Estimating Uncertainty in Time-difference and Doppler Estimates

Gabrielle Flood, Anders Heyden and Kalle Åström
Centre for Mathematical Sciences, Lund University, Lund, Sweden

Keywords: Time-difference of Arrival, Sub-sample Methods, Doppler Effect, Uncertainty Measure.

Abstract: Sound and radio can be used to estimate the distance between a transmitter and a sender by correlating the emitted and received signal. Alternatively by correlating two received signals it is possible to estimate distance difference. Such methods can be divided into methods that are robust to noise and reverberation, but give limited precision and sub-sample refinements that are sensitive to noise, but give higher precision when initialized close to the real translation. In this paper we develop stochastic models that can explain the limits in the precision of such sub-sample time-difference estimates. Using such models we provide new methods for precise estimates of time-differences as well as Doppler effects. The method is verified on both synthetic and real data.

1 INTRODUCTION

Audio and radio sensors are becoming ubiquitous among our everyday tools, e.g. smartphones, laptops, and tablet pc's. They also form the backbone of internet-of-things, e.g. small low-power units that can run for years on batteries or use energy harvesting to run for extended periods. Given the location of each one of these units, it is possible to use them as an ad-hoc acoustic sensor network. Such sensor networks can be used for many interesting applications. One use-case is localization, cf. (Brandstein et al., 1997; Cirillo et al., 2008; Cobos et al., 2011; Do et al., 2007). Another application is to improve the sound quality using so called beam-forming, (Anguera et al., 2007). A third application is so called speaker diarization, i.e. to determine who spoke when, (Anguera et al., 2012). If the sensor positions are unknown or only known to a certain accuracy, the results of such applications are inferior as is shown in (Plinge et al., 2016). However, even without any prior information it is possible to estimate both sender and receiver positions up to a choice of coordinate system, (Pollefeys and Nister, 2008; Crocco et al., 2012; Kuang et al., 2013; Kuang and Åström, 2013a; Zhayida et al., 2014), thus providing accurate sensor positions. A key component for all of these methods is the process of obtaining features such as time-difference estimates from pairwise channels. In this paper we will primarily focus on sound. However the same principles are applicable for radio (Batstone et al., 2016).

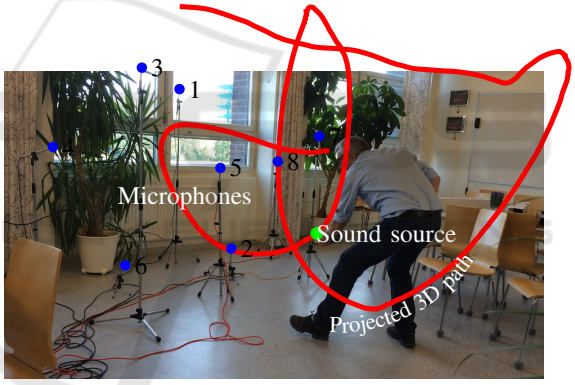

Figure 1: Precise time-difference of arrival estimation can be used for many purposes, e.g. beam-forming, diarization, anchor free node calibration and positioning. The figure illustrates its use for anchor free node calibration, sound source movement and room reconstruction.

All of these applications depend on accurate ways of comparing the sound (or radio) signals so as to extract information. The most common approach is to estimate time-differences, which are then used for subsequent processing. For the applications it is of interest to obtain the most precise estimate possible. In (Zhayida and Åström, 2016) sub-sample methods were used to improve on the time-difference estimates and it is empirically shown to give better estimates of the receiver-sender configurations. However, no analysis of the sub-sample time-difference uncertainties was provided.

Flood, G., Heyden, A. and Åström, K.
Estimating Uncertainty in Time-difference and Doppler Estimates.
DOI: 10.5220/0006722002450253
In *Proceedings of the 7th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2018)*, pages 245-253
ISBN: 978-989-758-276-9

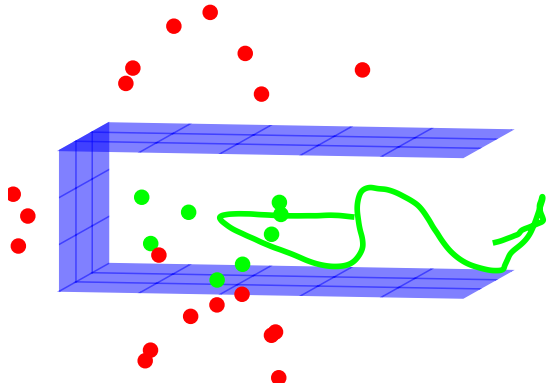

Figure 2: The figure examplifies one usage of precise time-difference of arrival estimation. The image illustrates the estimated microphone positions (green dots), estimated mirrored microphone positions (red dots) and sound source motion (green curve) from Fig 1. The estimated reflective planes are also shown in the figure. These three planes correspond to the floor, the ceiling and the wall.

The main contributions of this paper are:

- A framework for estimating time-difference estimates and for estimating the precision of such time-difference estimates.
- An extension of the method to also estimate minute Doppler effects.
- In practice there is also significant difference in amplitude of the signals. Accurate results require that also these amplitude changes are estimated and accounted for.
- A synthetic evaluation that demonstrates both the validity of the models, but also provides knowledge on the failure modes of the method.
- Evaluation on real data that demonstrate that there is useful information not only in the time-difference estimates but also in the amplitude changes and in the minute Doppler effects, even for speeds as small as 0.1 m/s.

2 MODELING PARADIGM

2.1 Measurement and Error Model

In this paper, we study discretely sampled signals, such as audio signals. We assume that signals have been sampled at a fixed and known sampling rate. A reasonable measurement model is that the measured signal y is the result of ideal sampling and added noise, i.e.

$$y(k) = Y(k) + e(k). \tag{1}$$

Here $Y : \mathbb{R} \mapsto \mathbb{R}$ is the original, continuous signal and e is a discrete stationary stochastic process.

Let $\mathbb{B}$ denote the set of continuous functions $Y : \mathbb{R} \to \mathbb{R}$ that are also square integrable and such that the Fourier transform is zero outside the interval $[-\pi, \pi]$. Let ℓ denote the set of discrete functions $y : \mathbb{Z} \to \mathbb{R}$ that are also square integrable. For such functions we introduce the discretization operator $D : \mathbb{B} \to \ell$, such that

$$y(i) = D(Y)(i) = Y(i).$$

We also introduce the interpolation operator $I_g : \ell \to \mathbb{B}$, such that

$$Y(x) = I_g(y)(x) = \sum_{i=-\infty}^{\infty} g(x-i)y(i).$$

The ideal interpolation operator $I : \ell \to \mathbb{B}$ is interpolation using the normalized sinc function, with $g(x) = \mathrm{sinc}(x)$. We have that ideal interpolation restores the sampled function, (Shannon, 1949), i.e.

$$I(D(Y)) = Y.$$

Other interpolation methods can often be written in a similar way. E.g. by substituting sinc by

$$G_a(x) = \frac{1}{\sqrt{2\pi a^2}} e^{x^2/2a^2}$$

we get Gaussian interpolation.

2.2 Scale-space Smoothing and Ideal Interpolation

The measured and interpolated signal can be smoothed to decrease the impact of measurement noise. This also makes it easier to capture patterns on a coarser scale, (Lindeberg, 1994). We have used a Gaussian kernel G_{a_2}, with standard deviation a_2, for the smoothing. Later, a_2 will also be referred to as the *smoothing parameter*.

Given a sampled signal y, the ideally interpolated and smoothed signal can be expressed

$$Y(x) = (G_{a_2} * I_{\mathrm{sinc}}(y))(x) = I_{G_{a_2} * \mathrm{sinc}}(y)(x).$$

For a sufficiently large a_2 we have $G_{a_2} * \mathrm{sinc} \approx G_{a_2}$. Thus ideal interplation followed by Gaussian smoothing can be approximated to interpolation using the Gaussian (Åström and Heyden, 1999), s.t.

$$Y(x) = I_{G_{a_2} * \mathrm{sinc}}(y)(x) \approx (I_{G_{a_2}}(y))(x). \tag{2}$$

What *sufficiently large* means will be discussed later, in Section 4.1.

Furthermore we will use that discrete w.s.s. Gaussian noise interpolates to continuous w.s.s. Gaussian noise and vice versa, as is shown in (Åström and Heyden, 1999).

3 TIME-DIFFERENCE AND DOPPLER ESTIMATION

Assume that we have two signals, $W(t)$ and $\bar{W}(t)$. The signals are measured and interpolated as described in the previous section. Also assume that the two signals are similar, but not identical. This can e.g. arise when a single audio signal is picked up by two different receivers. Then, one of the signals can be described by the other and a few parameters. We describe the relation as follows

$$W(t) = \bar{W}(\alpha t + h). \tag{3}$$

Here, h is a translation or the time-difference of arrival. If the distances from the sound source to the two microphones are the same, $h = 0$. The second parameter, α, is a Doppler factor. This is interesting if either the sound source or the microphones are moving. For a stationary setup $\alpha = 1$.

When the two signals are picked up by the microphones they are disturbed by Gaussian w.s.s. noise. Thus, the received signals are better described by

$$V(t) = W(t) + E(t), \quad \bar{V}(t) = \bar{W}(t) + \bar{E}(t), \tag{4}$$

where $E(t)$ and $\bar{E}(t)$ denotes the two independent noise signals after interpolation.

Now, assume that the signals V and $\bar{V}$ are given. Also, denote $\boldsymbol{z} = \begin{bmatrix} z_1 & z_2 \end{bmatrix}^T = \begin{bmatrix} h & \alpha \end{bmatrix}^T$. Then, the parameters for which (3) is true can be estimated by the $\boldsymbol{z}$ that minimizes the integral

$$F(\boldsymbol{z}) = \int_t (V(t) - \bar{V}(z_2 t + z_1))^2 \, \mathrm{d}t. \tag{5}$$

3.1 Estimating Standard Deviation of the Parameters

If $\boldsymbol{z}_T = \begin{bmatrix} h_T & \alpha_T \end{bmatrix}^T$ is the true parameter and $\hat{\boldsymbol{z}}$ is the parameter that has been estimated by (5), the estimation error can be expressed as

$$X = \hat{\boldsymbol{z}} - \boldsymbol{z}_T.$$

Assume, without loss of generality, that $\boldsymbol{z}_T = \begin{bmatrix} 0 & 1 \end{bmatrix}^T$. The standard deviation of $\hat{\boldsymbol{z}}$ will be the same as the standard deviation of X and the mean of those two will only differ by $\boldsymbol{z}_T$. Thus, we can study X to get statistical information about $\hat{\boldsymbol{z}}$.

Linearizing $F(\boldsymbol{z})$ around the true displacement $\boldsymbol{z}_T = \begin{bmatrix} 0 & 1 \end{bmatrix}^T$ we get

$$F(X) = \frac{1}{2} X^T a X + bX + f,$$

with

$$a = \nabla^2 F(\boldsymbol{z}_T), \qquad b = \nabla F(\boldsymbol{z}_T), \qquad f = F(\boldsymbol{z}_T).$$

Using (4) and (3), this gives

$$\begin{aligned} f =& F(\begin{bmatrix} 0 & 1 \end{bmatrix}^T) = \int_t (V(t) - \bar{V}(1 \cdot t - 0))^2 \, \mathrm{d}t \\ =& \int_t (W(t) + E(t) - (\bar{W}(t) + \bar{E}(t)))^2 \, \mathrm{d}t \\ =& \int_t (E - \bar{E})^2 \, \mathrm{d}t = \int_t E^2 + 2E\bar{E} + \bar{E}^2 \, \mathrm{d}t. \end{aligned}$$

Straightforward calculations give

$$b = -2 \begin{bmatrix} \int_t \hat{\phi} \, \mathrm{d}t \\ \int_t t \hat{\phi} \, \mathrm{d}t \end{bmatrix},$$

with

$$\hat{\phi} = E\bar{W}' + E\bar{E}' - \bar{E}\bar{W}' - \bar{E}\bar{E}'.$$

Furthermore,

$$\nabla^2 F = \begin{bmatrix} \int_t \phi \, \mathrm{d}t & \int_t t\phi \, \mathrm{d}t \\ \int_t t\phi \, \mathrm{d}t & \int_t t^2 \phi \, \mathrm{d}t \end{bmatrix},$$

using

$$\begin{aligned} \phi(\boldsymbol{z}) = &(\bar{V}'(\alpha t + h))^2 - \\ &V(t)\bar{V}''(\alpha t + h) + \bar{V}(\alpha t + h)\bar{V}''(\alpha t + h). \end{aligned}$$

If we let

$$\begin{aligned} \hat{\phi} =& \phi(\boldsymbol{z}_T) = (\bar{W}'(t) + \bar{E}'(t))^2 \\ &- (W(t) + E(t))(\bar{W}''(t) + \bar{E}''(t)) \\ &+ (\bar{W}(t) + \bar{E}(t))(\bar{W}''(t) + \bar{E}''(t)) \\ =& (\bar{W}')^2 + 2\bar{W}'\bar{E}' + (\bar{E}')^2 - E\bar{W}'' \\ &- E\bar{E}'' + \bar{E}\bar{W}'' + \bar{E}\bar{E}'' \end{aligned}$$

we get

$$a = \nabla^2 F(\boldsymbol{z}_T) = \begin{bmatrix} \int_t \hat{\phi} \, \mathrm{d}t & \int_t t\hat{\phi} \, \mathrm{d}t \\ \int_t t\hat{\phi} \, \mathrm{d}t & \int_t t^2 \hat{\phi} \, \mathrm{d}t \end{bmatrix}.$$

We have that $F(X) = 1/2 \cdot X^T a X + bX + f$. To minimize this function, one should find the X for which the derivative of $F(X)$ is zero.

$$\nabla F(X) = aX + b = 0 \quad \Leftrightarrow \quad X = g(a,b) = -a^{-1}b.$$

In the calculations we will assume that a is invertible.

Now we want to find the mean and covariance of X. For this, Gauss' approximation formula is used. If we denote the expected value of a and b with $\mu_a = \mathbf{E}[A]$ and $\mu_b = \mathbf{E}[b]$ respectively the expected value of X can be approximated to

$$\begin{aligned} \mathbf{E}[X] =& \mathbf{E}[g(a,b)] \approx \mathbf{E}[g(\mu_a,\mu_b) + (a - \mu_a)g'_a(\mu_a,\mu_b) \\ &+ (b - \mu_b)g'_b(\mu_a,\mu_b)] \\ =& g(\mu_a,\mu_b) + (\mathbf{E}[a] - \mu_a)g'_a(\mu_a,\mu_b) \\ &+ (\mathbf{E}[b] - \mu_b)g'_b(\mu_a,\mu_b) \\ =& g(\mu_a,\mu_b) = -\mu_a^{-1}\mu_b = -\mathbf{E}[a]^{-1}\mathbf{E}[b]. \end{aligned} \tag{6}$$

In the same manner the covariance of X is

$$\begin{aligned}\mathbf{C}[X] =&\mathbf{C}[g(a,b)] \approx g'_a(\mu_a,\mu_b)^T\mathbf{C}[a]g'_a(\mu_a,\mu_b)\\ &+ g'_b(\mu_a,\mu_b)^T\mathbf{C}[b]g'_b(\mu_a,\mu_b)\\ &+ 2g'_a(\mu_a,\mu_b)^T\mathbf{C}[a,b]g'_b(\mu_a,\mu_b),\end{aligned} \tag{7}$$

where $\mathbf{C}[a,b]$ denotes the cross-covariance between a and b. For further computations $g'_a(a,b) = -(a^{-1})^2 b$, $g'_b(a,b) = -a^{-1}$, $\mathbf{E}[a]$, $\mathbf{E}[b]$, $\mathbf{C}[b]$ and $\mathbf{C}[a,b]$ are needed.

By computing the expected value of $\hat{\varphi}$

$$\begin{aligned}\mathbf{E}[\hat{\varphi}] =&\mathbf{E}[E\bar{W}' + E\bar{E}' - \bar{E}\bar{W}' - \bar{E}\bar{E}']\\ =&\mathbf{E}[E]\bar{W}' + \mathbf{E}[E]\mathbf{E}[\bar{E}'] - \mathbf{E}[\bar{E}]\bar{W}' - \mathbf{E}[\bar{E}]\mathbf{E}[\bar{E}']\\ =&0\end{aligned}$$

we get

$$\begin{aligned}\mathbf{E}[b] =&\mathbf{E}\left[-2\begin{bmatrix}\int_t \hat{\varphi}\, dt\\ \int_t t\hat{\varphi}\,\mathrm{d}t\end{bmatrix}\right] = -2\begin{bmatrix}\int_t \mathbf{E}[\hat{\varphi}]\,\mathrm{d}t\\ \int_t t\mathbf{E}[\hat{\varphi}]\,\mathrm{d}t\end{bmatrix}\\ =&-2\begin{bmatrix}\int_t 0\,\mathrm{d}t\\ \int_t t\cdot 0\,\mathrm{d}t\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix}.\end{aligned}$$

In the second step of the computation of $\mathbf{E}[\hat{\varphi}]$ we have used that for a weakly stationary process the process and its derivative at a certain time are uncorrelated, and thus $\mathbf{E}[\bar{E}\bar{E}'] = \mathbf{E}[\bar{E}]\mathbf{E}[\bar{E}']$, (Lindgren et al., 2013). Hence,

$$\mathbf{E}[X] = -\mathbf{E}[a]^{-1}\mathbf{E}[b] = -\mathbf{E}[a]^{-1}\begin{bmatrix}0\\0\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix}.$$

Since $\mathbf{E}[b] = \mathbf{0}$, we get that $g'_a(\mu_a,\mu_b) = \mathbf{0}$. Thus the first and the last term in (7) cancel, leaving

$$\begin{aligned}\mathbf{C}[X] =&g'_b(\mu_a,\mu_b)^T\mathbf{C}[b]g'_b(\mu_a,\mu_b)\\ =&(-\mathbf{E}[a]^{-1})^T\mathbf{C}[b](-\mathbf{E}[a]^{-1})\\ =&(\mathbf{E}[a]^{-1})^T\mathbf{C}[b]\mathbf{E}[a]^{-1}.\end{aligned} \tag{8}$$

To find the expected value of a we need the expected value of $\hat{\phi}$. This is

$$\begin{aligned}\mathbf{E}[\hat{\phi}] =&(\bar{W}')^2 + 2\bar{W}'\mathbf{E}[\bar{E}'] + \mathbf{E}[(\bar{E}')^2] - \bar{W}''\mathbf{E}[E]\\ &- \mathbf{E}[E]\mathbf{E}[\bar{E}''] + \bar{W}''\mathbf{E}[\bar{E}] + \mathbf{E}[\bar{E}\bar{E}'']\\ =&(\bar{W}')^2 + \mathbf{E}[(\bar{E}')^2] + \mathbf{E}[\bar{E}\bar{E}''] = (\bar{W}')^2.\end{aligned}$$

In the last equality we have used that $\mathbf{E}[\bar{E}\bar{E}''] = -\mathbf{E}[(\bar{E}')^2]$, (Lindgren et al., 2013). Thus, the two last terms cancel. The expected value of a is therefore

$$\begin{aligned}\mathbf{E}[a] =&2\begin{bmatrix}\int_t \mathbf{E}[\hat{\phi}]\,\mathrm{d}t & \int_t \mathbf{E}[t\hat{\phi}]\,\mathrm{d}t\\ \int_t \mathbf{E}[t\hat{\phi}]\,\mathrm{d}t & \int_t \mathbf{E}[t^2\hat{\phi}]\,\mathrm{d}t\end{bmatrix}\\ =&2\begin{bmatrix}\int_t (\bar{W}')^2\,\mathrm{d}t & \int_t t(\bar{W}')^2\,\mathrm{d}t\\ \int_t t(\bar{W}')^2\,\mathrm{d}t & \int_t t^2(\bar{W}')^2\,\mathrm{d}t\end{bmatrix}.\end{aligned}$$

Now, since the expected value of b is zero, the covariance of b is

$$\mathbf{C}[b] = (-2)^2\begin{bmatrix}C_{11} & C_{12}\\ C_{21} & C_{22}\end{bmatrix},$$

with

$$\begin{aligned}C_{11} &= \mathbf{E}\left[\int_{t_1}\hat{\varphi}(t_1)\,\mathrm{d}t_1\cdot\int_{t_2}\hat{\varphi}(t_2)\,\mathrm{d}t_2\right]\\ C_{12} &= \mathbf{E}\left[\int_{t_1}t_1\hat{\varphi}(t_1)\,\mathrm{d}t_1\cdot\int_{t_2}\hat{\varphi}(t_2)\,\mathrm{d}t_2\right]\\ C_{21} &= \mathbf{E}\left[\int_{t_1}\hat{\varphi}(t_1)\,\mathrm{d}t_1\cdot\int_{t_2}t_2\hat{\varphi}(t_2)\,\mathrm{d}t_2\right]\\ C_{22} &= \mathbf{E}\left[\int_{t_1}t_1\hat{\varphi}(t_1)\,\mathrm{d}t_1\cdot\int_{t_2}t_2\hat{\varphi}(t_2)\,\mathrm{d}t_2\right].\end{aligned}$$

Note that by changing the order of the terms in C_{21} it is clear that $C_{21} = C_{12}$. Furthermore we get

$$\begin{aligned}C_{11} =&\mathbf{E}\left[\int_{t_1}\hat{\varphi}(t_1)\,\mathrm{d}t_1\cdot\int_{t_2}\hat{\varphi}(t_2)\,\mathrm{d}t_2\right]\\ =&\mathbf{E}\Big[\Big(\int_{t_1}(E-\bar{E})(\bar{W}'+\bar{E}')\,\mathrm{d}t_1\Big)\\ &\cdot\Big(\int_{t_2}(E-\bar{E})(\bar{W}'+\bar{E}')\,\mathrm{d}t_2\Big)\Big]\\ =&\mathbf{E}\Big[\int_{t_1}\int_{t_2}(E(t_1)-\bar{E}(t_1))(\bar{W}'(t_1)+\bar{E}'(t_1))\\ &\cdot(E(t_2)-\bar{E}(t_2))(\bar{W}'(t_2)+\bar{E}'(t_2))\,\mathrm{d}t_1\mathrm{d}t_2\Big].\end{aligned}$$

Denoting $\mathbf{E}[(E(t_1)-\bar{E}(t_1))(E(t_2)-\bar{E}(t_2))] = r_{E-\bar{E}}(t_1-t_2)$ and assuming that $\mathbf{E}[\bar{E}'(t_1)\bar{E}'(t_2)]$ is small yields

$$\begin{aligned}C_{11} =&\mathbf{E}\left[\int_{t_1}\hat{\varphi}(t_1)\,\mathrm{d}t_1\cdot\int_{t_2}\hat{\varphi}(t_2)\,\mathrm{d}t_2\right]\\ =&\int_{t_1}\int_{t_2}\mathbf{E}[(E(t_1)-\bar{E}(t_1))(E(t_2)-\bar{E}(t_2))]\\ &\cdot(\bar{W}'(t_1)\bar{W}'(t_2)\\ &+\bar{W}'(t_1)\mathbf{E}[\bar{E}'(t_2)] + \mathbf{E}[\bar{E}'(t_1)]\bar{W}'(t_2)\\ &+\mathbf{E}[\bar{E}'(t_1)\bar{E}'(t_2)])\,\mathrm{d}t_2\mathrm{d}t_1\\ \approx&\int_{t_1}\int_{t_2}r_{E-\bar{E}}(t_1-t_2)\bar{W}'(t_1)\bar{W}'(t_2)\,\mathrm{d}t_2\mathrm{d}t_1\\ =&\int_{t_1}\bar{W}'(t_1)(\bar{W}'*r_{E-\bar{E}})(t_1)\,\mathrm{d}t_1.\end{aligned}$$

The time t is not stochastic and the other elements in $\mathbf{C}[b]$ can be computed similarly. Finally

$$\begin{aligned}C_{11} &= \int_t \bar{W}'(t)(\bar{W}'*r_{E-\bar{E}})(t)\,\mathrm{d}t\\ C_{12} = C_{21} &= \int_t t\bar{W}'(t)(\bar{W}'*r_{E-\bar{E}})(t)\,\mathrm{d}t\\ C_{22} &= \int_t t\bar{W}'(t)((t\bar{W}')*r_{E-\bar{E}})(t)\,\mathrm{d}t\end{aligned}$$

and through (8) an expression for the variance and thus also the standard deviation of X is found.

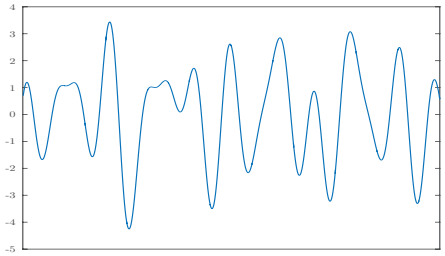

Figure 3: The simulated signal that was used for the experimental validation. Later, noise of different levels was added to achieve a more realistic signal.

3.2 Expanding the Model

The model (3) can easily be changed or expanded to contain more (or fewer) parameters. One example is the addition of an amplitude parameter γ, such that

$$W(t) = \gamma \bar{W}(\alpha t + h). \tag{9}$$

The integral (5) would then be changed accordingly and one would instead optimize over $\boldsymbol{z} = \begin{bmatrix} z_1 & z_2 & z_3 \end{bmatrix} = \begin{bmatrix} h & \alpha & \gamma \end{bmatrix}$.

In practice, the parameter computations will not be harder for more parameters. However the analysis carried out in the previous section does get more complex.

4 EXPERIMENTAL VALIDATION

For validation we use both synthetic data and real data. When we use synthetic data, the purpose is both to demonstrate the validity of the model and the approximations, but also to explore at what signal-to-noise ratio the approximations no longer hold. For the real data experiments, the purpose is to show that there is useful information in the estimated parameters. For time-difference this is well-known, but for the Doppler effects it is less known.

4.1 Synthetic Data - Validation of Method

Initially we tested the model on simulated data. This was done to investigate when the approximations in the model are valid. Examples of such are the linearization from Gauss' approximation formula, e.g. (3.1) and (7), and that ideal interpolation followed by convolution with a Gaussian can be approximated to Gaussian interpolation (2).

To do this we compared the theoretical standard deviations of the parameters calculated according to Section 3.1 with empirically computed standard deviations. While these agree our approximations are assumed to be valid.

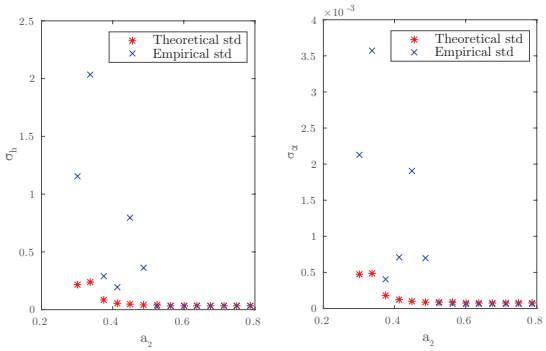

Figure 4: The plots show the standard deviation of the parameters in $\boldsymbol{z}$ for different values of the smoothing parameter a_2. The red stars ($*$) represent the theoretical values $\sigma_{\boldsymbol{z}}$ and the blue crosses (x) the empirical values $\hat{\sigma}_{\boldsymbol{z}}$. The left plot shows the results for the translation $z_1 = h$ and the right plot for the Doppler factor $z_2 = \alpha$. It is clear that the approximation is valid approximately when $a_2 > 0.55$.

An original continuous signal $W(x)$ was simulated, see Fig 3. The second signal was created according to (3) s.t. $\bar{W} = W(1/\alpha \cdot (x-h))$. Thereafter both signals were ideally sampled and then Gaussian white discrete noise with standard deviation σ_n was added to the discrete signals. During the analysis both signals were interpolated using a Gaussian kernel with standard deviation a_2, according to Section 2.2. At this point the signals could be described by $V(t)$ and $\bar{V}(t)$ as before.

To study the effect of a_2 and σ_n the signals V and $\bar{V}$ were re-simulated 1000 times using the same original signals W and $\bar{W}$, but different noise. The theoretical standard deviation of the parameter vector $\boldsymbol{z}$, $\sigma_{\boldsymbol{z}} = \begin{bmatrix} \sigma_h & \sigma_\alpha \end{bmatrix}$, was then computed according to the presented theory, while the empirical standard deviation, $\hat{\sigma}_{\boldsymbol{z}} = \begin{bmatrix} \hat{\sigma}_h & \hat{\sigma}_\alpha \end{bmatrix}$, was computed from the 1000 achieved parameter estimations.

First the effect of changing a_2 was studied. The noise level was set to $\sigma_n = 0.03$, the translation was $h = 3.63$ and the Doppler factor was set to $\alpha = 1.02$ – though the actual numbers are not of importance. The standard deviation of $\boldsymbol{z}$ was then computed according to above. This was done for several different $a_2 \in [0.3, 0.8]$.

The results are shown in Fig 4. It can be seen that the theoretical and empirical values $\sigma_{\boldsymbol{z}}$ and $\hat{\sigma}_{\boldsymbol{z}}$ disagree for both parameters when a_2 is lower than $a_2 \approx 0.55$. Thus the approximation (2) of ideal interpolation should only be used for a smoothing parameter $a_2 > 0.55$.

Secondly we investigated the impact of the noise.

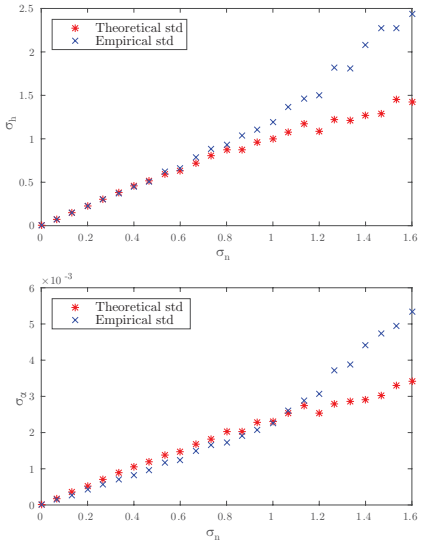

Figure 5: The standard deviation of the translation (on top) and Doppler factor (bottom) for different levels of noise in the signal. The red stars (∗) mark the theoretical values σ_z and the blue crosses (x) the empirical $\hat{\sigma}_z$. For the translation the values agree for signals with a noise level up to $\sigma_n \approx 0.8$. For the Doppler factor the theoretical values follow the empirical values for $\sigma_n < 1.1$.

The smoothing parameter was now fixed at $a_2 = 2$ and the translation and the Doppler factor were again set to $h = 3.63$ and $\alpha = 1.02$. Then σ_z and $\hat{\sigma}_z$ were computed as before. This time it was done for several different $\sigma_n \in [0, 1.6]$.

The results are shown in Fig 5. The uppermost plot contains the results for the translation parameter h and the plot below the corresponding results for the Doppler parameter α. Each plot shows the standard deviation of the parameter for different levels of noise. The theoretical and empirical standard deviation of the translation agree for values lower than $\sigma_n \approx 0.8$, after which they start to differ. Concerning the Doppler factor the theoretical and empirical values agree well until $\sigma_n \approx 1.1$, after which the estimation is poor.

Thus, by studying the plots we can conclude that the system can handle noise with a standard deviation up to $\sigma_n \approx 0.8$. The amplitude for the original signal varies between 1 and 3.5. Using the standard deviation of the original signal, σ_W, the signal-to-noise ratio is

$$\text{SNR} = \frac{\sigma_n^2}{\sigma_W^2} \approx 0.21.$$

4.2 Real Data - Validation of Method

For the real data experiments we used 8 T-Bone MM-1 microphones, connected to an audio interface (M-Audio Fast Track Ultra 8R), which was connected to a computer. The sound was recorded at $f = 96$ kHz and we made the experiments in an anaechoic chamber. The eight microphones were placed approximately 0.3-1.5 meters away from each other and so that they spanned 3D. We generated sounds by playing a song on a mobile phone connected to a simple loudspeaker, while moving around in the room.

Using the sound generated from all eight microphones we estimated the sound source path, i.e. a 3D trajectory $s(t)$ and the eight 3D positions of the microphones $r_1, \ldots, r_8$. This was done using the technique described in (Zhayida et al., 2014) and refined in (Zhayida et al., 2016). The method builds on RANSAC algorithms based on minimal solvers (Kuang and Åström, 2013b) to find initial estimates of the sound trajectory $s(t)$ and microphone positions $r_1, \ldots, r_8$. These estimates are then refined using nonlinear optimization of a robust error norm, but also including a smooth motion prior.

For the validation of the method presented in this article we used data from two of the microphones. The played song lasted for approximately 29 s and the loudspeaker moved during the whole time. Furthermore, the motion was not constant – both the speed and direction were changed.

Since our method assume a constant parameter z in a window the recording was divided into patches for which the parameters were approximately constant. We divided the signal into 2834 patches of 1000 samples (i.e. approximately 0.01 s) each. Each of these patches was then investigated separately. For each patch i a constant loudspeaker position was given from ground truth. Thus, we had one sender position $s^{(i)}$, its derivative $\frac{\partial s^{(i)}}{\partial t}(i)$ and two receiver positions r_1 and r_2 to compare our results with.

4.2.1 Estimating the Parameters

As mentioned in Section 3.2 it is in practice not harder to estimate three model parameters. Therefore, to get a more precise solution, we have used (9) as model for the received signals. In this case $V^{(i)}(t)$ will be signal patch i from the first microphone and $\bar{V}^{(i)}(t)$ the corresponding signal patch from the second microphone.

Our method is developed to estimate small translations, s.t. $h \in [-10, 10]$ samples and the delays in the experiments were larger than that. Therefore we pre-estimated the translation using GCC-PHAT. For a description of the method, see (Knapp and Carter, 1976). The pre-estimation gave an integer delay $\tilde{h}^{(i)}$, whereas our method did a subsample refinement, estimated the Doppler parameter and the amplitude fac-

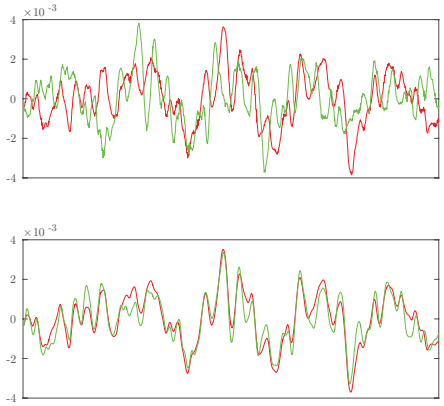

Figure 6: One example of the received signal patches at a certain time – the first signal in red and the second in green. The top image shows the signals before any modification. The bottom image shows the same patches after modifications using the optimal parameters.

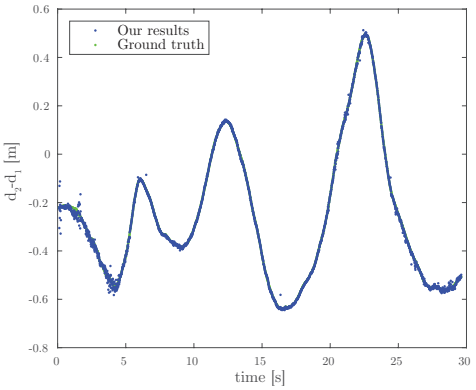

Figure 7: The figure shows the difference between the distances from receiver 1 to the sender (d_1) and receiver 2 to the sender (d_2) over time. Each dot represents the value for one signal patch. The ground truth is plotted in green and our estimations in blue. It is hard to distinguish any green dots since the estimations agree well with ground truth.

tor. This was done by minimizing the intergral

$$\int_t (V^{(i)}(t) - \gamma^{(i)}\bar{V}^{(i)}(\alpha^{(i)}t + \tilde{h}^{(i)} + h^{(i)}))^2 \, \mathrm{d}t.$$

Note that $\tilde{h}^{(i)}$ should be viewed as a constant in the equation above, while we were optimizing over $h^{(i)}$, $\alpha^{(i)}$ and $\gamma^{(i)}$.

This optimization was performed for all different patches. The result from one of these can be seen in Fig 6.

4.2.2 Comparing to Ground Truth

Using r_1, r_2 and $s^{(i)}$ we calculated the distances $d_1^{(i)}$ and $d_2^{(i)}$ from the loudspeaker to the microphones,

$$d_1^{(i)} = |r_1 - s^{(i)}|, \qquad d_2^{(i)} = |r_2 - s^{(i)}|.$$

The difference between these two distances, $\Delta d^{(i)} = d_2^{(i)} - d_1^{(i)}$, is connected to the time difference of arrival, and thus our computed translation $h^{(i)}$. While $h^{(i)}$ in measured in samples, $\Delta d^{(i)}$ is a measure in meters. Thus, we multiplied $h^{(i)}$ with a scaling factor c/f, where $c = 340$ m/s is the speed of sound and $f = 96$ kHz is the recording frequency. By that we had an estimation of $\Delta d^{(i)}$,

$$\Delta d^{(i)} = \frac{c}{f} \cdot h^{(i)}.$$

In a similar manner the time in samples can be expressed in seconds using f. Thereafter we could compare our estimation to ground truth. In Fig 7 ground truth $\Delta d^{(i)}$ is plotted together with our estimations for all patches, i.e. over time. It is clear that these agree well.

Concerning the Doppler parameter this is a measure of the change of distance differences, i.e.

$$\frac{\partial \Delta d}{\partial t} = \frac{\partial d_2}{\partial t} - \frac{\partial d_1}{\partial t}.$$

Here d_1 and d_2 denotes the distances over time, i.e. for all patches. If we look at one of the derivatives we have that $d_1(t) = |r_1 - s(t)|$ and

$$\frac{\partial d_1}{\partial t} = \frac{r_1 - s}{|r_1 - s|} \cdot \frac{\partial s}{\partial t},$$

where $\cdot$ denotes the scalar product between the two time dependent vectors. The derivative of d_2 is found similarly. Let $n_1^{(i)}$ and $n_2^{(i)}$ be unit vectors in the direction from $s^{(i)}$ to r_1 and r_2 respectively, i.e.

$$n_1^{(i)} = \frac{r_1 - s^{(i)}}{|r_1 - s^{(i)}|}, \qquad n_2^{(i)} = \frac{r_2 - s^{(i)}}{|r_2 - s^{(i)}|}.$$

Then, for a certain time step, the derivatives will be

$$\frac{\partial d_1^{(i)}}{\partial t} = n_1^{(i)} \cdot \frac{\partial s^{(i)}}{\partial t}, \qquad \frac{\partial d_2^{(i)}}{\partial t} = n_2^{(i)} \cdot \frac{\partial s^{(i)}}{\partial t}$$

and thus

$$\frac{\partial \Delta d^{(i)}}{\partial t} = n_2^{(i)} \cdot \frac{\partial s^{(i)}}{\partial t} - n_1^{(i)} \cdot \frac{\partial s^{(i)}}{\partial t}.$$

These ground truth Doppler values show how much Δd changes each second. However, our estimated Doppler factor α is a unitless constant. The relation between the two values is

$$\frac{\partial \Delta d}{\partial t} = (\alpha - 1) \cdot c,$$

with c still denoting the speed of sound. In Fig 8 we have plotted the ground truth value in green together

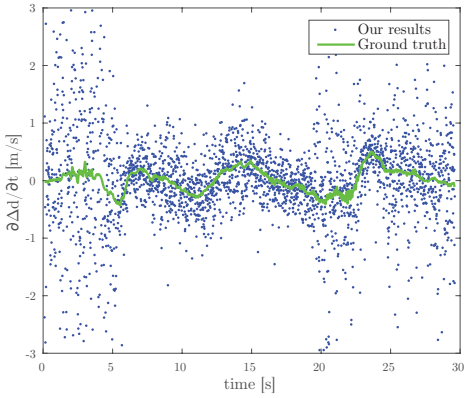

Figure 8: The derivative of the distance differences Δd plotted over time. The blue dots are our estimations and the solid green line is computed from ground truth. We see that even though the estimations are noisy they agree with ground truth.

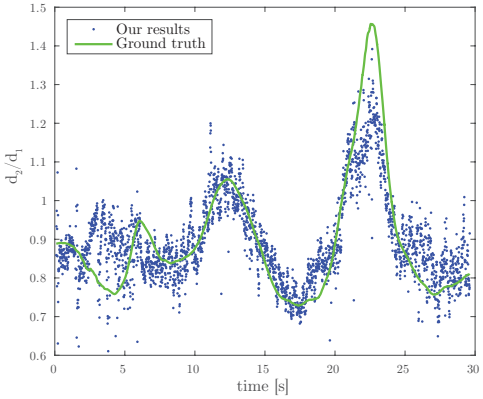

Figure 9: The distance quotient d_2/d_1 plotted over time. The green solid line represents the ground truth and each blue dot is the estimation for a certain patch. While the estimations are somewhat noisy there is no doubt that the pattern is the same.

with our estimations in blue. Even though the estimations are noisy the pattern and the similarities are well distinguishable.

There is also a relation between the amplitude factor for the two received signals and the distances $d_1^{(i)}$ and $d_2^{(i)}$. Our estimated amplitude $\gamma^{(i)}$ should be related to the distance quotient $d_2^{(i)}/d_1^{(i)}$. Furthermore, the sound from the loudspeaker spreads as on the surface of a sphere. Thus, the distance quotient is proportional to the square root of the amplitude $\gamma^{(i)}$,

$$\frac{d_2^{(i)}}{d_1^{(i)}} = C \cdot \sqrt{\gamma^{(i)}}.$$

We estimated the proportionality constant using data and got $C = 1.3$.

The distance quotient is plotted over time in Fig 9 – our estimations as blue dots and ground truth as a green line. Again we see that they clearly follow the same pattern.

The results from above show that our estimated parameters do contain relevant information. Though, the estimations in Fig 7, 8 and 9 are all quite noisy. This can be reduced by computing a moving average of the estimations. Fig 10 is similar to Fig 8 but instead of plotting the estimated derivative for each patch we have plotted the 20-patch moving average of the distance difference derivative. This means we have averaged over approximately 0.2 s. A comparison between the two figures shows that the moving average substantially reduces the noise in the estimates. The same method can be applied to the estimations of the translation h and the amplitude γ to reduce noise.

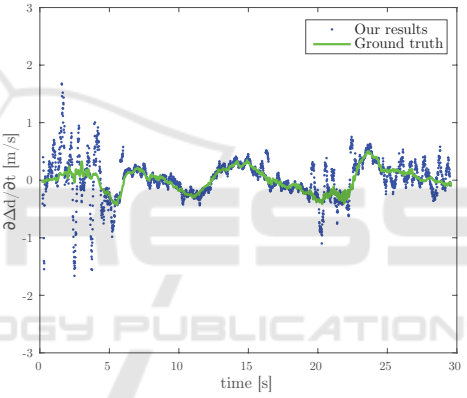

Figure 10: This plot shows essentially the same thing as Fig 8, i.e. $\partial\Delta d/\partial t$, but with a 20-patches moving average over the estimations. The averaging substantially reduces the noise.

Still after averaging, the estimations are noisy and poor in the beginning of the signal. This can be explained by the song that was played. The song starts with occasional drum beats with silence in between. Since the sound is not persistent the information is not sufficient to make good estimates. When the song is more continuous, after 5-6 s, this shows as the estimates get better.

5 CONCLUSIONS

In this paper we study the estimation of time-differences, amplitude changes and minute Doppler effects from two signals. We also study how to estimate the uncertainty in these estimated parameters.

The results are useful both for simultaneous determination of sender and receiver positions, but also for localization, beam-forming and diarization. In the paper we use previous results on stochastic analysis of interpolation and smoothing in order to give explicit formulas for the covariance matrix of the estimated parameters. We show that the approximations used are valid as long as the smoothing is at least 0.55 sample points and as long as the signal-to-noise ratio is smaller than 0.21. Furthermore we show in experimental studies on real data that these estimates provide useful information for subsequent analysis.

ACKNOWLEDGEMENTS

This work is supported by the strategic research projects ELLIIT and eSSENCE, Swedish Foundation for Strategic Research project "Semantic Mapping and Visual Navigation for Smart Robots" (grant no. RIT15-0038) and Wallenberg Autonomous Systems and Software Program (WASP).

REFERENCES

Anguera, X., Bozonnet, S., Evans, N., Fredouille, C., Friedland, G., and Vinyals, O. (2012). Speaker diarization: A review of recent research. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(2):356–370.

Anguera, X., Wooters, C., and Hernando, J. (2007). Acoustic beamforming for speaker diarization of meetings. *IEEE Transactions on Audio, Speech, and Language Processing*, 15(7):2011–2022.

Åström, K. and Heyden, A. (1999). Stochastic analysis of scale-space smoothing. *Advances in Applied Probability*, 30(1).

Batstone, K., Oskarsson, M., and Åström, K. (2016). Robust time-of-arrival self calibration and indoor localization using wi-fi round-trip time measurements. In *proc. of International Conference on Communication*.

Brandstein, M., Adcock, J., and Silverman, H. (1997). A closed-form location estimator for use with room environment microphone arrays. *Speech and Audio Processing, IEEE Transactions on*, 5(1):45 –50.

Cirillo, A., Parisi, R., and Uncini, A. (2008). Sound mapping in reverberant rooms by a robust direct method. In *Acoustics, Speech and Signal Processing, IEEE International Conference on*, pages 285 –288.

Cobos, M., Marti, A., and Lopez, J. (2011). A modified SRP-PHAT functional for robust real-time sound source localization with scalable spatial sampling. *Signal Processing Letters, IEEE*, 18(1):71 –74.

Crocco, M., Del Bue, A., Bustreo, M., and Murino, V. (2012). A closed form solution to the microphone position self-calibration problem. In *ICASSP*.

Do, H., Silverman, H., and Yu, Y. (2007). A real-time SRP-PHAT source location implementation using stochastic region contraction(SRC) on a large-aperture microphone array. In *ICASSP 2007*, volume 1, pages 121–124.

Knapp, C. and Carter, G. (1976). The generalized correlation method for estimation of time delay. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 24(4):320–327.

Kuang, Y. and Åström, K. (2013a). Stratified sensor network self-calibration from tdoa measurements. In *EUSIPCO*.

Kuang, Y. and Åström, K. (2013b). Stratified sensor network self-calibration from tdoa measurements. In *21st European Signal Processing Conference 2013*.

Kuang, Y., Burgess, S., Torstensson, A., and Åström, K. (2013). A complete characterization and solution to the microphone position self-calibration problem. In *ICASSP*.

Lindeberg, T. (1994). Scale-space theory: A basic tool for analyzing structures at different scales. *Journal of applied statistics*, 21(1-2):225–270.

Lindgren, G., Rootzén, H., and Sandsten, M. (2013). *Stationary Stochastic Processes for Scientists and Engineers*. CRC Press.

Plinge, A., Jacob, F., Haeb-Umbach, R., and Fink, G. A. (2016). Acoustic microphone geometry calibration: An overview and experimental evaluation of state-of-the-art algorithms. *IEEE Signal Processing Magazine*, 33(4):14–29.

Pollefeys, M. and Nister, D. (2008). Direct computation of sound and microphone locations from time-difference-of-arrival data. In *Proc. of ICASSP*.

Shannon, C. E. (1949). Communication in the presence of noise. *Proceedings of the IRE*, 37(1):10–21.

Zhayida, S., Andersson, F., Kuang, Y., and Åström, K. (2014). An automatic system for microphone self-localization using ambient sound. In *22st European Signal Processing Conference*.

Zhayida, S. and Åström, K. (2016). Time difference estimation with sub-sample interpolation. *Journal of Signal Processing*, 20(6):275–282.

Zhayida, S., Segerblom Rex, S., Kuang, Y., Andersson, F., and Åström, K. (2016). An Automatic System for Acoustic Microphone Geometry Calibration based on Minimal Solvers. *ArXiv e-prints*.