SUB-SAMPLE TIME DELAY ESTIMATION VIA AUXILIARY-FUNCTION-BASED ITERATIVE UPDATES

Kouei Yamaoka, Robin Scheibler, Nobutaka Ono, Yukoh Wakabayashi

Graduate School of System Design, Tokyo Metropolitan University
6-6, Asahigaoka, Hino-shi, Tokyo 191-0065, Japan
{yamaoka-kouei@ed., robin@, onono@, wakayuko@}tmu.ac.jp

ABSTRACT

We propose an efficient iterative method to estimate a sub-sample time delay between two signals. We formulate it as the optimization problem of maximizing the generalized cross correlation (GCC) of the two signals in terms of a continuous time delay parameter. The maximization is carried out with an auxiliary function method. First, we prove that when written as a sum of cosines, the GCC can be lower bounded at any point by a quadratic function. By repeatedly maximizing this lower-bound, an alternative update algorithm for the estimation of the time delay is derived. We follow through with numerical experiments highlighting that given a reasonable initial estimate, the proposed method converges quickly to the maximum of the GCC. In addition, we show that the method is robust to noise and attains the Cramér–Rao lower bound (CRLB).

Index Terms— Time delay estimation, time difference of arrival, generalized cross correlation, auxiliary function, majorization-minimization

1. INTRODUCTION

Audio array signal processing uses multiple microphones and exploits spatial cues for improved processing. Popular applications of microphone arrays include speech enhancement, source localization and separation [1]. Spatial cues come in two flavors: variations in amplitudes and time delays at different microphones. The time delay, or equivalently, phase difference between channels, is particularly important and underpins several important techniques in localization [2, 3], resampling [4], and synchronization [5]. For these, any improvement in time delay estimation directly translates to better performance.

A naive estimate of the time delay between two sampled signals is given by the location of the maximum of their discrete cross correlation (CC) [6]. Without further processing, the accuracy of this method is limited by the sampling frequency. For compact arrays, this can be a serious problem. For example, the maximum time difference of arrival (TDOA) for two microphones spaced by $4\,\mathrm{cm}$ is less than $0.12\,\mathrm{ms}$, i.e., less than 2 samples at $16\,\mathrm{kHz}$. Clearly, sample level accuracy is insufficient for many important applications, not just in audio processing, but also, for example, sonar [7], radar [8], or reflection seismology [9].

A popular method for time delay estimation is based on the generalized cross correlation (GCC) [10, 11], where the estimate is given by the location of its maximum. Applying interpolation in the vicinity of the maximum is effective to attain sub-sample estimates. Various schemes have been proposed, for example, parabolic [12] and Gaussian curve fitting [13], among others [14–16]. Yet another interpolation method is zero padding in the frequency domain, which corresponds to Dirichlet kernel interpolation [17] of the GCC. The ratio of non-padded to padded signal lengths is the attainable sub-sample accuracy. Finally, it is possible to try to find the maximum of the continuous GCC directly. For band-limited signals, following the Nyquist-Shannon sampling theorem [18, 19], the continuous GCC is obtained by sinc-interpolation of its discrete counterpart. Its maximization is a non-convex problem with no known closed-form solution. Nevertheless, a locally optimal solution can be found with a search algorithm such as the golden-section search (GSS) [20]. GSS-based estimation searches the maximum (or minimum) of a unimodal function by narrowing the interval where the maximum value is known to exist.

In this paper, we propose a new method for sub-sample time delay estimation based on maximizing the continuous GCC via the iterative maximization of an auxiliary function, which is also known as majorization-minimization (MM) method [21]. We first show that the objective function can be globally bounded by a quadratic auxiliary function. This auxiliary function can then be repeatedly maximized for guaranteed convergence to a local maximum. We evaluate the method through numerical experiments and find that when initialized in the vicinity of the true maximum, it converges there in only a few iterations. We confirm that the proposed method achieves the Cramér–Rao lower bound (CRLB) in the presence of additive Gaussian noise and evaluate its performance on reverberant signals. Compared to GSS-based estimation, we find that the proposed method typically converges in fewer iterations.

This work was supported by JSPS under grants 16H01735, 17F17049, and 19J20420.

The rest of this paper is organized as follows. Section 2 formulates the problem and sets notation. The algorithm is described in Section 3. Numerical experiments and their results are the topic of Section 4. Section 5 concludes.

2. PROBLEM FORMULATION

Let x_n and y_n be real discrete signals. The GCC of x_n and y_n is written as

$$\Phi^{(xy)}(t) = \frac{1}{N} \sum_{k=-N/2+1}^{N/2} W_k S_k^{(xy)} e^{j2\pi kt/N}, \quad (1)$$

where k is a discrete frequency index and $S_k^{(xy)}$ is the cross spectrum of x_n and y_n. The variable $W_k \in \mathbb{R}$, $W_k \geq 0$, is an arbitrary weight function. Finally, $t \in \mathbb{R}$ denotes the continuous time delay between x_n and y_n.

The GCC [10, 11] is a well-known method for estimating the discrete time delay t that maximizes (1), and suitable weight functions W_k have been proposed, e.g., GCC-phase transform (PHAT) and GCC-smoothed coherence transform (SCOT);

$$W_k^{\mathrm{PHAT}} = |S_k^{(xy)}|^{-1}, \quad W_k^{\mathrm{SCOT}} = \left(S_k^{(xx)} S_k^{(yy)}\right)^{-\frac{1}{2}}. \quad (2)$$

The GCC with $W_k = 1$ is equivalent to the ordinary CC. In typical implementations, the above GCC is only computed at discrete time delays given by the sampling frequency F_s of the input signals, i.e. $t \in \left\{ \frac{k}{F_s} \mid k = -\frac{N}{2}+1, \ldots, \frac{N}{2} \right\}$. To improve the accuracy, it is necessary to lift this restriction.

We consider finding a continuous variable $t \in \mathbb{R}$ maximizing (1), which we can rewrite as a sum of cosines using the conjugate symmetry of $S_k^{(xy)}$,

$$\Phi^{(xy)}(t) = \sum_{k=0}^{N/2} A_k \cos(\omega_k t + \phi_k), \quad (3)$$

where $A_k = \frac{\beta_k}{N} |W_k S_k^{(xy)}|$, $\phi_k = \angle S_k^{(xy)}$, $\omega_k = 2\pi \frac{k}{N}$, and $\beta_0 = \beta_{N/2} = 1$, and $\beta_k = 2$ for $k \notin \{0, N/2\}$. Our goal is thus to compute the sub-sample accuracy time delay estimate

$$\hat{t} = \operatorname*{arg\,max}_{t \in \mathbb{R}} \Phi^{(xy)}(t). \quad (4)$$

Unfortunately, there is no closed-form solution to this optimization problem. Since $\Phi^{(xy)}(t)$ denoted by the summation of the sinusoids, is a strictly unimodal function in short interval, the GSS [20] can find the peak. In practice, $\Phi^{(xy)}(t)$ is an unimodal function in the vicinity of its maximum, and thus, the GSS [20] can be applied. A drawback of this method is that it requires an initial interval guaranteed to contain the maximum. Concretely, the left and right side of the interval must be chosen to the increasing and decreasing parts of the unimodal range around the optimum, respectively.

In this paper, we solve the optimization problem of maximizing the objective function (3) using an auxiliary function that has a closed-form solution. It is then only required that the initial estimate falls in the basin of attraction of the true maximum, which is usually the case. Better initial estimation usually leads to faster convergence.

3. SUB-SAMPLE TIME DELAY ESTIMATION USING AUXILIARY FUNCTION

The auxiliary function method (also known as MM algorithm [21]) is a well-known due to the *expectation-maximization* and various other algorithms [22, 23]. Adapted to our problem, we would like to find an auxiliary function $Q(t, \boldsymbol{\theta})$ such that

- $\Phi^{(xy)}(t) \geq Q(t, \boldsymbol{\theta})$ for any t and $\boldsymbol{\theta}$,
- For any t_0, $\exists \boldsymbol{\theta}_0 = f(t_0)$ such that $\Phi^{(xy)}(t_0) = Q(t_0, \boldsymbol{\theta}_0)$,

where $\boldsymbol{\theta} = (\theta_0, \theta_1, \cdots, \theta_{N/2})$ are auxiliary variables. Provided such a $Q(t, \boldsymbol{\theta})$ exists, and given an initial estimate $\hat{t}_0$, the following sequence of updates is guaranteed to converge to a local maximum

$$\boldsymbol{\theta}^{(\ell)} = f(\hat{t}^{(\ell)}), \quad \hat{t}^{(\ell+1)} = \operatorname*{arg\,max}_{t \in \mathbb{R}} Q(t, \boldsymbol{\theta}^{(\ell)}), \quad (5)$$

where ℓ is the iteration index.

3.1. Quadratic Auxiliary Function for Continuous GCC

This section provides an auxiliary function for $\Phi^{(xy)}(t)$.

Theorem 1 *The following is an auxiliary function for* $\Phi^{(xy)}(t)$,

$$Q(t, \boldsymbol{\theta}) = \sum_{k=0}^{N/2} -\frac{A_k}{2} \cdot \frac{\sin \theta_k}{\theta_k} (\omega_k t + \phi_k + 2n_k \pi)^2 + C, \quad (6)$$

where C is a constant term that does not include t, and $n_k \in \mathbb{Z}$ is such that $|\omega_k t + \phi_k + 2n_k\pi| \leq \pi$. The auxiliary variables are θ_k and n_k and $Q(t, \boldsymbol{\theta}) = \Phi^{(xy)}(t)$ when

$$\theta_k = \omega_k t + \phi_k + 2n_k \pi. \quad (7)$$

This theorem is a direct consequence of the following inequality for a cosine function, which is of general interest.

Proposition 1 *Let $|\theta_0| \leq \pi$. For any real number θ, the following inequality is satisfied*

$$\cos\theta \geq -\frac{1}{2} \frac{\sin\theta_0}{\theta_0} \theta^2 + \left(\cos\theta_0 + \frac{1}{2} \theta_0 \sin\theta_0 \right). \quad (8)$$

When $|\theta_0| < \pi$, equality holds if and only if $|\theta| = |\theta_0|$. When $|\theta_0| = \pi$, equality holds if and only if $\theta = (2n+1)\pi$, $n \in \mathbb{Z}$.

Proof: Let

$$f(\theta) = \cos\theta + \frac{1}{2}\frac{\sin\theta_0}{\theta_0}\theta^2 - \left(\cos\theta_0 + \frac{1}{2}\theta_0\sin\theta_0\right). \quad (9)$$

Then, we have

$$f'(\theta) = -\sin\theta + \frac{\sin\theta_0}{\theta_0}\theta = -\theta\left(\frac{\sin\theta}{\theta} - \frac{\sin\theta_0}{\theta_0}\right). \quad (10)$$

We separately consider the following three cases.

Case 1: $0 < |\theta_0| < \pi$

Because $\sin\theta/\theta$ is monotonically decreasing in $0 \le \theta \le \pi$,

$$f'(\theta)\begin{cases} < 0 & (0 \le \theta < |\theta_0|), \\ = 0 & (\theta = |\theta_0|), \\ > 0 & (|\theta_0| < \theta \le \pi). \end{cases} \quad (11)$$

It thus appears that $f(\theta)$ attains its minimum at $|\theta_0|$. Moreover, $f(\theta_0) = 0$ and thus $f(\theta) \ge 0$ in $0 \le \theta \le \pi$. Since $f(\theta)$ is an even function, its minimum value in $-\pi \le \theta \le \pi$ is also 0. Because $\cos\theta$ is periodic but $-\theta^2$ is not, $f(\theta + 2n\pi) > f(\theta)$ for any $-\pi \le \theta \le \pi$ and integer $n \ne 0$. Therefore, $f(\theta) \ge 0$, with equality if and only if $|\theta| = |\theta_0|$.

Case 2: $\theta_0 = 0$

In this case, for $0 \le \theta \le \pi$, we have

$$f'(\theta)\begin{cases} = 0 & (\theta = |\theta_0| = 0) \\ > 0 & (|\theta_0| < \theta \le \pi) \end{cases}, \quad (12)$$

which means f takes its minimum value at $f(0) = 0$ in $-\pi \le \theta \le \pi$. Similarly to case 1, we obtain $f(\theta) \ge 0$, with equality if and only if $\theta = 0$.

Case 3: $\theta_0 = \pi$ or $\theta_0 = -\pi$

In this case, $f(\theta) = \cos\theta + 1$. Therefore $f(\theta) \ge 0$, with equality if and only if $\theta = (2n+1)\pi$ for any $n \in \mathbb{Z}$. ■

Then, Theorem 1 is proved as follows.

Proof of Theorem 1: Because $\cos(\omega_k t + \phi_k) = \cos(\omega_k t + \phi_k + 2\pi n_k)$ with $n_k \in \mathbb{Z}$, and because $A_k \ge 0$, we can apply Proposition 1 separately to each term of the sum in (3). ■

3.2. Derivation of Auxiliary Function and Update Rules

Since $Q(t, \boldsymbol{\theta})$ is a quadratic function, it is easily maximized with respect to t by putting its derivative to zero

$$\frac{\partial Q}{\partial t} = -\sum_{k=0}^{N/2} A_k\omega_k \frac{\sin\theta_k}{\theta_k}(\omega_k t + \phi_k + 2n_k\pi) = 0. \quad (13)$$

Therefore, the maximizer is

$$\hat{t} = \frac{\sum_{k=0}^{N/2} A_k\omega_k^2(\sin\theta_k/\theta_k)(-(\phi_k + 2n_k\pi)/\omega_k)}{\sum_{k=0}^{N/2} A_k\omega_k^2(\sin\theta_k/\theta_k)}. \quad (14)$$

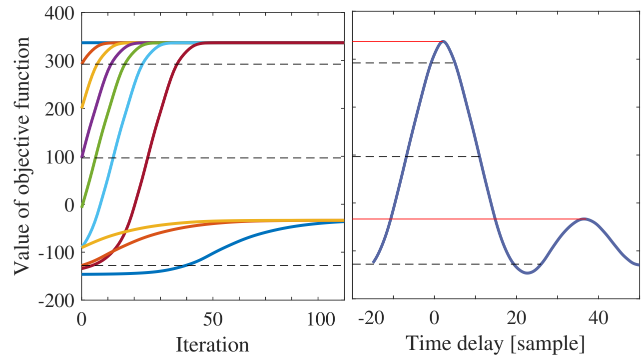

Figure 1: Left: the value of objective function via the proposed method over initial values. Different curves represent different values of the initial time delay estimate, taken every 3 samples from 2 to 29. Right: objective function.

Now under the condition for equality (7), we can substitute $\phi_k + 2n_k\pi = \theta_k - \omega_k t$ and obtain the final update rules

$$n_k^{(\ell)} \leftarrow \underset{n\in\mathbb{Z}}{\arg\min}\left|\omega_k t^{(\ell)} + \phi_k + 2n\pi\right|, \quad (15)$$

$$\theta_k^{(\ell)} \leftarrow \omega_k t^{(\ell)} + \phi_k + 2n_k^{(\ell)}\pi, \quad k = 0, \dots, \frac{N}{2}, \quad (16)$$

$$t^{(\ell+1)} \leftarrow t^{(\ell)} - \frac{\sum_{k=0}^{N/2} A_k\omega_k^2\left(\sin\theta_k^{(\ell)}/\theta_k^{(\ell)}\right)\frac{\theta_k^{(\ell)}}{\omega_k}}{\sum_{k=0}^{N/2} A_k\omega_k^2\left(\sin\theta_k^{(\ell)}/\theta_k^{(\ell)}\right)}. \quad (17)$$

Interestingly, the second term of (17) is a weighted sum of the auxiliary variables scaled by the frequency, i.e., $\theta_k^{(\ell)}/\omega_k$.

4. NUMERICAL EXPERIMENTS

To evaluate the effectiveness of the proposed method, we investigated the performance of the sub-sample time delay estimation in terms of the convergence in subsection 4.1 and 4.2 and robustness to noise in subsection 4.3.

4.1. Empirical Convergence

In this experiment, we created stereo observations with simulated time delays ranging from -5 to 5 samples and added white Gaussian noise to each microphone with signal-to-noise ratios (SNRs) from -10 dB to 30 dB. We used four different target signals, Japanese male/female and English male/female speech, sampled at 16 kHz. We evaluated the performance of the proposed method compared to that of conventional GSS-based estimation [20]. The proposed method was initialized with the sample maximizing the discrete GCC. The initial interval for GSS was the two samples around that maximum. The weight W_k of the GCC was set to 1 for all frequencies.

Figure 1 shows the convergence of the objective function with the proposed method for different initial values with the

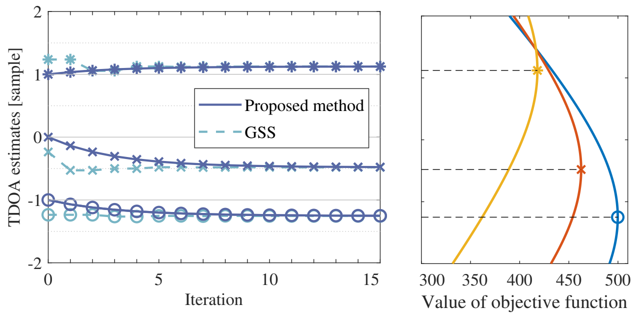

Figure 2: Left: TDOA estimates in the reverberant environment with the SNR of 10 dB; $\circ$, DOA is 30°; $\times$, DOA is 70°; $*$, DOA is 315°. Right: value of the objective function corresponding to respective DOAs.

ground truth at 2.0996 sample. The SNR is 10 dB. We set the initial value of the proposed method to every 3 samples form 2 to 29. This shows that if the initial value is close enough to the ground truth, between -20 to 22 samples in this figure, the proposed method will converge to the global maximum. Here, the error was 0.035 sample. Moreover, we empirically confirm the guaranteed monotonic increase of the objective function of the proposed method for any initial value. In addition, the better the initial estimate, the faster the convergence is. In contrast, GSS is initialized with an interval that must be unimodal and contain the maximum. Given a better initial estimate, it is not clear how to produce a narrower such interval to hasten convergence, especially in noisy conditions.

4.2. Convergence in Reverberant Environment

Next, we convolved a signal with room impulse responses simulated by the *pyroomacoustics* Python package [24]. We used the same experimental conditions as in subsection 4.1 except for the time delay and reverberation time. We estimated the TDOA for three directions of arrival (DOAs) at 30°, 70°, and 135°, with a reverberation time of 300 ms. The distance between the microphones was 4 cm and the source placed 1.5 m away. We averaged the resulting TDOA estimates over different speakers.

Figure 2 shows the TDOA estimates by the proposed method and GSS over iterations. Markers '$\circ$', '$\times$', and '$*$' denote the DOAs of 30°, 70°, and 135°, respectively. Although the TDOA estimation in reverberant environment is a difficult problem [25], both the proposed method and GSS converge. Considering these results, it can be concluded that our proposed method shows enough performance as an alternative method for sub-sample time delay estimation.

4.3. Robustness to Noise

In this experiment, we compare the performance of the GSS and proposed methods with respect to the initial value and the number of iterations in the presence of noise. For this purpose, we use the same speech signal as in subsection 4.1 and a copy randomly delayed by $t \sim U[0, 1]$. The two signals are corrupted by additive white Gaussian noise at SNR ranging from -12 dB to 40 dB. Given the input signal and the variance of the noise σ^2, the CRLB for $\hat{t}$ is

$$\mathrm{Var}\{\hat{t}\} \geq \left\{ 8\pi^2 \sum_{k=0}^{N/2} \frac{(|S_k|^2/(N\sigma^2))^2}{1+2(|S_k|^2/(N\sigma^2))} (kF_s/N)^2 \right\}^{-1},$$

where S_k is the k-th point of the discrete Fourier transform (DFT) of the input signal, and F_s is the sampling frequency. We compare two initialization schemes: from the location of the peak of the discrete GCC, and from the result of parabolic interpolation around it. The experiment is repeated for 100 realizations of the delay and the noise.

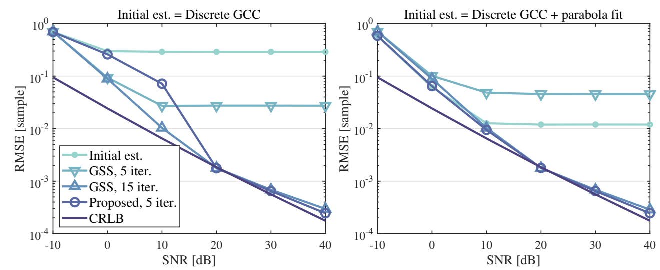

Figure 3: RMSE of TDOA estimates in noise.

Figure 3 shows that both GSS and the proposed method reach the CRLB as the SNR increases. However, the number of iterations needed vary. At low SNR, GSS performs well with 5 iterations, but at high SNR at least 15 iterations are needed to reach the CRLB. This result does not depend on the initialization. In contrast, 5 iterations are always sufficient for the proposed method, provided a good enough starting point such as given by parabolic interpolation.

5. CONCLUSIONS

This paper presents efficient iterative update rules that maximize the continuous GCC for sub-sample time delay estimation based on auxiliary function technique. The objective function, that is, the continuous GCC is represented as a sum of sinusoids and a quadratic auxiliary function can be derived. By updating the time delay estimates and auxiliary variable alternatively, the objective function is monotonically maximized. This optimization procedure should extend to multichannel signals by designing an appropriate objective function, whereas the conventional GSS can only be applied to pairs of signals.

In the experiments, we confirmed that the derived update rules achieve faster convergence compared to GSS and attains the CRLB. We thus conclude that the proposed method could be an alternative method for sub-sample time delay estimation.

6. REFERENCES

[1] M. Brandstein and D. Ward, *Microphone arrays: signal processing techniques and applications*, Springer, 2001.

[2] A. Beck, P. Stoica, and J. Li, "Exact and approximate solutions of source localization problems," *IEEE Trans. on Signal Process.*, vol. 56, no. 5, pp. 1770–1778, Apr. 2008.

[3] T.-K. Le and N. Ono, "Closed-form and near closed-form solutions for TDOA-based joint source and sensor localization," *IEEE Trans. on Signal Process.*, vol. 65, no. 5, pp. 1207–1221, Dec. 2016.

[4] S. Miyabe, N. Ono, and S. Makino, "Blind compensation of interchannel sampling frequency mismatch for ad hoc microphone array based on maximum likelihood estimation," *Signal Processing*, vol. 107, pp. 185–196, Feb. 2015.

[5] A. J. Coulson, "Maximum likelihood synchronization for OFDM using a pilot symbol: algorithms," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 12, pp. 2486–2494, Dec. 2001.

[6] J. Chen, Y. Huang, and J. Benesty, "Time delay estimation," in *Audio Signal Processing for Next-Generation Multimedia Communication Systems*, pp. 197–227, Kluwer Academic Publishers, Boston, 2004.

[7] G. Carter, "Time delay estimation for passive sonar signal processing," *IEEE Trans. on Acoust., Speech, Signal Process.*, vol. 29, no. 3, pp. 463–470, June 1981.

[8] P. Protiva, J. Mrkvica, and J. Machac, "Estimation of wall parameters from time-delay-only through-wall radar measurements," *IEEE Trans. on Antennas Propag.*, vol. 59, no. 11, pp. 4268–4278, Nov. 2011.

[9] J. Capon, "Applications of detection and estimation theory to large array seismology," *Proc. IEEE*, vol. 58, no. 5, pp. 760–770, May 1970.

[10] C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. on Acoust., Speech, Signal Process.*, vol. 24, no. 4, pp. 320–327, 1976.

[11] I. J. Tashev, *Sound Capture and Processing*, Practical Approaches. John Wiley & Sons, 2009.

[12] G. Jacovitti and G. Scarano, "Discrete time techniques for time delay estimation," *IEEE Trans. on Signal Process.*, vol. 41, no. 2, pp. 525–533, 1993.

[13] L. Zhang and X. Wu, "On the application of cross correlation function to subsample discrete time delay estimation," *Dig. Signal Process.*, vol. 16, pp. 682–694, 2006.

[14] F. Viola and W. F. Walker, "Computationally efficient spline-based time delay estimation," *IEEE Trans. on UFFC*, vol. 55, no. 9, pp. 2084–2091, 2008.

[15] B. Qin, H. Zhang, Q. Fu, and Y. Yan, "Subsample time delay estimation via improved GCC PHAT algorithm," *International Conference on Signal Process.*, pp. 2579–2582, 2008.

[16] R. Tao, X.-M. Li, Y.-L. Li, and Y. Wang, "Time-delay estimation of chirp signals in the fractional fourier domain," *IEEE Trans. on Signal Process.*, vol. 57, no. 7, pp. 2852–2855, 2009.

[17] M. Vetterli, J. Kovačević, and V. K. Goyal, *Foundations of signal processing*, Cambridge Univ. Press, 2014.

[18] C. E. Shannon, "Communication in the presence of noise," *Proc. IRE*, vol. 37, pp. 10–21, 1949.

[19] H. Ogawa, "Sampling theory and Isao Someya: A historical note," *Sampling theory in signal and image process.*, vol. 5, no. 3, pp. 247–256, 2006.

[20] J. C. Kiefer, "Sequential minimax serach for a maximum," *Proc. of the American Mathematical Society*, vol. 4, no. 3, pp. 502–506, 1953.

[21] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *The Am. Stat.*, vol. 58, no. 1, pp. 30–37, 2004.

[22] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Proc. NIPS*, 2000, pp. 556–562.

[23] H. Kameoka, T. Nishimoto, and S. Sagayama, "A multipitch analyzer based on harmonic temporal structured clustering," *IEEE Trans. on Audio, Speech, Lang. Process.*, vol. 15, no. 3, pp. 982–994, Mar. 2007.

[24] R. Scheibler, E. Bezzam, and I. Dokmanić, "Pyroomacoustics: A Python package for audio room simulations and array processing algorithms," in *Proc. IEEE ICASSP*, Calgary, CA, Apr. 2018, pp. 351–355.

[25] B. Champagne, S. Bedard, and A. Stephenne, "Performance of time-delay estimation in the presence of room reverberation," *IEEE Trans. on Speech Audio Process.*, vol. 4, no. 2, pp. 148–152, Mar. 1996.