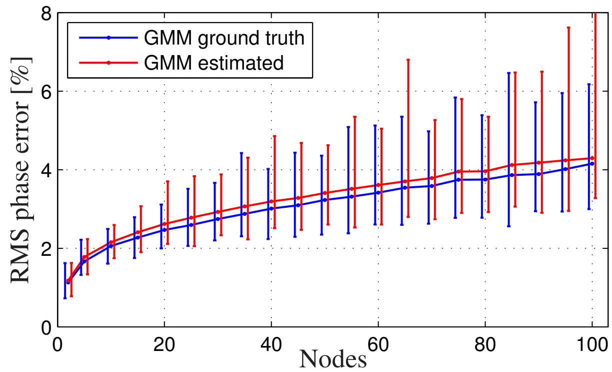

Figure 5: RMS phase errors between nodes in the network after initialization phase

5 Conclusions

We have presented an online expectation maximization approach for estimating observation error models, which is applicable even on limited hardware resources. To this end the Kalman filter predictions are integrated into the expectation step of the EM algorithm. Simulations of large random networks showed that the remaining error caused by the non-optimal error model only slightly degrade the network synchronization performance.

References

[1] D. L. Mills, “Internet time synchronization: the network time protocol,” *IEEE Transactions on Communications*, vol. 39, pp. 1482–1493, 1991.

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

[3] S. Miyabe, N. Ono, and S. Makino, “Blind compensation of inter-channel sampling frequency mismatch with maximum likelihood estimation in the STFT domain,” in *Proc. ICASSP*, (Vancouver), May 2013.

[4] 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.

[5] J. Schmalenstroeer and R. Haeb-Umbach, “Sampling rate synchronisation in acoustic sensor networks with a pre-trained clock skew error model,” in *Proc. 21th European Signal Processing Conference, EUSIPCO’13*, 2013.

[6] J. Schmalenstroeer, P. Jebramcik, and R. Haeb-Umbach, “A gossiping approach to sampling clock synchronization in wireless acoustic sensor networks,” in *IEEE International Conf. on Acoustics, Speech and Signal Processing (ICASSP’14)*, (Florence, Italy), May 2014.

[7] K. Murphy, “Switching Kalman filters,” tech. rep., U.C. Berkeley, 1998.