

# On Frequency Tracking in Harmonic Acoustic Signals

Martin Lindfors, Gustaf Hendeby, Fredrik Gustafsson and Rickard Karlsson

The self-archived version of this journal article is available at Linköping University Electronic Press:

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-139796>

**N.B.: When citing this work, cite the original publication.**

Lindfors, M., Hendeby, G., Gustafsson, F., Karlsson, R., (2017), On Frequency Tracking in Harmonic Acoustic Signals, *Proceedings of the 2017 20th International Conference on Information Fusion (FUSION)*.

Original publication available at:

Copyright: IEEE <http://www.ieee.org>



# On Frequency Tracking in Harmonic Acoustic Signals

Martin Lindfors, Gustaf Hendeby, Fredrik Gustafsson and Rickard Karlsson

Dept. of Electrical Engineering, Linköping University, Sweden

E-mail: martin.lindfors@liu.se, {hendeby, fredrik, rickard}@isy.liu.se.

**Abstract**—Acoustic frequency tracking of a harmonic signal with continuously varying frequency is considered. The Rao-Blackwellized point mass filter (RBPMF), previously proposed by the authors for mechanical vibration tracking, is applied to the problem. The RBPMF is compared with two periodogram-based methods, and the similarities and differences between them are explained. Both experimental and simulation results in a Doppler frequency tracking scenario are presented, and the results show that the RBPMF can have significantly less estimation error than the competing methods.

## I. INTRODUCTION

Frequency and harmonics tracking of a periodic signal in noise is a well studied problem; for an overview, see [1]. One approach includes methods which assume that the frequency is approximately constant over a certain interval (batch). Frequency estimates over that batch can then be computed based on maximum likelihood [2], periodogram maximization, [3], or the autocorrelation function [4, 5]. These approaches work well in high *signal-to-noise ratio* (SNR) conditions [6], but depend on the constant frequency assumption. In addition, the phase and amplitude coherence over different batches is largely ignored.

As an alternative to batch methods, it is possible to model a changing source frequency with Markovian state-space models. Since the resulting estimation problem is nonlinear and high-dimensional, approximate filtering methods have to be used. When the measurement noise is sufficiently white, and the SNR is sufficiently high, linearization-based filters such as the extended Kalman filter, [7], can be used, as in [8–10]. When the SNR is worse, the problem may become multimodal and unsuited to the Gaussian approximations inherent in Kalman filtering. Instead, it is preferable to use methods that approximate the recursive Bayesian filtering equations directly. These approaches include *maximum a posteriori filters* [11], *particle filters* [12, 13], and *point mass filters* (PMFs) [14].

In this paper the RBPMF, previously proposed by the authors of this paper in [14] for frequency tracking of mechanical vibrations, is applied to a new scenario encompassing acoustic frequency tracking. The RBPMF uses a deterministic grid over the frequency, and approximates the full Bayesian filtering equations. Thus, the RBPMF is capable of full multi-hypothesis frequency tracking.

The proposed method is evaluated on simulated and experimental data. In the studied scenario, a vehicle is passing through an area where microphones with known positions

measure the sound from, among others, the vehicle engine. The measurements are Doppler shifted due to the movement of the sound-emitting vehicle, and this shift is different for microphones situated in different positions. Other works, such as [15], have shown that this effect can be used for estimating the vehicle trajectory. The RBPMF is compared with two periodogram-based methods and evaluated on both experimental and simulated data sets from the scenario.

The paper is organized as follows: Sec. II describes frequency and harmonics estimation using maximum likelihood and periodogram methods, while Sec. III presents extensions to tracking a varying frequency of a harmonic signal. Sec. IV and Sec. V evaluate the aforementioned methods in the Doppler tracking scenario using experimental and simulated data, respectively, and Sec. VI concludes the paper.

## II. FREQUENCY ESTIMATION

In this paper, frequency tracking of a harmonic signal, with first harmonic or fundamental frequency  $\omega$  and higher harmonics at frequencies  $m\omega$ ,  $m = 2, \dots, M$ , is considered. In order to understand the fundamental properties of the approach, frequency estimation, *i.e.*, the case when the frequency, amplitude and phase of the harmonic signal does not vary with time, is considered first.

### A. Maximum Likelihood Frequency Estimation

The problem of harmonics estimation, see *e.g.* [1], comes down to estimation of the angular frequency  $\omega$  of the signal

$$\begin{aligned} y_k &= \sum_{m=1}^M \operatorname{Re} \left( e^{im\omega kT} (\alpha^{(m)} + i\beta^{(m)}) \right) + e_k \\ &= \sum_{m=1}^M \left( \alpha^{(m)} \cos(m\omega kT) - \beta^{(m)} \sin(m\omega kT) \right) + e_k, \end{aligned} \quad (1)$$

where  $k$  is the discrete time index,  $T$  is the sample time, and  $M$  is the number of harmonics. The parameters  $\alpha^{(m)}$  and  $\beta^{(m)}$  are the real and imaginary components of the phasor  $\alpha^{(m)} + i\beta^{(m)}$ , representing the phase offset and amplitude of each harmonic  $m = 1, \dots, M$ , as illustrated in Fig. 1. The variable  $e_k$  represents the measurement noise, which is assumed to be Gaussian and white,  $e_k \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$ .

Since  $\alpha^{(m)}$  and  $\beta^{(m)}$  are unknown, they must also be estimated or marginalized away. Since they occur linearly in (1), one approach is to estimate them conditioned on  $\omega$ , and

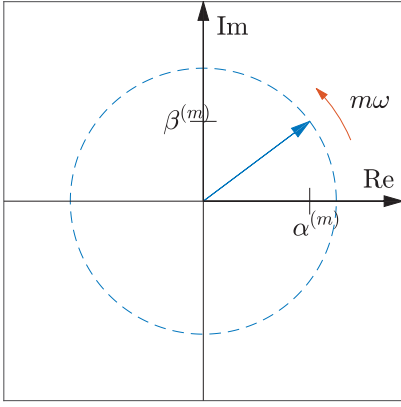


Fig. 1. Illustration of the representation of phase and amplitude of each harmonic in the signal. A phasor is used, such that the amplitude of harmonic  $m$  is  $|a^{(m)} + ib^{(m)}|$  and the phase is  $\arg(a^{(m)} + ib^{(m)})$ .

then estimate  $\omega$  using this estimate; the separable least squares approach. Stacking  $N$  instances of (1) in a matrix equation yields

$$Y_N = A(\omega)Z + E, \quad (2)$$

where  $Y_N = [y_1, \dots, y_N]^T$ ,  $Z = [\alpha^{(1)}, \beta^{(1)}, \dots, \alpha^{(M)}, \beta^{(M)}]^T$ , and

$$(A(\omega))_{k,2m-1} = \cos(m\omega kT), \quad (3a)$$

$$(A(\omega))_{k,2m} = -\sin(m\omega kT). \quad (3b)$$

The corresponding joint least squares problem for  $(Z, \omega)$  is

$$(\hat{Z}, \hat{\omega}) = \arg \min_{Z, \omega} (Y_N - A(\omega)Z)^T (Y_N - A(\omega)Z), \quad (4)$$

and given  $\omega$ , it is possible to solve for  $\hat{Z}(\omega) = D(\omega)^{-1}F(\omega)$ , where

$$D(\omega) = A(\omega)^T A(\omega), \quad F(\omega) = A(\omega)^T Y_N. \quad (5)$$

Substitution of this solution into (4), followed by simplifications, yields the least squares problem for  $\omega$  given by

$$\hat{\omega} = \arg \min_{\omega} -F(\omega)^T D(\omega)^{-1} F(\omega). \quad (6)$$

### B. Periodogram Frequency Estimation

It is now shown that (6) is connected to the well known periodogram maximization approach. Under suitable conditions (e.g.,  $m\omega T \leq \pi$ ) [1], it can be shown that  $D(\omega)$  approaches a scaled identity matrix,  $\frac{2}{N}D(\omega) \rightarrow I$  as  $N \rightarrow \infty$ . Thus, (6) may be approximately minimized when  $N$  is large by maximizing the expression

$$\frac{2}{N}F(\omega)^T F(\omega) = \frac{2}{N} \sum_{m=1}^M |\mathcal{Y}(m\omega)|^2, \quad (7)$$

where each summand is an input-scaled truncated *discrete-time Fourier transform* (DTFT),  $\mathcal{Y}(\omega) = \sum_{k=1}^N y_k e^{-ik\omega T}$ , so that  $|\mathcal{Y}(\omega)|^2$  is the periodogram of  $Y_N$ . This is an intuitively appealing result since peaks in the periodogram are known to indicate frequencies of periodic signals in the data. The case when  $e_k$  is colored with known spectrum results in a frequency

weighting of the summands in (7) [1]. The periodogram  $|\mathcal{Y}(\omega)|^2$  is often smoothed, see [3]; this makes the estimator asymptotically less efficient but may improve small-sample properties or noise sensitivity.

### C. Discrete Periodogram Maximization (DPM)

One way to implement the method described in Sec. II-B is to use the computationally efficient *fast Fourier transform* (FFT) instead of the DTFT. This approach is named *discrete periodogram maximization* (DPM); the discretization means that the maximization cannot get exactly carried out over the continuous range of  $\omega$ . Instead, only the information given from the FFT grid points is used;  $\omega_l = l\omega_S/N$ , where  $\omega_S$  is the sampling frequency and  $l$  is an integer index. The authors of [15] evaluate a smoothed sum on the discrete frequency grid points of the FFT and pick the fundamental frequency as the maximum of that sum;  $\hat{\omega} = \omega_l$  where

$$\hat{l} = \arg \max_l \sum_{m=1}^M \sum_{p=-1}^1 c_p |\mathcal{Y}(m\omega_{l+p})|^2, \quad (8)$$

where the maximization is carried out for  $l$  such that  $\omega_l$  lies within a given range  $\omega_{\min} \leq \omega_l \leq \omega_{\max}$ . The parameters  $c_p$  are smoothing weights.

### D. Subharmonic Summation (SHS)

The method of *subharmonic summation* (SHS) [3] is popular in acoustic frequency estimation for, e.g., speech. Instead of maximizing over only the grid points, it interpolates the FFT  $\mathcal{Y}(\omega_l)$  to a logarithmic frequency scale  $\check{\mathcal{Y}}(\log \omega)$ , and uses the approximate objective function given by

$$\sum_{m=1}^M c_m |\mathcal{Y}(m\omega)|^2 \approx \sum_{m=1}^M c_m |\check{\mathcal{Y}}(\log m + \log \omega)|^2, \quad (9)$$

where  $c_m$  are harmonics scaling parameters added for tuning. They can be viewed as approximation of the frequency weighting mentioned above. Hence, SHS computes the magnitude spectrum of the signal and interpolates the frequency to a logarithmic scale. Translated, scaled versions of the spectrum are then summed together, and the estimate of the fundamental frequency is given by the frequency which maximizes the said sum.

## III. FREQUENCY TRACKING

In many applications, the relevant frequency is not constant over time, but varies. One way to approach this problem is to assume that the frequency is constant in each batch, another is to model it as varying between each sample instant.

### A. Batched Frequency Tracking

Given measurements of a signal where the fundamental frequency is sufficiently slowly varying, and the SNR sufficiently high, it is clear that one approach is to group the measurements into batches of length  $N_b$ . The fundamental frequency in each batch can then be estimated as described in Sec. II. This way, the frequency variation within each batch is ignored, and

information about phase and amplitude coherence between batches is discarded. In order to retain partial information on phase and amplitude coherence it is possible to let the batches partially overlap. However, this places further demands on low frequency variation in the measurement sequence. This is because overlapping batches, with the same spacing between each initial batch sample, are larger than ones which do not overlap. Further information on the batched approach is available in *e.g.* [1, 3, 15].

### B. Recursive Frequency Tracking

Instead of assuming that the frequency is constant over batches, it can be modeled as varying between each sample instant instead. This can be cast as a state-space model, and one way to formulate it, which will be suitable to the upcoming method, is to look at a way to reformulate (1), where  $\alpha^{(m)} + i\beta^{(m)}$  was a phasor representing the *initial* phase and amplitude of the harmonic  $m$ , as in Fig. 1. Instead, let  $\alpha_k^{(m)} + i\beta_k^{(m)}$ , now with time index  $k$ , be the phasor representation of the *instant* phase and amplitude. The dynamic behavior of this phasor with constant angular frequency  $m\omega_0$  is given by

$$\begin{aligned}\alpha_k^{(m)} + i\beta_k^{(m)} &= e^{im\omega_0 T}(\alpha_{k-1}^{(m)} + i\beta_{k-1}^{(m)}) \\ &= e^{im\omega_0 kT}(\alpha_0^{(m)} + i\beta_0^{(m)}),\end{aligned}\quad (10)$$

and if the measurement equation is chosen as

$$\begin{aligned}y_k &= \sum_{m=1}^M \alpha_k^{(m)} + e_k \\ &= \sum_{m=1}^M \operatorname{Re}\left(e^{im\omega_0 kT}(\alpha_0^{(m)} + i\beta_0^{(m)})\right) + e_k,\end{aligned}\quad (11)$$

it can be seen that this is an equivalent formulation of (1). However, this formulation is better suited for introduction of varying amplitudes, phase offsets and fundamental frequency through process noises. In a state-space framework, the dynamics (10) can be extended as

$$\omega_{k+1} = \omega_k + \sqrt{T}v_k^\omega, \quad (12a)$$

$$\begin{bmatrix} \alpha_{k+1}^{(m)} \\ \beta_{k+1}^{(m)} \end{bmatrix} = R(m\omega_k T) \begin{bmatrix} \alpha_k^{(m)} \\ \beta_k^{(m)} \end{bmatrix} + \sqrt{T} \begin{bmatrix} v_k^{\alpha^{(m)}} \\ v_k^{\beta^{(m)}} \end{bmatrix}, \quad (12b)$$

where independent Gaussian process noises  $v_k^\omega$ ,  $v_k^{\alpha^{(m)}}$ ,  $v_k^{\beta^{(m)}}$  have been added in order to adapt the model to changes in amplitude, frequency, and phase. The two dimensional rotation matrix

$$R(\psi) = \begin{bmatrix} \cos \psi & -\sin \psi \\ \sin \psi & \cos \psi \end{bmatrix} \quad (13)$$

implements the complex exponential multiplication in (10).

Given  $\omega_k$ , the phasor parameters

$$Z_k = [\alpha_k^{(1)}, \beta_k^{(1)}, \alpha_k^{(2)}, \beta_k^{(2)}, \dots, \alpha_k^{(M)}, \beta_k^{(M)}]^\top \quad (14)$$

occur in a linear-Gaussian manner, which makes it suitable for Rao-Blackwellization, see *e.g.* [16]. The full state vector is  $[\omega_k, Z_k^\top]^\top$ .

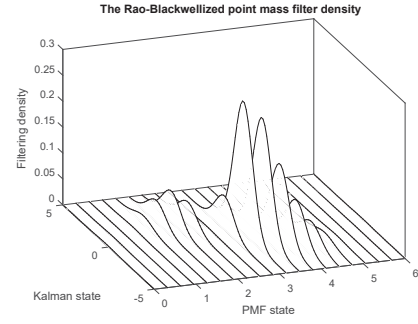


Fig. 2. Illustration of the RBPMF density approximation for a two-dimensional density, *i.e.*, when  $Z_k$  is a scalar in (15). The PMF is run on one state, yielding a discrete approximate density, while the Kalman filter is run on the other state yielding a Gaussian density conditional on the PMF state; thus, each curve represents one Kalman filter. The heights of the curves correspond to the weights  $\gamma_k^{(j)}$  of each point mass.

### C. The Rao-Blackwellized Point Mass Filter

It is possible to apply an RBPMF, as described by [14, 17], to the problem with the given model structure. The RBPMF can handle the time-varying behavior of the tracked fundamental frequency directly without resorting to a batch formulation, and is capable of full multi-hypothesis tracking of the fundamental frequency. However, it is more computationally expensive than the batch methods. In order to understand the RBPMF, the PMF must first be understood.

The PMF is one approach to compute approximate Bayesian solutions to filtering problems [18]. It computes integrals over the probabilistic state space using a deterministic grid of points. This makes it possible to compute an approximate posterior distribution of the state over the said grid.

When parts of the state vector occur linearly with Gaussian noise, the technique called Rao-Blackwellization can be applied, see *e.g.* [16, 19]. This approach infers the distribution of the linear-Gaussian components (*e.g.*,  $Z_k$ ) optimally, using Kalman filters [20], conditional on the remaining components (*e.g.*,  $\omega_k$ ), which are in turn inferred approximately using a nonlinear filter. In the case that the phasor parameters  $Z_k$  are handled by a Kalman filter, and the fundamental frequency  $\omega_k$  is handled by a PMF, this is called the RBPMF. The corresponding factorization of the joint filtering density is given by

$$\begin{aligned}p(\omega_k, Z_k | Y_k) &= p(\omega_k | Y_k)p(Z_k | \omega_k, Y_k) \\ &\approx \sum_{j=1}^J \gamma_k^{(j)} \delta(\omega_k - \omega^{(j)}) \mathcal{N}(Z_k | \hat{Z}_k^{(j)}, P_k^{(j)}),\end{aligned}\quad (15)$$

illustrated in Fig. 2 for a two-dimensional example. Each weighted Dirac function  $\gamma_k^{(j)} \delta(\omega_k - \omega^{(j)})$  represents one of the  $J$  grid points of the PMF estimating the marginal density  $p(\omega_k | Y_k)$ . Each Gaussian density  $\mathcal{N}(Z_k | \hat{Z}_k^{(j)}, P_k^{(j)})$  represents the Kalman filter running on  $Z_k$  approximating the conditional density  $p(Z_k | \omega^{(j)}, Y_k)$ . For further details, see [14, 17].

As a point estimate, the filtering mean

$$\hat{\omega}_k = \mathbb{E}[\omega_k | Y_k] = \sum_{j=1}^J \gamma_k^{(j)} \omega^{(j)} \quad (16)$$

is used. Furthermore, the filtering variance

$$\text{Var}(\omega_k) = \mathbb{E}[(\omega_k - \hat{\omega}_k)^2 | Y_k] = \sum_{j=1}^J \gamma_k^{(j)} (\omega^{(j)} - \hat{\omega}_k)^2 \quad (17)$$

can be computed and used as an activation criterion  $\text{Var}(\omega_k) \leq L_{\text{RBPMF}}$  in order to determine when the estimate is sufficiently accurate.  $L_{\text{RBPMF}}$  is a tuning parameter.

The RBPMF returns a frequency estimate at every sample point. In order to compare its output to the batch-based methods with the same time fidelity, an average over  $N_b$  instants of the expectation  $\hat{\omega}_k$  is computed.

#### IV. EXPERIMENTAL EVALUATION

For experimental evaluation, a data set described in [21] is used. The experimental data was collected in the scenario illustrated in Fig. 3. An *all-terrain vehicle* (ATV) traversed the circular segment at constant speed. Microphones with known positions in the area measured the sound emitted by the ATV. The recorded sound contains clearly audible harmonics from the ATV's engine. Since no ground truth frequency is available, a reference frequency is estimated using the fact that the different microphones measure different, Doppler-shifted, delayed versions of the same engine sound with continuously varying frequency.

##### A. Experimental Setup

The experimental data used for evaluation of the methods was collected using a setup of eight single microphones (M1–M8), and one acoustic array (A) with four microphones (A1–A4), as described in [21]. Acoustic data from an ATV was collected while the vehicle passed through the area at approximately constant speed and engine speed. The acoustic data from the microphones was recorded at a sample rate of 1200 Hz. It contains multiple harmonics generated by the vehicle's internal combustion engine, frequency shifted by the Doppler effect. The area used for testing is shown in Fig. 3.

The positions of the microphones are known, and in order to validate the method, a *differential global positioning system* (DGPS) receiver measures the position of the vehicle with high accuracy.

##### B. The Doppler Effect

In order to compute a reference frequency from the DGPS positions, the Doppler effect, [22], needs to be understood. The Doppler effect occurs when the transmitter and receiver of a signal are moving relative to each other. Let  $c$  be the transmission speed of the signal in the medium, and  $r^{(i)}(t)$  be the distance between the transmitter and the stationary receiver  $i$  at time  $t$ . Then, the propagation delay  $\Delta^{(i)}(t)$  is given by the implicit equation

$$r^{(i)}(t - \Delta^{(i)}(t)) = c\Delta^{(i)}(t). \quad (18)$$

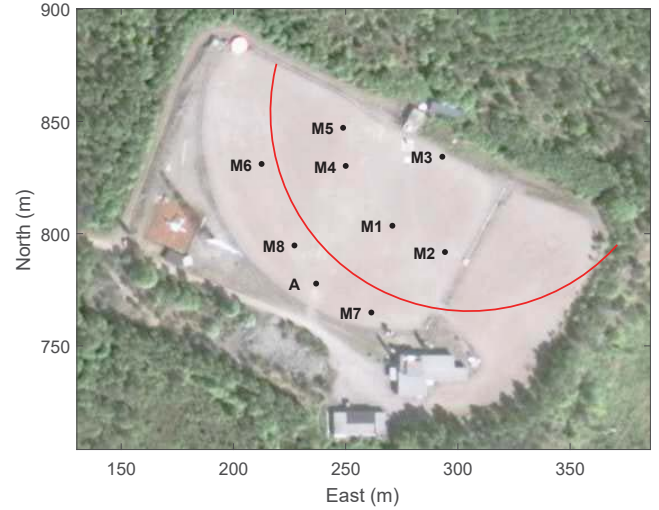


Fig. 3. The test area used for experiments. The vehicle trajectory is a red curve, traversed at approximately constant velocity. The black dots denote microphone locations. The microphone array (A) consists of four microphones (A1, A2, A3, A4), yielding 12 microphones in total.

Assume that the transmitter broadcasts an acoustic signal with instant phase  $\Phi^{(0)}(t)$  and frequency  $f^{(0)}(t) = \frac{d}{dt}\Phi^{(0)}(t)$ . Then, the  $i$ th receiver will measure the Doppler-shifted signal with instant phase

$$\Phi^{(i)}(t) = \Phi^{(0)}(t - \Delta^{(i)}(t)) \quad (19)$$

and thus received frequency

$$f^{(i)}(t) = \frac{d}{dt}\Phi^{(i)}(t) = f^{(0)}(t - \Delta^{(i)}(t)) \cdot (1 - \dot{\Delta}^{(i)}(t)). \quad (20)$$

$\dot{\Delta}^{(i)}(t)$  can be computed by differentiating (18) w.r.t. time, yielding

$$\dot{\Delta}^{(i)}(t) = \frac{v^{(i)}(t - \Delta^{(i)}(t))}{c + v^{(i)}(t - \Delta^{(i)}(t))}, \quad (21)$$

where  $v^{(i)}(t) = \frac{d}{dt}r^{(i)}(t)$  is the radial velocity of the transmitter to microphone  $i$ .

With known DGPS measurements and microphone positions, it is possible to compute the Doppler shift. For each microphone, the transmission delay  $\Delta^{(i)}(t)$  is computed by iterating (18) until convergence, as given by

$$\Delta^{(i)}(t) \leftarrow r^{(i)}(t - \Delta^{(i)}(t))/c, \quad (22)$$

initialized as  $\Delta^{(i)}(t) \equiv 0 \forall t$ . Since the target speed is subsonic, the right hand side is a contraction with respect to  $\Delta^{(i)}$ , hence the iterations are guaranteed to converge by Banach's fixed-point theorem [23]. Then,  $\dot{\Delta}^{(i)}(t)$  can be computed using (21), where the speed  $v^{(i)}(t)$  is computed through numerical differentiation with spline interpolation of the DGPS positions.

##### C. Reference Estimation

A fundamental problem with evaluation of the performance of different frequency tracking methods using the experimental data detailed above is that the ground truth, *i.e.*, the source frequency  $f^{(0)}(t)$  emitted by the ATV engine, is unavailable. In order to evaluate the frequency tracking methods with the experimental data, this reference frequency must be estimated.

This can be done using the fact that all measurements originate from the same acoustic sound signal with continuously varying source frequency. However, the measurements are frequency shifted by the Doppler effect, and delayed due to the transmission delay from the ATV to the microphone in question.

The values of the source frequency  $f^{(0)}$  at particular times  $f_k^{(0)} = f^{(0)}(t_k - \Delta_k^{(i)})$  are of interest, where  $\Delta_k^{(i)} = \Delta^{(i)}(t_k)$ . Let  $\bar{t}_k = t_k - \Delta_k^{(i)}$  and  $\tilde{t}_k = \bar{t}_{k+1} - \bar{t}_k$ , and note that the samples are not equidistant in time. Assuming a ‘‘constant velocity’’ model for the source frequency, its discrete-time state-space formulation with state  $z_k^{(0)} = [f_k^{(0)} \dot{f}_k^{(0)}]^\top$  is given by

$$z_{k+1}^{(0)} = \begin{bmatrix} 1 & \tilde{t}_k \\ 0 & 1 \end{bmatrix} z_k^{(0)} + v_k^{(0)}, \quad (23a)$$

$$f_k^{(i)} = (1 - \dot{\Delta}_k^{(i)}) [1 \quad 0] z_k^{(0)} + e_k^{(i)}, \quad 1 \leq i \leq S, \quad (23b)$$

where the outputs  $f_k^{(i)}$  of the considered frequency tracking algorithm from microphone  $i$  are considered measurements of the Doppler shifted and delayed source frequency, and  $S$  is the total number of microphones. The frequency tracker errors  $e_k^{(i)}$  are approximated as independent Gaussian white noise signals for  $i = 1, \dots, S$ . The process noise  $v_k^{(0)}$  is Gaussian and white with covariance

$$Q_k = \text{cov } v_k^{(0)} = \sigma_{cv}^2 \begin{bmatrix} \tilde{t}_k^3/3 & \tilde{t}_k^2/2 \\ \tilde{t}_k^2/2 & \tilde{t}_k \end{bmatrix},$$

where  $\sigma_{cv}^2$  is a tuning parameter controlling the variability of the source frequency. The values  $\Delta_k^{(i)}$  and  $\dot{\Delta}_k^{(i)}$  are computed using the DGPS measurements as described in the previous section. For evaluation purposes, it is assumed that they are known without uncertainty. A Kalman smoother as described in [24] is applied to this model, and the posterior mean estimate

$$\hat{f}_k^{(0)} = \mathbb{E}[f_k^{(0)} | F^{(1)}, \dots, F^{(S)}] \quad (24)$$

of the source frequency given the set of all frequency tracker estimate time series  $F^{(1)}, \dots, F^{(S)}$ , where  $F^{(i)} = \{f_k^{(i)}\}_{k=1}^N$ , is obtained. Then, for evaluation, the expected frequency is compared to the sensor value for microphone  $i$  using the *root mean square error* (RMSE) value

$$\widehat{\text{RMSE}}(f_k^{(i)}) = \sqrt{\frac{1}{N} \sum_{k=1}^N \left( f_k^{(i)} - (1 - \dot{\Delta}_k^{(i)}) \hat{f}_k^{(0)} \right)^2}, \quad (25)$$

where the  $\hat{\phantom{x}}$  notation is used since the comparison is not made to a measured ground truth value.

This comparison is performed for each frequency tracker individually. That is, each frequency tracker outputs a set of time series  $F^{(1)}, \dots, F^{(S)}$ , and a separate posterior mean  $\hat{f}_k^{(0)}$  and corresponding  $\widehat{\text{RMSE}}$  is computed for each frequency tracker from those values. This is in order to make the comparisons equitable for different methods.

#### D. Experimental Results

The three tracking methods (RBPMF, DPM, and SHS) described above were compared and tested on experimental data using the parameters shown in Table I. An example of the

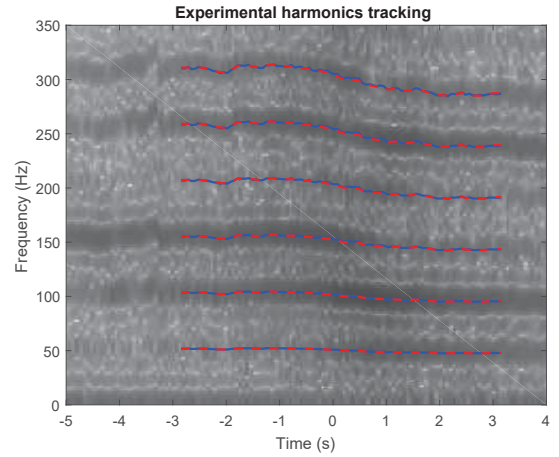


Fig. 4. An example of a comparison of a frequency estimate (blue) to an estimated expected frequency (red) computed as detailed in Sec. IV-C. Since the frequency tracking methods consider 6 harmonics, 6 multiples of the curves are illustrated for easy comparison with the underlying spectrogram, where darker shades indicate more energy at that time and frequency band. The method accurately captures the fundamental frequency, since the blue and red curves almost coincide, and agree with the energy peaks in the spectrogram.

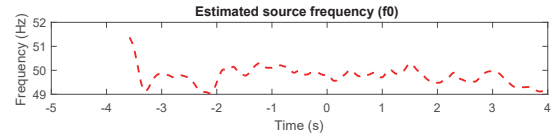


Fig. 5. Estimated source frequency from the same data sequence as Fig. 4. It can be seen that the source frequency is clearly varying.

experimental results is given in Fig. 4, where the spectrogram of the experimental microphone data is shown. Superimposed is the output of the SHS algorithm and the reference frequency computed as described in Sec. IV-C. The estimated source frequency is shown in Fig. 5, and it can be seen that the variation of the source frequency is comparable in magnitude to the Doppler frequency shift, motivating the relevance of estimating the source frequency. All results use a batch length of  $N_b = 80$  and  $M = 6$  harmonics unless otherwise noted, and for evaluation,  $\sigma_{cv}^2 = 10^3$ .

In Fig. 6, the experimental error of the three algorithms in comparison to the estimated reference frequency is shown for one data sequence and one sensor. It can be observed that the estimation error for the RBPMF has the lowest variation, although all methods appear to be biased around  $t = 0$  s. For the same data sequence, Table II shows estimated RMSE values for the different sensors as determined in Sec. IV-C; the values are presented for 4 sensors for brevity, but the results are representative for all 9 sensors. For all sensors in this data sequence, the RBPMF indeed has the lowest RMSE, and on average it is more than 25% lower than the periodogram-based methods. Clearly, in this case tracking performance is gained when the frequency is considered inherently time-varying.

In order to evaluate the methods independently of the source frequency estimation, the sensor array A can be used. Since its four microphones are located in essentially the same place,

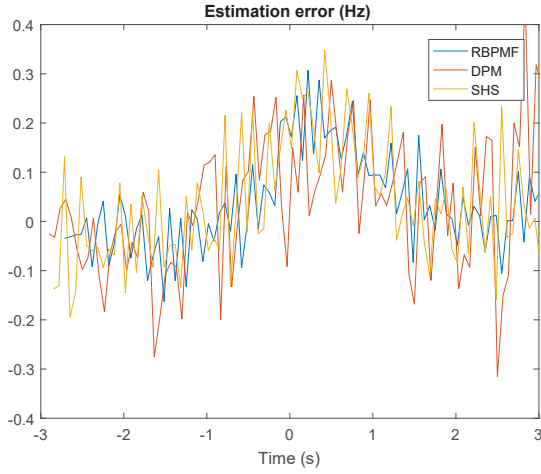


Fig. 6. Estimation errors for frequency for the three compared methods over an experimental data set. Note that the RBPMF does not provide frequency estimates for a few fractions of a second until its frequency estimate has sufficiently low variance.

TABLE I  
PARAMETER VALUES USED FOR THE DIFFERENT ALGORITHMS. FOR MORE DETAIL ABOUT DPM, SEE [15].

RBPMF Par.	Value	DPM Par.	Value	SHS Par.	Value
$L_{\text{RBPMF}}$	250	$N_b$	80	$N_b$	80
$\text{cov}(e_k)$	1	$c_p$	$1 -  p /2$	$c_m$	$0.99^m$
$\text{cov}(v_k^\omega)$	$10^4$	$\omega_{\min}$	$2\pi \cdot 40$		
$\text{cov}(v_k^{\alpha(m)})$	0.1	$\omega_{\max}$	$2\pi \cdot 60$		
$\text{cov}(v_k^{\beta(m)})$	0.1				
$\omega^{(j)}$	$2\pi(20 + 40j/J)$				
$J$	250				

they experience the same Doppler shift and the estimates are directly comparable. RMSE values in this case are shown in Table III, where similar results as in the previous paragraph can be seen. The relative ranking of the methods are the same, with the RBPMF outperforming the other methods by 35%, although the absolute errors are smaller. This may be because parts of the noise (e.g., the wind noise) affecting the microphones are correlated, biasing the RMSE.

In the case that frequency estimates are not required as often, the batch length may be set to, e.g.,  $N_b = 600$  instead of  $N_b = 80$  which was used above. In this case, the methods are more comparable, as seen in Table IV, with the SHS method having the best RMSE. One explanation for this is that a large amount of the performance degradation for the periodogram-based methods in Tables II–III is due to them ignoring phase and amplitude coherence between batches. This is less detrimental when the batch size  $N_b$  is large enough, since then, there is more information identifying the phase and amplitude of each harmonic in the data.

## V. SIMULATION

In order to evaluate the algorithms in a controlled environment performed in a similar scenario to Sec. IV-A,

TABLE II  
EXPERIMENTAL RMSE RESULTS (Hz) FOR ONE TRAJECTORY AND 4 DIFFERENT MICROPHONES M1–M4. THE SOURCE FREQUENCY WAS ESTIMATED USING ALL 12 MICROPHONES, HOWEVER, ONLY 4 RMSE VALUES ARE PRESENTED FOR BREVITY.

	M1	M2	M3	M4
RBPMF	0.11	0.11	0.11	0.11
DPM	0.15	0.14	0.13	0.15
SHS	0.14	0.14	0.14	0.14

TABLE III  
EXPERIMENTAL RMSE RESULTS (Hz) FOR ONE TRAJECTORY AND A MICROPHONE ARRAY OF 4 MICROPHONES LOCATED IN THE SAME SENSOR ARRAY A. IN THIS CASE THE ESTIMATION OF SOURCE FREQUENCY IS NOT REQUIRED, SINCE THE DOPPLER EFFECT IS THE SAME FOR ALL OF THEM.

	A1	A2	A3	A4
RBPMF	0.07	0.07	0.07	0.07
DPM	0.12	0.11	0.11	0.12
SHS	0.11	0.11	0.12	0.10

50 simulations were run. In the simulations, a vehicle with constant source frequency 50 Hz has been simulated like it was driving along the trajectory shown in Fig. 3. Output measurements from each microphone location in the figure were generated; thus, only one microphone was simulated located at the array position A, yielding 9 locations in total. Using physical principles as described in Sec. IV-B and in more detail in [15], the measurements  $y_k^{(i)}$  of microphone  $k$  are computed according to

$$y_k^{(i)} = \sum_{m=1}^M \frac{A_k^{(m,i)}}{r^{(i)}(t_k)} \cos(m\omega_0 \bar{t}_k + \phi_k^{(m,i)}) + e_k^{(i)}, \quad (26a)$$

where  $\bar{t}_k$ ,  $\tilde{t}_k$ , and  $r^{(i)}$  have been defined in Sec. IV-C. With a perfect source,  $A_k^{(m,i)}$  and  $\phi_k^{(m,i)}$  may be taken as constant initial phases. However, in order to take imperfections in the generating process into account, the amplitudes and phase offsets are allowed to change according to

$$A_{k+1}^{(m,i)} = (1 - \lambda)A_k^{(m,i)} + \lambda \left( A_0^{(m,i)} + \sqrt{\tilde{t}_k} v_k^A{}^{(m,i)} \right), \quad (26b)$$

$$\phi_{k+1}^{(m,i)} = \phi_k^{(m,i)} + \sqrt{\tilde{t}_k} v_k^\phi{}^{(m,i)}, \quad (26c)$$

where  $e_k^{(i)}$ ,  $v_k^A{}^{(m,i)}$ , and  $v_k^\phi{}^{(m,i)}$  are Gaussian white noise sequences. The parameters used are shown in Table V.

TABLE IV  
EXPERIMENTAL RMSE RESULTS (Hz) FOR ONE TRAJECTORY AND 4 DIFFERENT MICROPHONES M1–M4. THE SOURCE FREQUENCY WAS ESTIMATED USING ALL 12 MICROPHONES, HOWEVER, ONLY 4 RMSE VALUES ARE PRESENTED FOR BREVITY. FOR THESE VALUES,  $N_b = 600$  INSTEAD OF  $N_b = 80$ .

	M1	M2	M3	M4
RBPMF	0.053	0.053	0.049	0.051
DPM	0.063	0.049	0.058	0.061
SHS	0.045	0.045	0.045	0.042

TABLE V  
PARAMETER VALUES USED FOR SIMULATED DATA GENERATION.

Parameter	Value
$\omega_0$	50
$M$	6
$\text{cov}(v_k^{A(m,i)})$	$10^{-4}$
$\text{cov}(v_k^{\phi(m,i)})$	$10^{-4}$
$\text{cov}(e_k^{(i)})$	$10^{-3}$
$\lambda$	$10^{-3}$
$A_0^{(m,i)}$	0.2
$\phi_0^{(m,i)}$	0

TABLE VI  
SIMULATION RMSE RESULTS (Hz) OVER 50 NOISE REALIZATIONS. IT CAN BE SEEN THAT THE RBPMF ACHIEVES THE LOWEST RMSE. Avg. DENOTES THE AVERAGE RMSE OVER ALL 9 MICROPHONES.

	M1	M2	M3	M4	Avg.
RBPMF	0.057	0.058	0.068	0.058	0.058
DPM	0.169	0.171	0.172	0.177	0.180
SHS	0.139	0.176	0.150	0.161	0.209

### A. Simulation Results

The RMSE values from the simulation are given in Table VI. Notice that the RMSE magnitudes and the relative ordering of the methods are similar to the experimental evaluation, indicating that the simulation parameters are realistic. The RBPMF again has the best performance as quantified by the RMSE. This can be explained by the fact that the RBPMF considers phase and amplitude coherence over time, and inherently allows for variability in the source frequency. Since the batch length is  $N_b = 80$ , phase and amplitude coherence appears to be useful in order to improve the accuracy of the estimate. However, note that the RBPMF implementation in MATLAB requires roughly 100 times more computational effort than the corresponding SHS or DPM implementations, although it also provides an estimate 80 times more often.

In Fig. 7, the average RMSE over all sensors and all simulation runs is plotted against time. It can be seen that the RMSE is smallest in the middle of the trajectory. This is because the engine is closest to the microphones at this point in time, so that the SNR is highest then. This effect is most pronounced for the SHS method, and one explanation for this could be that its frequency interpolation is better than the DPM method. The RBPMF can again be seen to have the lowest average RMSE for each time instant as compared to its competitors; however, the decrease in the middle is less pronounced.

## VI. CONCLUSION

The problem of acoustic frequency tracking of a harmonic signal has been considered. The RBPMF, previously proposed for frequency tracking in a different domain (vibration tracking), has now successfully been applied for acoustic signals. Experimental and simulation results indicate that the RBPMF outperforms the other, periodogram-based methods in RMSE. Furthermore, the theoretical connections between the different

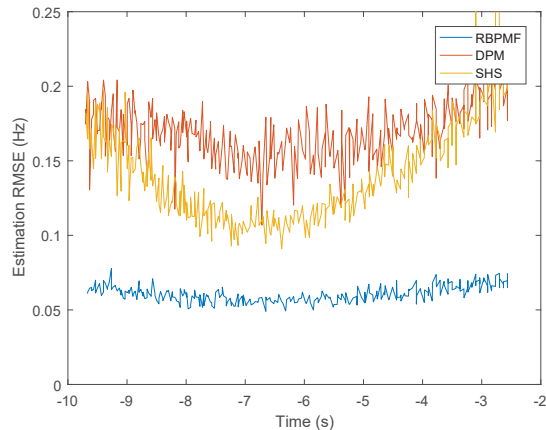


Fig. 7. Illustration of RMSE over time in simulation for 50 noise realizations, averaged over all microphones. It can be seen that the RBPMF has smaller errors than the alternative, periodogram-based approaches. The error is the lowest when the SNR is the highest, *i.e.*, when the simulated vehicle is closest to the microphone. This effect is less pronounced for the RBPMF due to its transient bias and its higher performance in lower SNR conditions.

methods have been explained in order to put them in a larger context. Future work includes investigating whether the possible improvements detailed in this paper can improve localization performance when used with the methods shown in [15].

## ACKNOWLEDGEMENTS

This work was partially supported by the Wallenberg Autonomous Systems Program (WASP). Thanks to David Lindgren (Swedish Defence Research Agency, FOI) for providing measurement data, database code, and a reference method.

## REFERENCES

- [1] B. G. Quinn and E. J. Hannan, *The estimation and tracking of frequency*. Cambridge University Press, 2001, vol. 9.
- [2] P. Stoica, P. Händel, and T. Söderström, "Approximate maximum likelihood frequency estimation," *Automatica*, vol. 30, no. 1, pp. 131–145, 1994.
- [3] D. Hermes, "Measurement of pitch by subharmonic summation," *Journal of the Acoustical Society of America*, vol. 83, no. 1, pp. 257–264, 1988.
- [4] R. Roy and T. Kailath, "ESPRIT — estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 7, pp. 984–995, 1989.
- [5] A. de Cheveigne and H. Kawahara, "YIN, a fundamental frequency estimator for speech and music," *Journal of the Acoustical Society of America*, vol. 111, no. 4, pp. 1917 – 1930, 2002.
- [6] R. C. Williamson, B. James, B. D. Anderson, and P. J. Kootsookos, "Threshold effects in multiharmonic maximum likelihood frequency estimation," *Signal Processing*, vol. 37, no. 3, pp. 309–331, 1994.
- [7] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, ser. Mathematics in Science and Engineering. Academic Press, Inc, 1970, vol. 64.
- [8] R. Kumar, "Optimum filters and smoothers design for carrier phase and frequency tracking," NASA Jet Propulsion Laboratory, Tech. Rep., 1987, JPL publication 87-10.
- [9] P. J. Parker and B. D. Anderson, "Frequency tracking of nonsinusoidal periodic signals in noise," *Signal Processing*, no. 20, pp. 127 – 152, 1990.
- [10] B. La Scala and R. Bitmead, "Design of an extended Kalman filter frequency tracker," *IEEE Transactions on Signal Processing*, vol. 44, no. 3, pp. 739 – 742, 1996.
- [11] J. Tabrikian, S. Dubnov, and Y. Dickalov, "Maximum a-posteriori probability pitch tracking in noisy environments using harmonic model,"

- IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 12, no. 1, pp. 76–87, 2004.
- [12] S. Kim, L. Holmstrom, and J. McNames, “Multiharmonic tracking using marginalized particle filters,” in *Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Vancouver, Canada, Aug. 2008, pp. 29–33.
- [13] W. Ng, J. Chunlin, W.-K. Ma, and H. C. So, “A study on particle filters for single-tone frequency tracking,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 45, no. 3, pp. 1111 – 1125, 2009.
- [14] M. Lindfors, G. Hendeby, F. Gustafsson, and R. Karlsson, “Vehicle Speed Tracking Using Chassis Vibrations,” in *Proceedings of the IEEE Intelligent Vehicles Symposium (IV)*, Gothenburg, Sweden, 2016, pp. 214–219.
- [15] D. Lindgren, G. Hendeby, and F. Gustafsson, “Distributed localization using acoustic Doppler,” *Signal Processing*, vol. 107, Feb. 2015.
- [16] F. Lindsten, “Rao-Blackwellised particle methods for inference and identification,” Licentiate Thesis No 1480, Department of Electrical Engineering, Linköpings universitet, Sweden, 2011.
- [17] V. Smidl and M. Gasperin, “Rao-Blackwellized point mass filter for reliable state estimation,” in *Proceedings of 16th IEEE International Conference on Information Fusion*, Istanbul, Turkey, Jul. 9–12 2013, pp. 312–318.
- [18] R. S. Bucy and K. D. Senne, “Digital synthesis of non-linear filters,” *Automatica*, vol. 7, no. 3, pp. 287–298, 1971.
- [19] Z. Chen, “Bayesian filtering: From Kalman filters to particle filters, and beyond,” *Statistics*, vol. 182, no. 1, pp. 1–69, 2003.
- [20] R. E. Kalman, “A new approach to linear filtering and prediction problems,” *Transactions of the American Society of Mechanical Engineering — Journal of Basic Engineering, Series D*, vol. 82, pp. 35–45, Mar. 1960.
- [21] V. Deleskog, H. Habberstad, G. Hendeby, and D. Lindgren, “Acoustic and visual sensor network measurements in Lilla Gåra 2012,” The Swedish Defence Research Agency (FOI), Linköping, Sweden, Tech. Rep., 2014, MEMO4833.
- [22] R. H. Randall, *An introduction to acoustics*. Courier Corporation, 2012.
- [23] E. Kreyszig, *Introductory functional analysis with applications*. Wiley New York, 1989, vol. 1.
- [24] S. Särkkä, *Bayesian filtering and smoothing*. Cambridge University Press, 2013, vol. 3.