Simultaneous Frequency Estimation for Three or More Sinusoids Based on Sinusoidal Constraint Differential Equation

Kenta Yamada, Yoshiki Masuyama, Yukoh Wakabayashi, Nobutaka Ono
*Tokyo Metropolitan University, Japan
E-mail: {yamada-kenta5@ed., masuyama-yoshiki@ed., wakayuko@, onono@}tmu.ac.jp

***Abstract*—In this paper, we present a short-time frequency estimation method that can handle multiple sinusoids simultaneously. Frequency estimation is a fundamental problem in audio analysis. For realizing high-temporal resolution, an approach based on a differential equation of a sinusoid, which is referred to as the sinusoidal constraint differential equation (SCDE), has been proposed. The SCDE-based method can efficiently and accurately estimate frequency even from a short-term signal. However, in terms of simultaneous estimation, up to two sinusoids have been considered so far. In this paper, we extend this approach to three or more sinusoids. Our experimental results show that our method outperformed existing methods based on the discrete Fourier transform.**

***Index Terms*—Frequency estimation, sinusoidal modeling, audio modeling, differential equation.**

I. Introduction

Frequency estimation of multiple sinusoids has been intensively studied as a fundamental problem in audio signal processing with various applications [1]–[7]. A popular method is the maximum likelihood estimation [8]–[12]. Although it can accurately estimate the frequencies, it requires an iterative optimization to solve a non-convex optimization problem.

More computationally efficient method is to leverage the discrete Fourier transform (DFT) [13]–[18]. To improve the estimation accuracy more than taking the spectral peak, various methods have been developed, including the parabolic interpolation [13], [14] and the spectral reassignment [15], [16]. However, their estimation accuracy decreases as the frequency resolution becomes low with a short-time signal. Fig. 1 shows a sum of three sinusoids of 100 ms whose frequencies are 440 Hz, 460 Hz, and 480 Hz. Its spectrum with different signal lengths are also depicted. The power spectrum of 10 ms cannot separate the peak corresponding to each sinusoid because they are in a single peak. In order to separate three peaks, a signal length of 40 ms is required in this case.

Our motivation is the precise real-time pitch analysis of polyphonic instrumental and vocal sounds. In music, their overtones overlap to form harmony. However, whether consciously or unconsciously, these overtones might mismatch slightly due to tuning, the skill of the singer or performer, musical intention, etc., and they can also fluctuate over time. Thus, in order to capture the temporal variation of such very close frequency components, a method with high-frequency resolution in a short time is required.

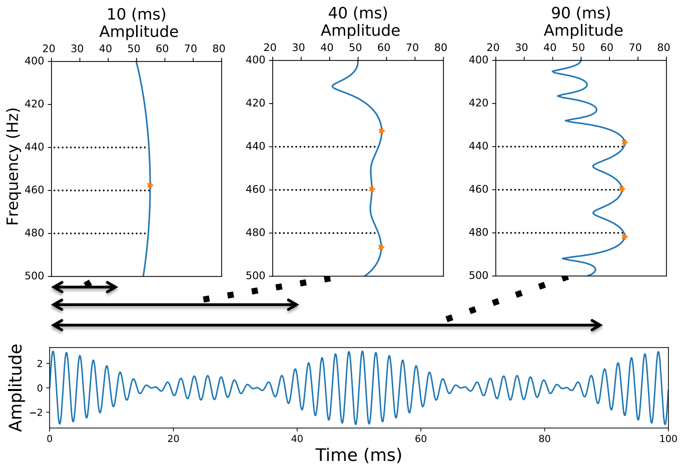

Fig. 1. The sum of three sinusoids of 100 ms whose frequencies are 440 Hz, 460 Hz, and 480 Hz (bottom). Its spectrum with different signal lengths are also depicted (top).

One promising method to estimate the frequency from a short-time signal is the sinusoidal constraint differential equation (SCDE)-based method [19]–[22]. When the signal comprises one sinusoid, its second-order derivative is given by the product of the squared frequency and the signal itself. The SCDE-based method exploits this relation and estimates the frequency efficiently. Although it handles only one sinusoid, it was extended to the case with multiple sinusoids by combining the DFT and applying SCDE to each sub-band. This method assumes spectral peaks corresponding to sinusoids are separable and thus require a sufficient-length signal, as shown in Fig. 1.

In this paper, we present an extension of the SCDE-based method that can handle multiple sinusoids simultaneously. While the SCDE with two sinusoids was already presented in [21], we explicitly extend it to the case of three or more sinusoids. Based on the SCDE with multiple sinusoids, our method estimates the frequencies by solving simultaneous equations and finding the roots of a polynomial equation. Since the method does not require extracting the sub-band that contains one sinusoid, it can work with a short-term signal. We experimentally investigate the relationship between the signal length and the estimation accuracy. The results show that the SCDE-based method outperformed the DFT-based methods when a signal is short.

II. Sinusoidal constraint differential equation

A. SCDE method for a single sinusoid

We first summarize the basis of SCDE method according to [19], [20]. Let us consider a sinusoid $x(t)$ given as

$$x(t) = A\cos(\omega t + \phi), \tag{1}$$

where t denotes continuous time, and A, ω, ϕ respectively are the amplitude, the frequency, and the phase of the sinusoid, respectively. We consider the following second-order differential equation:

$$\frac{d^2}{dt^2}x(t) + \omega^2 x(t) = 0. \tag{2}$$

This equation is called the SCDE [19], [20] and holds for any $t \in \mathbb{R}$. Since it does not contain A and ϕ, the frequency estimation can be performed independently of these parameters.

When a signal is a sum of sinusoids, the DFT is used to select the sub-band containing each sinusoid, and the SCDE is applied to each sub-band to estimate each frequency. This method assumes that the frequencies are well separated. Hence, it is not applicable to estimating the frequencies from a short-time signal where the spectral peaks are not separable.

B. SCDE method for two sinusoids

We summarize the basis of SCDE method for two sinusoids according to [21]. Let us consider a signal comprised of two sinusoids:

$$x(t) = A_1\cos(\omega_1 t + \phi_1) + A_2\cos(\omega_2 t + \phi_2), \tag{3}$$

where A_k, ω_k, and ϕ_k denote the amplitude, the frequency, and the phase of the kth sinusoid, respectively. Each sinusoid satisfies the following equation as in Eq. (2):

$$\left(\frac{d^2}{dt^2} + \omega_k^2\right) A_k\cos(\omega_k t + \phi_k) = 0, \tag{4}$$

for $k = 1,\ 2$. Hence, the following equation holds for any $t \in \mathbb{R}$ [21]:

$$\begin{aligned}&\left(\frac{d^2}{dt^2} + \omega_1^2\right)\left(\frac{d^2}{dt^2} + \omega_2^2\right) x(t) = 0\\ \Leftrightarrow &\frac{d^4}{dt^4}x(t) + \alpha\frac{d^2}{dt^2}x(t) + \beta x(t) = 0,\end{aligned} \tag{5}$$

where α and β are defined as

$$\begin{cases}\alpha = \omega_1^2 + \omega_2^2\\ \beta = \omega_1^2\omega_2^2\end{cases}. \tag{6}$$

We used the linearity and commutativity of the operator $(d^2/dt^2 + \omega_k^2)$. To obtain ω_1 and ω_2 that satisfy Eq. (5), we minimize the following objective function:

$$J(\alpha, \beta) = \int_\Gamma \left[\frac{d^4}{dt^4}x(t) + \alpha\frac{d^2}{dt^2}x(t) + \beta x(t)\right]^2 dt, \tag{7}$$

where Γ denotes definite integral.

III. Sinusoidal constraint differential equation for more than two sinusoids

A. SCDE method for three sinusoids

We extended the SCDE to the case with three or more sinusoids. Let us consider a signal consisting of three sinusoids. The SCDE with three sinusoids is defined by

$$\begin{aligned}&\left(\frac{d^2}{dt^2} + \omega_1^2\right)\left(\frac{d^2}{dt^2} + \omega_2^2\right)\left(\frac{d^2}{dt^2} + \omega_3^2\right) x(t) = 0\\ \Leftrightarrow\ &\frac{d^6}{dt^6}x(t) + \alpha\frac{d^4}{dt^4}x(t) + \beta\frac{d^2}{dt^2}x(t) + \gamma x(t) = 0,\end{aligned} \tag{8}$$

where α, β, and γ are defined as

$$\begin{cases}\alpha = \omega_1^2 + \omega_2^2 + \omega_3^2\\ \beta = \omega_1^2\omega_2^2 + \omega_2^2\omega_3^2 + \omega_1^2\omega_3^2\\ \gamma = \omega_1^2\omega_2^2\omega_3^2\end{cases}. \tag{9}$$

To obtain α, β, and γ that satisfy Eq. (8), we minimize the following objective function:

$$\begin{aligned}&J(\alpha, \beta, \gamma) =\\ &\int_\Gamma \left[\frac{d^6}{dt^6}x(t) + \alpha\frac{d^4}{dt^4}x(t) + \beta\frac{d^2}{dt^2}x(t) + \gamma x(t)\right]^2 dt.\end{aligned} \tag{10}$$

By setting the partial derivatives of $J(\alpha, \beta, \gamma)$ to zero, we can obtain the optimal α, β, and γ as follows

$$\begin{pmatrix}\alpha\\ \beta\\ \gamma\end{pmatrix} = -\begin{pmatrix}S_{2,2} & S_{1,2} & S_{0,2}\\ S_{2,1} & S_{1,1} & S_{0,1}\\ S_{2,0} & S_{1,0} & S_{0,0}\end{pmatrix}^{-1}\begin{pmatrix}S_{2,3}\\ S_{1,3}\\ S_{0,3}\end{pmatrix}. \tag{11}$$

Here, for simplicity, we introduce the following definition of the short-time covariance:

$$S_{m,n} = \int_\Gamma \left[\frac{d^{2m}x(t)}{dt^{2m}}\frac{d^{2n}x(t)}{dt^{2n}}\right] dt. \tag{12}$$

According to Eq. (9), ω_1^2, ω_2^2, and ω_3^2 are computed by solving the following cubic equation for Ω:

$$\Omega^3 - \alpha\Omega^2 + \beta\Omega - \gamma = 0. \tag{13}$$

This equation can also be solved in a closed-form.

B. General solution for SCDE method

For a sum of n sinusoids $x(t)$, the following SCDE holds:

$$\begin{aligned}&\left(\frac{d^2}{dt^2} + \omega_1^2\right)\cdots\left(\frac{d^2}{dt^2} + \omega_n^2\right) x(t) = 0\\ \Leftrightarrow\ &\frac{d^{2n}}{dt^{2n}}x(t) + \alpha_1\frac{d^{2(n-1)}}{dt^{2(n-1)}}x(t)\\ &+ \cdots + \alpha_{n-1}\frac{d^2}{dt^2}x(t) + \alpha_n x(t) = 0,\end{aligned} \tag{14}$$

where $\alpha_1, \ldots, \alpha_n$ denote coefficients of differential equation. To obtain $\alpha_1, \ldots, \alpha_n$ that satisfy Eq. (14), we minimize the following objective function:

$$J(\alpha_1, \ldots, \alpha_n) = \int_\Gamma \left[\frac{d^{2n}}{dt^{2n}}x(t) + \sum_{k=1}^{n}\alpha_k\frac{d^{2(n-k)}}{dt^{2(n-k)}}x(t)\right]^2 dt. \tag{15}$$

In order to optimize $\alpha_1, \ldots, \alpha_n$, we set the partial derivatives of $J(\alpha_1, \cdots, \alpha_n)$ to zero and solve the obtained simultaneous equations:

$$\begin{cases} S_{n,n-1} + \alpha_1 S_{n-1,n-1} & + \cdots + \alpha_n S_{0,n-1} & = 0 \\ S_{n,n-2} + \alpha_1 S_{n-1,n-2} & + \cdots + \alpha_n S_{0,n-2} & = 0 \\ \quad \vdots & & \\ S_{n,0} + \alpha_1 S_{n-1,0} & + \cdots + \alpha_n S_{0,0} & = 0 \end{cases}, \tag{16}$$

where Eq. (16) denotes the simultaneous equations, and we can obtain

$$\begin{pmatrix} \alpha_1 & \cdots & \alpha_n \end{pmatrix}^{\mathsf{T}} = -\boldsymbol{B}^{-1} \begin{pmatrix} S_{n,n-1} & S_{n,n-2} & \cdots & S_{n,0} \end{pmatrix}^{\mathsf{T}}, \tag{17}$$

where T denotes the transpose, and $\boldsymbol{B}$ is a symmetry covariance matrix such that the (i, j) entry is $S_{n-i,n-j}$. According to the relation between $\omega_1^2, \ldots, \omega_n^2$ and $\alpha_1, \ldots, \alpha_n$, $\omega_1^2, \ldots, \omega_n^2$ are computed by solving the following nth order equation for Ω:

$$\Omega^n + (-1)^1\alpha_1\Omega^{n-1} + (-1)^2\alpha_2\Omega^{n-2} + (-1)^3\alpha_3\Omega^{n-3} + \cdots + (-1)^{n-1}\alpha_n = 0. \tag{18}$$

Eq. (18) can be solved in a closed-form under fourth-order equations. The fifth-order or higher order equation can not be solved algebraically, but we can efficiently compute the solution by an iterative procedure.

C. Discussion on estimation of number of sinusoids

In this paper, we assume that the number of sinusoids is known in advance, but here we mention how it can be estimated. Suppose a set of $(n + 1)$ signals:

$$\left(\frac{d^{2n}}{dt^{2n}} x(t), \frac{d^{2(n-1)}}{dt^{2(n-1)}} x(t), \ldots, \frac{d^2}{dt^2} x(t), x(t) \right), \tag{19}$$

when $x(t)$ consists of only k sinusoids, which is less than n. As an example, suppose the case when $n = 3$ and $k = 2$. Adding to the fact that $\left(\frac{d^4}{dt^4}x(t), \frac{d^2}{dt^2}x(t), x(t)\right)$ satisfies Eq. (5), $\left(\frac{d^6}{dt^6}x(t), \frac{d^4}{dt^4}x(t), \frac{d^2}{dt^2}x(t)\right)$ also satisfies the same equation since it is easily obtained by applying the operator d^2/dt^2 to the both sides of Eq. (5). It indicates that two of four signals are linearly independent in this case. By generalizing this, only k of $(n+1)$ signals in Eq. (19) are linearly independent. Then, the rank of the covariance matrix of Eq. (19) becomes k, and we could estimate the number of sinusoids from the rank (or practically the number of large eigenvalues) of the covariance matrix of Eq. (19). We will investigate this in the future.

IV. Experiments

A. Comparisons with other methods on the signal length

In order to evaluate the relationship between the signal length and estimation accuracy, we compare the SCDE-based method with two DFT-based methods: the parabolic interpolation method [13], [14] and the spectral reassignment method [15], [16]. The parabolic interpolation method finds the spectral peak and using this main bin and its both side of neighbors, and fits a quadratic function to those bins. The spectral reassignment method calculates the derivative of the phase by using a time-differentiated window function. The DFT-based methods estimate multiple frequencies by applying the method to each spectral peak. For these methods, the number of DFT points was set to 65536 by zero padding. They were conducted only when the signal was long enough to find the spectral peaks corresponding to all sinusoids. The DFT-based methods were only applied when the signal was longer than 40 ms.

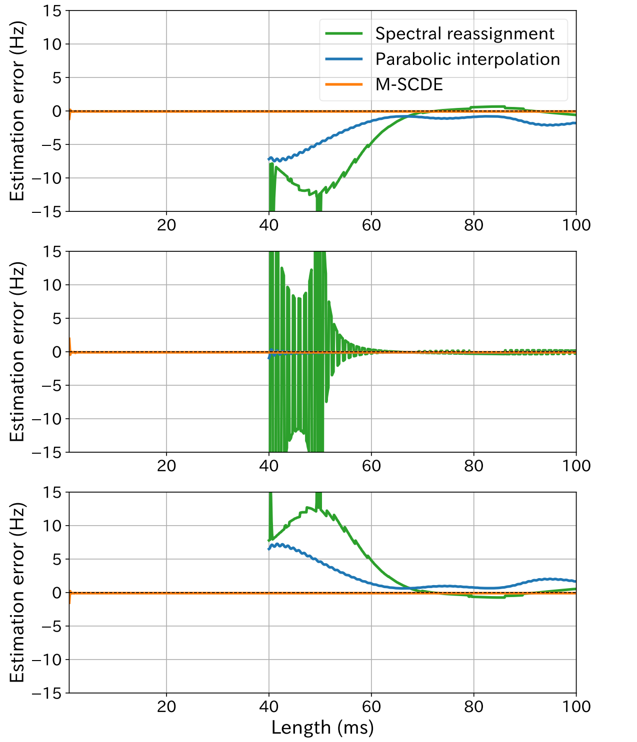

Fig. 2. Estimation error for the frequencies of 440 Hz (top), 460 Hz (middle), and 480 Hz (bottom) on the signal length by the three methods.

At first, we investigated the estimation accuracy with a sum of three sinusoids whose frequencies were 440 Hz, 460 Hz, and 480 Hz. The sampling frequency was 44.1 kHz. We used the centered difference scheme and the quadrature method to compute differential and integration. Fig. 2 shows the experimental results for the three sinusoids, where the multiple-SCDE (M-SCDE) indicates our method. The DFT-based methods resulted in a large estimation error. This should be because they are affected by the spectral leakage. In contrast, M-SCDE accurately estimated all frequencies even when the signal was 2 ms, which is shorter than the wavelength of the sinusoid of 440 Hz.

B. Evaluation on noisy conditions

We evaluated the estimation accuracy under noisy conditions. Assuming the combination of the proposed method with a rough frequency decomposition, such as [19] and [22], we used a band-limited noise with a bandwidth from 430 Hz to 490 Hz for a sum of three sinusoids whose frequencies are

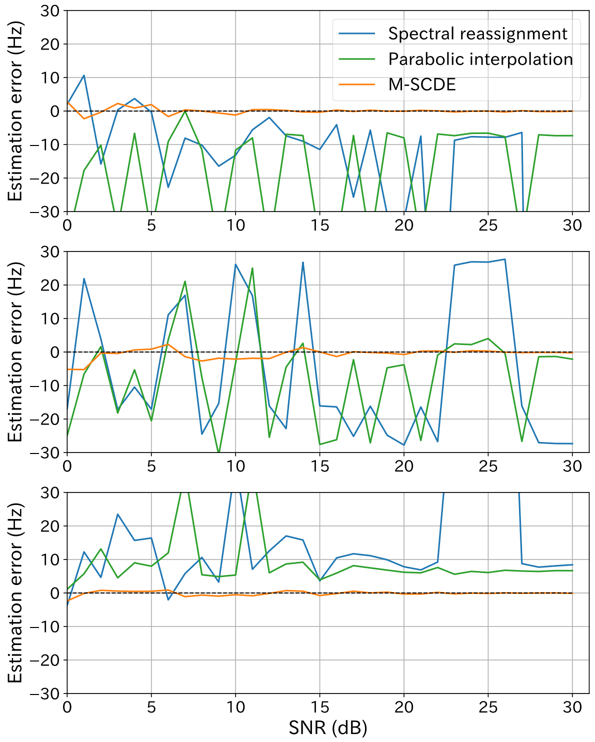

Fig. 3. Estimation accuracy for three sinusoids of 440 Hz (top), 460 Hz (middle), and 480 Hz (bottom) on the signal length by the three methods under noisy conditions.

440 Hz, 460 Hz, and 480 Hz. The signal length was 40 ms. Other experimental conditions were the same as in Section IV-A. Fig. 3 shows the estimation error for each sinusoid under noisy conditions. M-SCDE resulted in a smaller error than the DFT-based methods.

V. Conclusion

In this paper, we presented an extension of the SCDE-based methods for estimating the frequencies of three or more sinusoids. Based on the SCDE for multiple sinusoids, we solve the simultaneous equations related to the differentiation of the signal. Then, the frequencies are estimated by finding the roots of the polynomial equation on the basis of the simultaneous equations. Our experimental results confirmed that our method could accurately estimate the frequencies even when the DFT-based methods did not work.

Our future work includes the combination of the proposed method with frequency decomposition, such as [19] and [22]. The frequency decomposition will contribute to the noise robustness. Also, unlike conventional DFT-based methods, a rough decomposition, where one spectral peak contains a few sinusoids, should be sufficient for our method.

Acknowledgment

This work was supported by JST CREST Grant Number JPMJCR19A3, Japan.

References

[1] D. J. Hirst, "Form and function in the representation of speech prosody," *Speech, Commun.*, vol. 46, no. 3–4, pp. 334–347, 2005.

[2] M. Muller, D.P.W. Ellis, A. Klapuri, and G. Richard, "Signal processing for music analysis," *IEEE J. Sel. Top. Signal process.*, vol. 5, no. 6, pp. 1088–1110, 2011.

[3] J. Ventura, R. Sousa, and A. Ferreira, "Accurate analysis and visual feedback of vibrato in singing," in *Proc. IEEE Int. Symp. Commun., Control. Signal Process.*, 2012, pp. 1–6.

[4] E. Benetos, S. Dixon, Z. Duan, and S. Ewert, "Automatic music transcription: An overview," *IEEE Signal Process. Mag.*, vol. 36, no. 1, pp. 20–30, 2018.

[5] Y. Masuyama, T. Kusano, K. Yatabe, and Y. Oikawa, "Modal decomposition of musical instrument sounds via optimization-based non-linear filtering," *Acoust. Sci. Technol.*, vol. 40, no. 3, pp. 186–197, 2019.

[6] K. Yatabe and Y. Oikawa, "Phase corrected total variation for audio signals," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2018, pp. 656–660.

[7] Y. Masuyama, K. Yatabe, and Y. Oikawa, "Low-rankness of complex-valued spectrogram and its application to phase-aware audio processing," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2019, pp. 855–859.

[8] D. Rife and R. Boorstyn, "Single tone parameter estimation from discrete-time observations," *IEEE Trans. Inf. Theory*, vol. 20, no. 5, pp. 591–598, 1974.

[9] P. Stoica, R. L. Moses, B. Friedlander, and T. Soderstrom, "Maximum likelihood estimation of the parameters of multiple sinusoids from noisy measurements," *IEEE Trans. Acoust., Speech, Singal Process.*, vol. 37, no. 3, pp. 378–392, 1989.

[10] T. Abatzoglou, "A fast maximum likelihood algorithm for frequency estimation of a sinusoid based on newton's method," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 1, pp. 77–89, 1985.

[11] M. J. Turmon and M. I. Miller, "Maximum-likelihood estimation of complex sinusoids and toeplitz covariances," *IEEE Trans. Signal Process.*, vol. 42, no. 5, pp. 1074–1086, 1994.

[12] M. Ghogho, A.K. Nandi, and A. Swami, "Cramer-rao bounds and maximum likelihood estimation for random amplitude phase-modulated signals," *IEEE Trans. Signal Process.*, vol. 47, no. 11, pp. 2905–2916, 1999.

[13] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Minimization or maximization of functions*, Cambridge University Press, 1992.

[14] M. Abe and J. O. Smith, "Design criteria for simple sinusoidal parameter estimation based on quadratic interpolation of fft magnitude peaks," *J. Audio Eng. Soc.*, 2004.

[15] F. Auger and P. Flandrin, "Improving the readability of time-frequency and time-scale representations by the reassignment method," *IEEE Trans. Signal Process.*, vol. 43, no. 5, pp. 1068–1089, 1995.

[16] F. Auger, P. Flandrin, Y. Lin, S. McLaughlin, S. Meignen, T. Oberlin, and H. Wu, "Time-frequency reassignment and synchrosqueezing: An overview," *IEEE Signal Process. Mag.*, vol. 30, no. 6, pp. 32–41, 2013.

[17] M. D. Macleod, "Fast nearly ml estimation of the parameters of real or complex single tones or resolved multiple tones," *IEEE Trans. Signal Process.*, vol. 46, no. 1, pp. 141–148, 1998.

[18] F. Keiler and S. Marchand, "Survey on extraction of sinusoids in stationary sounds," in *Proc. Digit. Audio Effect. Conf.*, 2002, pp. 51–58.

[19] S. Ando and T. Nara, "An exact direct method of sinusoidal parameter estimation derived from finite Fourier integral of differential equation," *IEEE Trans. Signal Process.*, vol. 57, no. 9, pp. 3317–3329, 2009.

[20] S. Ando, "An exact direct method for short-time sinusoidal parameter estimation," in *Proc. IEICE Tech. Rep.*, 2007, vol. 107, pp. 25–30, in Japanese.

[21] S. Ando, "An exact direct method for short-time sinusoidal parameter estimation by means of finite fourier transform," in *Proc. IEICE Tech. Rep., EA*, 2007, vol. 107, pp. 37–42, in Japanese.

[22] S. Ando, "Frequency-domain prony method for autoregressive model identification and sinusoidal parameter estimation," *IEEE Trans. Signal Process.*, vol. 68, pp. 3461–3470, 2020.