On the Estimation of Sinusoidal Parameters Via Parabolic Interpolation of Scaled Magnitude Spectra

Marcelo Caetano, Philippe Depalle

ON THE ESTIMATION OF SINUSOIDAL PARAMETERS VIA PARABOLIC INTERPOLATION OF SCALED MAGNITUDE SPECTRA

Marcelo Caetano * *and Philippe Depalle* †

Schulich School of Music & CIRMMT
McGill University
Montreal, Quebec, Canada
marcelo.caetano@mcgill.ca | philippe.depalle@mcgill.ca

ABSTRACT

Sinusoids are widely used to represent the oscillatory modes of music and speech. The estimation of the sinusoidal parameters directly affects the quality of the representation. A parabolic interpolation of the peaks of the log-magnitude spectrum is commonly used to get a more accurate estimation of the frequencies and the amplitudes of the sinusoids at a relatively low computational cost. Recently, Werner and Germain [1] proposed an improved sinusoidal estimator that performs parabolic interpolation of the peaks of a power-scaled magnitude spectrum. For each analysis window type and size, a power-scaling factor p is pre-calculated via a computationally demanding heuristic. Consequently, the power-scaling estimation method is currently constrained to a few tabulated power-scaling factors for pre-selected window sizes, limiting its practical applications. In this article, we propose a method to obtain the power-scaling factor p for any window size from the tabulated values. Additionally, we investigate the impact of zero-padding on the estimation accuracy of the power-scaled sinusoidal parameter estimator.

1. INTRODUCTION

Sinusoids are widely used to represent the oscillatory modes of musical sounds [2] and speech [3]. The sound is typically modeled as a sum of time-varying sinusoids parameterized by their amplitudes, frequencies and phases. Sinusoidal parameter estimators play a crucial role in sinusoidal modeling because the quality of the representation depends on the accuracy of the estimation.

The estimation of the parameters of sinusoids in noise has a long history in the signal processing literature [4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1, 15]. Spectral estimators based on the interpolation of three DFT samples [14, 1, 15] commonly fit a parabola to local peaks of the magnitude spectrum. Parabolic interpolation on a logarithmic scale has been shown to outperform parabolic interpolation on linear scale for both stationary [16, 17] and non-stationary [18] sinusoids. Fig. 1 shows that the shape of the main lobe of the Hann window changes dramatically under linear or log scaling of the magnitude spectrum.

* This work has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 831852 (MORPH)

† Thanks to the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant (RGPIN- 2018-05662) for funding

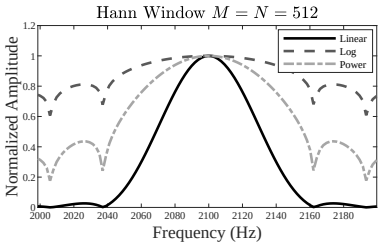


Figure 1: *Comparison of the shape of the magnitude spectrum of a Hann window under linear, logarithmic, and power scaling.*

High-resolution methods such as MUSIC [19] and ESPRIT [20] provide accurate estimation for multiple sinusoids in additive noise. Iterative estimation via least-squares [21] also converges to accurate values. However, these estimators suffer from the demanding computational cost of matrix algebra. Keiler and Marchand [22] compared the sinusoidal parameter estimators most commonly used for audio, including parabolic interpolation on a logarithmic scale. To the best of our knowledge, their performance has not been compared to recent state-of-the-art estimators [12, 13, 14, 1, 15].

Recently, Werner and Germain [1] developed an estimator based on parabolic interpolation of DFT samples of a power-scaled magnitude spectrum. They heuristically pre-calculated the power p that distorts the main lobe of the window to make its shape more parabolic and minimize the estimation error (see Fig. 1). Then, they compared the performance of their estimator with the state-of-the-art estimators of Duda [12] and Candan [14] in additive noise.

Werner and Germain determined p for 12 commonly used windows [23], showing that p depends on the window size M. Thus, obtaining p for non-tabulated M would require performing the computationally demanding heuristic. Consequently, power-scaling estimation is currently constrained to a few tabulated values (see Table 1 reproduced from [1]), severely limiting its practical applications. In this article, we propose a simple method to estimate p for any M starting from the existing tabulated values, bridging the gap between [1] and practical use of the proposed estimator. Additionally, we check the impact of zero-padding on spectral estimation via log and power-scaled parabolic interpolation.

The next section briefly reviews spectral estimation of sinusoidal parameters via parabolic interpolation of DFT samples of the magnitude spectrum. Then, the method for obtaining a power scaling parameter for any window size is presented, followed by a study into the impact of zero-padding on spectral estimation via

parabolic interpolation under log and power scaling. Finally, conclusions and perspectives are presented.

2. SPECTRAL ESTIMATION

Let us consider the discrete-time cisoid (i.e., complex sinusoid)

$$c\left(n\right) = a_0\, e^{j\left(2\pi\nu_0\, n + \phi_0\right)}. \tag{1}$$

where n is the sample index, a_0 is the amplitude, ν_0 is the normalized frequency, and ϕ_0 is the initial phase in radians. The sampling period $T_s = {}^1\!/_{f_s}$, where f_s is the sampling frequency, provides the connection between discrete-time samples n and time in seconds. So $t_n = n\,T_s$ are discrete-time values in seconds and $f_0 = \nu_0\, f_s$ is the fundamental frequency in Hertz. Since $c\left(n\right)$ has infinite support, let us consider further $x\left(n\right) = w\left(n\right) c\left(n\right)$, where $w\left(n\right)$ is a window function with M non-zero samples. The discrete-time Fourier transform (DTFT) of $x\left(n\right)$ is

$$X\left(\nu\right) = a_0\, e^{j\,\phi_0}\, W\left(\nu - \nu_0\right), \tag{2}$$

which is the DTFT of $w\left(n\right)$ shifted in frequency by ν_0 and scaled by the complex amplitude $a_0 e^{j\phi_0}$. Fig. 2 shows the main lobe of the Fourier transform of the Hann window modulated by a sinusoid with $a_0 = 1$ and $f_0 = 2100\,\text{Hz}$. The solid line in Fig. 2 is the magnitude of $X\left(\nu\right)$ *w-normalized* by $W\left(0\right) = \sum_0^{M-1} w\left(n\right)$. Since ν is a continuous variable in eq. (2), the estimations $\hat{\nu}_0$ and $\hat{a}_0$, of ν_0 and a_0 respectively, can be easily found at the maximum of $\left|X\left(\nu\right)\right|$ and lead to the unbiased values $\hat{\nu}_0 = \nu_0$ and $\hat{a}_0 = a_0$. However, the practical use of the discrete Fourier transform (DFT) only provides samples of $X\left(\nu\right)$ at discrete normalized frequencies

$$\nu_k = \frac{k}{N}, \quad k = 0, \ldots, N-1; \tag{3}$$

where N is the size of the DFT ($N \geq M$) and k is the discrete frequency index also called frequency bin. From now on we will consider the magnitude spectrum S defined as $S\left(k\right) = \left|X\left(\nu_k\right)\right|$, where $f_k = \nu_k f_s = \frac{k}{N} f_s$ is the frequency in Hz. Fig. 2 shows the *w-normalized* DFT samples $S\left(k\right)$ as *, $\left|X\left(\nu\right)\right|$ as the solid line, and the frequency bins k as the vertical dotted lines.

2.1. Nearest Neighbor Estimation

As shown in Fig. 2, ν_0 and a_0 can be estimated from the discrete-frequency spectrum $S\left(k\right)$ as a spectral peak (i.e., a local maximum). A local maximum of $S\left(k\right)$ is found at k_m whenever $S\left(k_m - 1\right) < S\left(k_m\right) > S\left(k_m + 1\right)$. Then, $\hat{\nu}_0 = \nu_{k_m}$ and $\hat{a}_0 = S\left(k_m\right)$ are the *nearest neighbor* estimates of ν_0 and a_0. The maximum of $\left|X\left(\nu\right)\right|$ actually lies at ν_0 in between two DFT samples. So, using eq. (3), it can be written as

$$\nu_0 = \frac{k_m + \delta k_m}{N} = \frac{k_0}{N}, \tag{4}$$

where k_m is an integer bin number and δk_m is the correction term called *bin offset* that, when added together, result in the *fractional bin* k_0. The estimation errors ε_ν and ε_a are defined as

$$\varepsilon_\nu\left(\nu_0\right) = \nu_0 - \hat{\nu}_0, \tag{5}$$

$$\varepsilon_a\left(a_0\right) = \frac{a_0 - \hat{a}_0}{a_0}. \tag{6}$$

So, according to eq. (4), nearest neighbor estimation yields $\varepsilon_\nu\left(\nu_0\right) = {}^{\delta k_m}\!/_N$ with $\left|\delta k_m\right| \leq 0.5$.

2.2. Parameter Estimation by Parabolic Interpolation

Parabolic interpolation improves the estimation by fitting a parabola to the main lobe of $\left|X\left(\nu\right)\right|$ to estimate δk_m from k_m and its immediate neighbors k_{m-1} and k_{m+1}. Fig. (2a) illustrates parabolic interpolation with the dashed line resulting from the fit of the parabola to the three points surrounding the maximum $P_{k_{m-1}}, P_{k_m}, P_{k_{m+1}}$, where point $P_{k_i} = \left(k_i, S\left(k_i\right)\right)$.

Given the shape of $W(\nu)$ for each window, the logarithm of the magnitude spectrum usually results in a better fit. Fig. (2b) illustrates such a case where a parabolic interpolation is performed after a logarithmic scaling of $S\left(k\right)$, i.e. on $\log_{10} S\left(k\right)$. Note how the parabola fits the main lobe better than in the linear case shown in Fig. (2a). However, there are still errors in frequency $\hat{\nu}_0$ and amplitude $\hat{a}_0$ estimates.

To further improve the parabolic fit, Werner and Germain [1] showed that a *well-chosen* power scaling of the magnitude spectrum $S\left(k\right)$ leads to better estimations of $\hat{\nu}_0$ and $\hat{a}_0$. They heuristically found the values of power p that distort the main lobe of the spectrum of the window $W(\nu)$ to make its shape more parabolic and thus decrease the estimation error ε_ν. However, p depends on the window type and window size M and the optimization procedure is complex and computationally demanding, requiring precomputation of p for practical use. For example, Table 1 lists the optimum value of p for four windows of different sizes as found by Werner and Germain [1]. Fig. (2c) illustrates parabolic interpolation with power scaling, where the parabola is fit to $S^p\left(k\right)$ with p taken from Table 1.

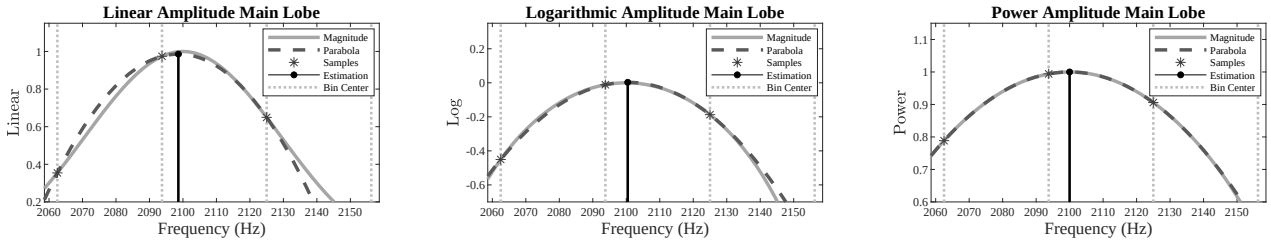
Table 1: *Optimum power scaling factor p for four window sizes M (in samples) and four window types. Values reproduced from [1].*

M	**512**	**1024**	**2048**	**4096**
Hann	0.22903	0.22911	0.22915	0.22917
Gaussian	0.12024	0.12074	0.12099	0.12112
Blackman-Harris	0.08552	0.08553	0.08553	0.08554
Dolph-Chebychev	0.08403	0.08403	0.08404	0.08404

Note that the mathematical operation denoted *exponentiation* written as $f\left(\beta, p\right) = \beta^p$, where both β and p are real positive values, has an opposite effect whether $\beta \in [0, 1]$ or $\beta \in (1, \infty)$ for a given value of p. Thus, we ensure that $S\left(k\right)$ lies in the interval $[0, 1]$ after being *w-normalized*. Since $p \in [0, 1]$ (cf. Table 1), $S^p\left(k\right)$ systematically bends outward above the identity line $g\left(\beta\right) = \beta$ for $\beta \in [0, 1]$.

2.3. Analysis Windows and Parabolic Estimation

The shape of the main lobe of the DTFT $W(\nu)$ of a window has an important impact on the parabolic fit, which, in turn, impacts the estimation. For example, the DTFT W_G of a Gaussian window w_G is also Gaussian. So its main lobe on a logarithmic scale is a prefect parabola and thus parabolic interpolation of $\log(|W_G|)$ is the best possible fit [17, 18]. However, w_G has infinite support and thus must be truncated in practice. This truncation distorts the parabolic shape of the main lobe of $\log|W_G|$. Another important example is the rectangular window w_R, whose main lobe width only spans two frequency bins when $N = M$. Since parabolic interpolation requires three samples, the use of w_R with parabolic estimation should be avoided unless a zero-padding step is added.



(a) Linear scaling. (b) Logarithmic scaling. (c) Power scaling.

Figure 2: *Parabolic interpolation for multiple scalings. The figure shows the main lobe of a* $N = 512$ *DFT of the Hann window with* $M = 512$ *samples modulated by a sinusoid with* $a_0 = 1$ *and* $f_0 = 2.1$ *kHz, (*$f_s = 44.1$ *kHz). Table 1 shows the value of p used in Fig. (2c).*

Finally, the estimation of multiple sinusoids is achieved simply by independently estimating the parameters of each sinusoid [7, 17] provided that the spectral peaks are properly resolved [23, 7, 17].

The choice of the windows for this study used two of the many possible features of the windows (see [23] for details), namely the width of the main lobe and the height of the side lobes. Thus, we considered the *Hann* window for its relatively narrow main lobe, the (truncated) *Gaussian* window ($\sigma^2 = 2.5$) for the theoretically approaching parabolic shape on a log scale, the minimum four-sample *Blackman-Harris* window for its very low side-lobe structure (-92 dB), and the *Dolph-Chebychev* window for the same reason (-100 dB for the chosen one).

2.4. Estimation Bias

In the absence of noise, the only source of error ε_a and ε_ν is the *bias* of the estimator, which is the systematic error due to its structure. Fig. 3 shows the estimation bias for ε_a and ε_ν respectively as a function of the bin offset $|\delta k_m| \leq 0.5$ for the *log* and *power* parabolic estimators using the *Hann*, *Gaussian*, *Blackman-Harris*, and *Dolph-Chebychev* [23] windows. We can see that both amplitude and frequency bias vary smoothly. Note the different orders of magnitude for the log and power cases, indicating an increasing order of estimation accuracy. Finally, these curves are independent of the values of ν_0 and a_0, and also of M as long as $N = M$.

The estimation bias curves shown in Fig. 3 allow us to define the *maximum* estimation error $\tilde{\varepsilon}$ as

$$\tilde{\varepsilon} = \max_{\delta k \in [0,\, 0.5]} |\varepsilon\left(\delta k\right)|, \tag{7}$$

where the interval for δk is restricted to $[0,\, 0.5]$ because the estimation bias curves are symmetrical around 0. Note that the definition of the maximum estimation error is the same for the amplitude error $\tilde{\varepsilon}_a$ and the frequency bin error $\tilde{\varepsilon}_\nu$. Additionally, the mean estimation error $\bar{\varepsilon}$ is given as

$$\bar{\varepsilon} = 2 \int_0^{0.5} |\varepsilon\left(\delta k\right)| \, d\delta k. \tag{8}$$

Fig. 3 shows $\bar{\varepsilon}$ and $-\bar{\varepsilon}$ as solid horizontal lines delimiting the range of ε_ν and of ε_a, and $\bar{\varepsilon}$ as the dashed horizontal line.

3. POWER SCALE FOR ANY WINDOW SIZE

As already mentioned, Werner and Germain [1] provide the power scaling factors p shown in Table 1, evaluated via an optimization process for a few window sizes. Consequently, estimation with power scaling is currently constrained to the four tabulated window sizes M. They speculated that p depends on M in a structured way, as shown in Fig. 4 reproduced from [1]. Fig. 4 shows four curves $p\left(M\right)$ corresponding to the the maximum frequency estimation error $\tilde{\varepsilon}_\nu$, the maximum amplitude estimation error $\tilde{\varepsilon}_a$, the mean frequency estimation error $\bar{\varepsilon}_\nu$, and the mean amplitude estimation error $\bar{\varepsilon}_a$ for the Hann window. The horizontal axis of Fig. 4 is labeled with values of M but varies linearly in octaves as $m = \log_2 M$. They also suggested to use piece-wise linear interpolation of $p\left(M\right)$ for values of M not found in Table 1.

We note, however, that the curves in Fig. 4 visually resemble the step response of a first order linear time-invariant system. Therefore, in this work, we propose to model $p(m)$ as

$$p\left(m\right) = \alpha\tau \left[1 - e^{\left(-\frac{m - m_0}{\tau}\right)}\right], \tag{9}$$

parameterized by the arbitrary constants m_0, τ, and α. In general terms, m_0 shifts the curves along the horizontal axis, τ controls the rate of growth and $\alpha\tau$ determines the horizontal asymptote. The next step is to fit the curve in eq. (9) that models $p\left(m\right)$ to get an estimate of the optimal p for any window size M (not limited to the tabulated values) without having to perform the costly optimization. We propose to first reformulate the relationship between p and M into a linear expression, and then perform a linear fit of the form $y = ax + b$ whose coefficients would allow us to predict p for arbitrary window sizes.

3.1. Linear Regression

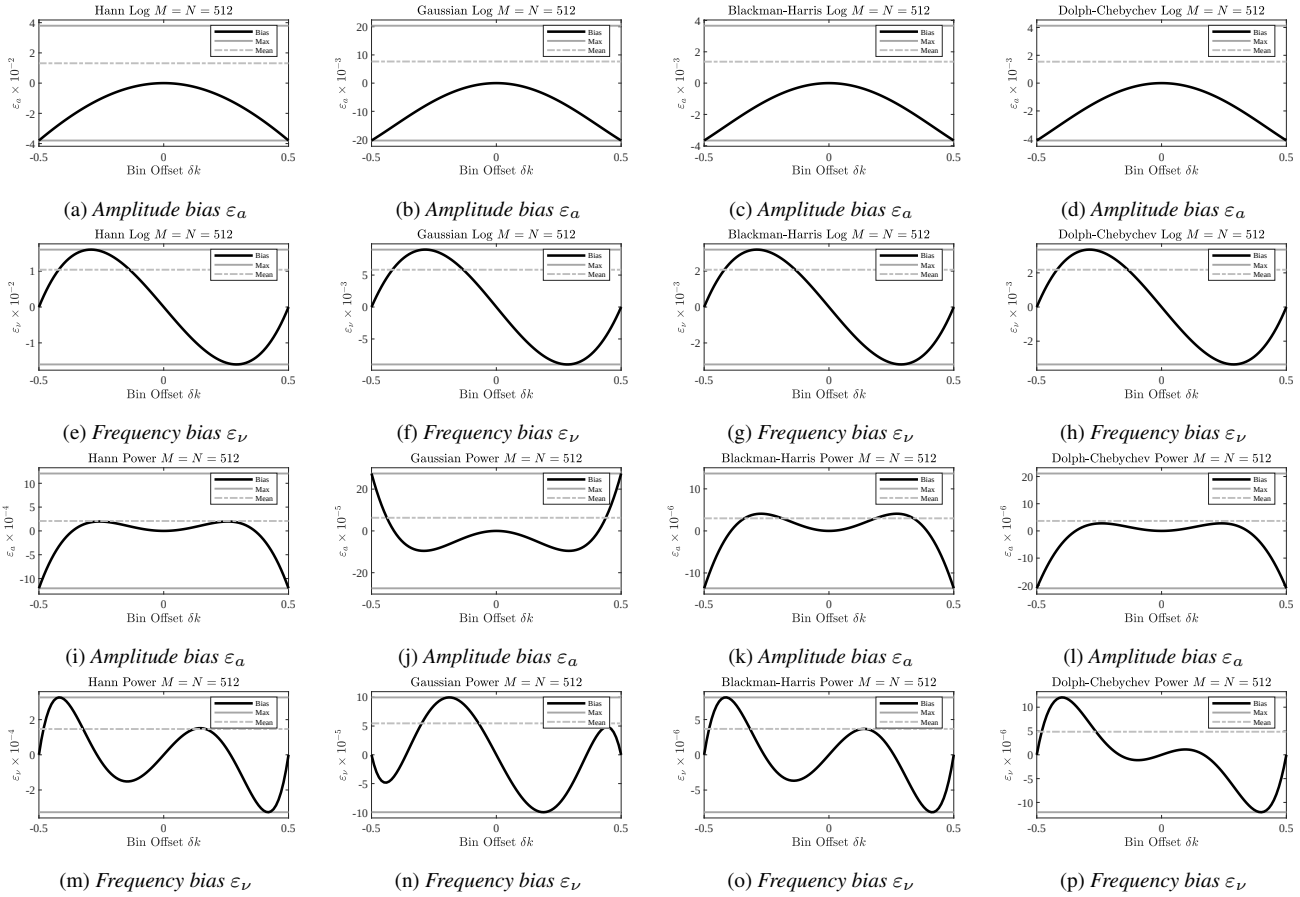
As noted above, the horizontal axis is linearized by

$$m = \log_2\left(M\right). \tag{10}$$

If we assume that the general expression of eq. (9) appropriately models $p\left(m\right)$, we can rewrite it as

$$\ln\left[1 - \frac{p\left(m\right)}{\kappa}\right] = a\, m + b, \tag{11}$$

where $a = -\frac{1}{\tau}$, $b = \frac{m_0}{\tau}$, and $\kappa = \alpha\tau$. We will fit a straight line to the relationship in eq. (11) by considering *observed* values of p and M as the ones from Table 1. However, eq. (11) depends on an unknown constant κ. In this study, we will determine the optimal value of κ as the one that best explains the linear dependence that links the values as in eq. (11). In what follows, we describe how κ is empirically determined to fulfill this constraint.



(a) *Amplitude bias* ε_a (b) *Amplitude bias* ε_a (c) *Amplitude bias* ε_a (d) *Amplitude bias* ε_a

(e) *Frequency bias* ε_ν (f) *Frequency bias* ε_ν (g) *Frequency bias* ε_ν (h) *Frequency bias* ε_ν

(i) *Amplitude bias* ε_a (j) *Amplitude bias* ε_a (k) *Amplitude bias* ε_a (l) *Amplitude bias* ε_a

(m) *Frequency bias* ε_ν (n) *Frequency bias* ε_ν (o) *Frequency bias* ε_ν (p) *Frequency bias* ε_ν

Figure 3: *Amplitude and frequency estimation bias for log scaling (top) and power scaling (botom) with optimal value of p.*

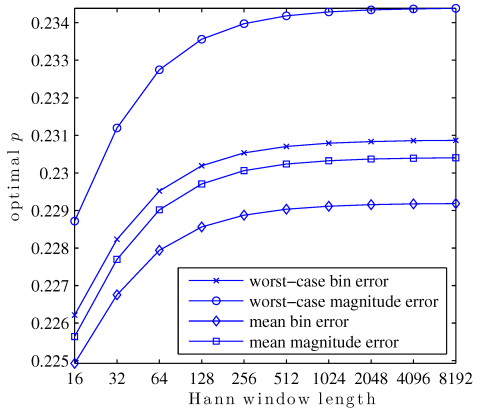


Figure 4: Original curves $p\,(M)$ from [1]

We know from eq. (9) that κ is the upper asymptote of $p(m)$, thus $\kappa > p\,(m)$. Note, that this is also the condition that keeps the argument of the logarithm in eq. (11) strictly positive. Therefore, we seek the value of $\kappa > p\,(m)$ that maximizes the linearity of eq. (11). The empirical method relies on the use of the coefficient of determination R^2 to measure the linearity of the model. This is performed by sweeping the value of κ (starting at $\kappa = p\,(12) + \mu$, with μ arbitrarily small) and selecting the one that corresponds to the highest R^2 for each window in Table 1. The solid line in each panel of Fig. 5 is $R^2\,(\kappa)$ and the vertical dashed line indicates the value of κ that maximizes R^2 listed in the first column of Table 2.

Fig. 5 reveals that $R^2\,(\kappa)$ for the Hann and Gaussian windows have a clear maximum, whereas $R^2\,(\kappa)$ has no clear maximum for the Blackman-Harris and Dolph-Chebychev windows. For Blackman-Harris, $R^2\,(\kappa)$ reaches a plateau at approximately $R^2 = 0.9$, suggesting that increasing κ above an initial threshold does not improve the linearity of eq. (11). For this window, Table 2 shows the highest value of κ used in the sweep, but Fig. 5 suggests that any $\kappa > 0.1$ would suffice. For the Dolph-Chebychev window, we get a similar behavior than the one observed for the Blackman-Harris window, with a $R^2\,(\kappa)$ that is constant at around $R^2 = 0.8$ with respect to κ. These low R^2 values for the two last windows suggest that the model of $p(m)$ might be different from the one proposed in eq.(9), as discussed below.

Table 2 shows the values of κ that maximize the linearity of eq. (11) and the corresponding a and b for the estimated model of $p\,(M)$. Using eq. (9) and eq. (10), the final model be written as

$$p\,(M) = \kappa \left[1 - e^{(a\,\log_2(M)+b)}\right]. \tag{12}$$

Fig. 6 shows the line fit on the top panel and the corresponding $p\,(m)$ on the bottom panel. The dots are predicted by the model

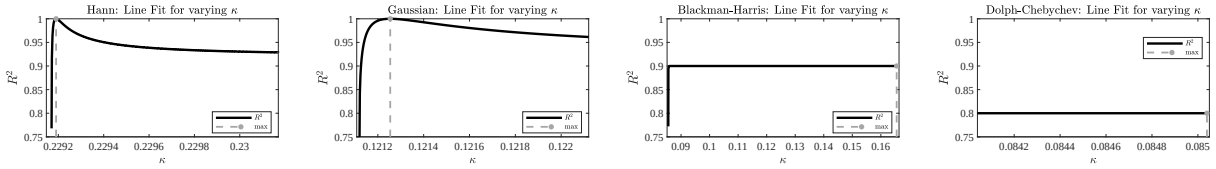


Figure 5: R^2 *as a function of* κ *for the windows under investigation.*

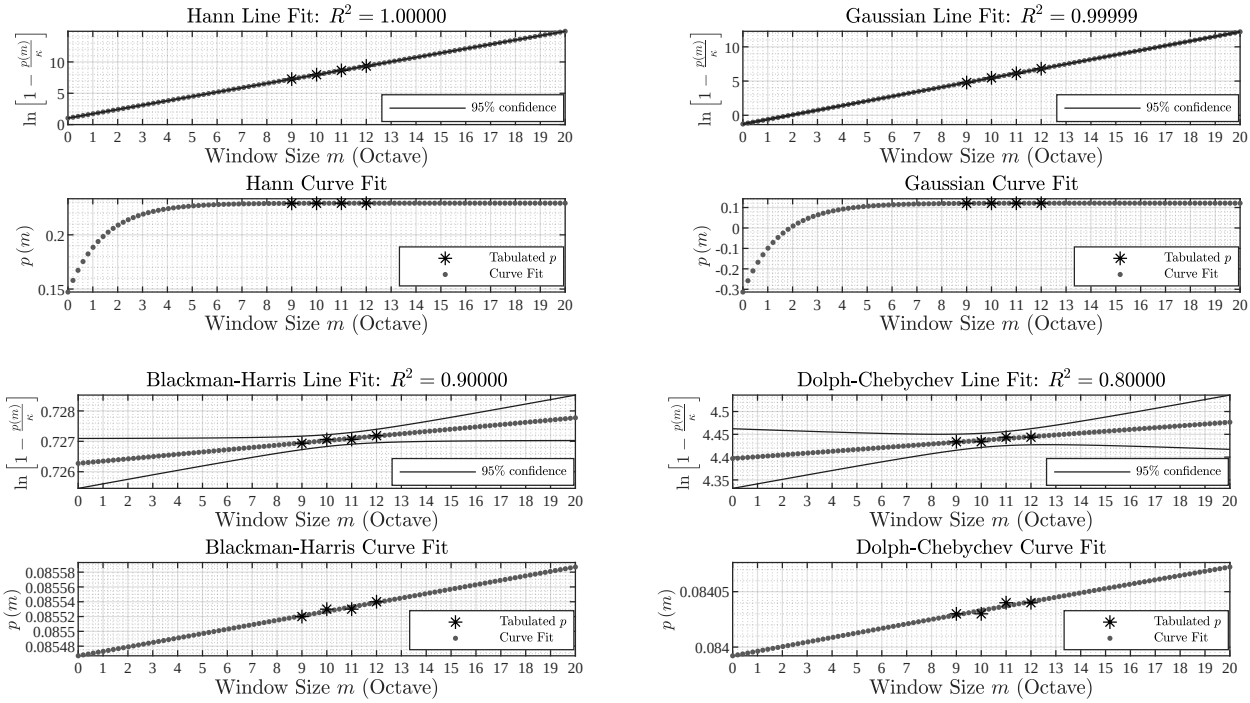


Figure 6: *Interpolation of optimal p for the windows under investigation. The top panel shows the line fit with the corresponding* R^2 *and the* 95% *confidence interval. The bottom panel shows the curve fit* $p\,(m)$*. The optimal values of p from Table 1 are marked as * whereas the values predicted by the model are represented with dots.*

Table 2: *Line fit parameters for the windows under investigation.*

Window	κ	a	b
Hann	0.22919	-0.69315	-1.0288
Gaussian	0.12125	-0.67356	1.2765
Blackman-Harris	0.16554	-7.4991×10^{-5}	-0.72627
Dolph-Chebychev	0.085039	-3.9824×10^{-3}	-4.3969

whereas * represents the optimal p data from Table 1. The top panel in Fig. 6 also shows the 95% confidence interval (solid lines) around the linear fit. Note that $p\,(M)$ is an analytic function of M, allowing to get p for any M, not only between tabulated values (interpolation) but also beyond them (extrapolation), as shown in Fig. 6. Naturally, the prediction power of $p\,(M)$ for each window depends on the line fit. The values of R^2 can be interpreted as an indication of how well p and M in Table 1 fit the model of eq. (9), or how well $p\,(M)$ in eq. (9) using values from Table 2 captures the relationship between the *optimal* values of p and the tabulated values of M in Table 1.

Fig. 6 shows that the Hann window fits the model perfectly while the Gaussian window has nearly perfect fit. However, the Blackman-Harris and Dolph-Chebychev windows have lower fits, which can be visually confirmed on the top panel of Fig. 6 by inspecting the alignment of *. Additionally, a for the Blackman-Harris and Dolph-Chebychev windows are respectively 4 and 2 orders of magnitude smaller than those for the Hann and Gaussian windows. Note that $a = -1/\tau$, so a is inversely proportional to the rate of growth τ in eq. (9). Given the poor fit (low R^2 and a) to the model of eq. (12) for the Blackamn-Harris and Dolph-Chebychev windows (see Fig. 6), we conclude that a linear model of $p(m)$ is a better fit for these two windows. This conclusion is confirmed by visual inspection of $p(m)$ on the bottom panel of Fig. 6.

Although not formally proven, analyses performed on other types of windows seem to confirm that windows with a Fourier transform that exhibits a narrow main lobe fit well the model of eq. (12) while those that exhibit a wide main lobe and a low side-lobe structure fit a linear model better than the one in eq. (12)

in the considered range of M ($\leq 2^{20}$). See the results for all 12 windows at `https://marcelo-caetano.github.io/dafx2021.html`, which also contains sound examples.

4. IMPACT OF ZERO-PADDING

This section reports on the impact of zero-padding for the estimation of frequencies and amplitudes of sinusoids via parabolic interpolation for logarithmic and power scale. We perform parabolic interpolation of the three consecutive bins corresponding to the spectral peak of the scaled and zero-padded magnitude DFT. More specifically, we want to compare the use of zero-padding and power-scale for the quality of the estimation.

As is widely known and briefly reviewed at the beginning of Sec. 2, the DTFT of a signal x of duration M is $X(\nu)$, and its DFT for a size $N \geq M$ is simply the sampling of $X(\nu)$ at evenly spaced frequencies $\nu_k = \frac{k}{N}$ with $0 \leq k < N$. Increasing N results in a finer sampling of the same shape $X(\nu)$ and provides spectral samples that are closer together. This is known as *zero-padding* due to its practical implementation by padding a vector of M non-zero samples with $M-N$ zeroes before computing the DFT on the resulting vector of size N. Overall, we expect to get better results for higher values of N as the three points used for the parabolic interpolation get closer to the maximum, and consequently are a better fit to a Taylor expansion of order 2 around 0 of a scaled version of $X(\nu)$.

In theory, we can make ε_ν and ε_a arbitrarily small by increasing N via zero-padding while holding M fixed. However, even if we use the FFT algorithm, this *brute force* nearest-neighbor estimator will become too computationally demanding as N grows. In practice, it is common to combine zero-padding with parabolic interpolation [16].

Defining $L = N/M$ as the *zero-padding factor*, Smith and Serra [16] suggest to use $L \approx 5$ and a parabolic interpolation on the log-scaled magnitude spectrum to achieve values of ε_ν of the order of 0.05% of the width of the main lobe. Hereafter, this estimator will be denoted *ZP-Log-PI* for the sequence of operations: Zero-Padding, Log scaling, and Parabolic Interpolation. By extension, the estimator *ZP-Pow-PI* uses *power* scaling. Additionally, the absence of the prefix *ZP* will denote *PI* when $N = M$. For example, *Pow-PI* denotes the parabolic interpolation with power scaling and no zero-padding ($N = M$) as proposed by Werner and Germain [1]. Next, we determine the zero-padding factor L for ZP-Log-PI that results in performance comparable to Pow-PI.

4.1. Zero-Padding and Log-Scaled Parabolic Interpolation

In this section, we will determine how much zero-padding is required by the ZP-Log-PI estimator to get estimation errors below those of Pow-PI. The motivation is to decide whether ZP-Log-PI can be a practical substitute for Pow-PI. For such, we calculate ε_ν and ε_a for ZP-Log-PI as a function of L and compare with Pow-PI, as illustrated in Fig. 7 for the Hann window. The solid line in Fig. 7 is ε_a as a function of L for ZP-Log-PI and the horizontal dashed line is ε_a for Pow-PI when $N = M$ (so $L = 1$). Note how zero-padding dramatically decreases the estimation error ε_a for the Hann window. As the figure shows, ε_a for ZP-Log-PI goes below that of Pow-PI for $L \approx 2$. Therefore, we need to set $N \geq 2M$ to guarantee that parabolic interpolation on a log scale will result in ε_a below the one obtained with power scaling when $N = M$.

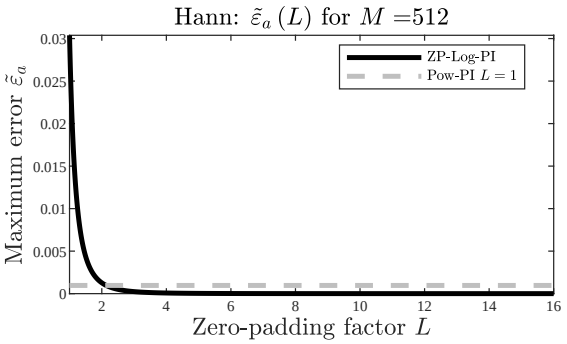


Figure 7: *Maximum amplitude estimation error as a function of the zero-padding factor for the Hann window. The figure illustrates how to determine the minimum amount of zero-padding with the ZP-Log-PI estimator to achieve estimation errors below those of the Pow-PI estimator.*

Table 3: *Minimum zero-padding factor where ZP-Log-PI results in estimation error below that of Pow-PI.*

Window	min L for ε_a	min L for ε_ν
Hann	2	6
Gaussian	3	9
Blackman-Harris	4	> 16
Dolph-Chebychev	4	> 16

Table 3 shows the minimum value of L for which ZP-Log-PI results in ε_a and ε_ν below those of Pow-PI. Firstly, notice that the values of L for the amplitude error are consistently smaller than those for the frequency error. Additionally, we note that the Blackman-Harris and Dolph-Chebychev windows require $L > 16$ for ε_ν, making ZP-Log-PI an impractical alternative to Pow-PI when we require comparable frequency estimation errors. Naturally, we must use the value of L for ε_ν if we want to guarantee that both ε_a and ε_ν for ZP-Log-PI will fall below those of Pow-PI. In that case, ZP-Log-PI is not a beneficial alternative for Pow-PI for the windows shown in Table 3. Both Blackman-Harris and Dolph-Chebychev require $L > 16$, which is impractical in most situations. The Gaussian window requires $L = 9$, which is also rather impractical. For the Hann window, $L = 6$ would result in fairly high N depending on the minimum required value for M. For example, the analysis of a *C3* note corresponding to a fundamental frequency $f_0 \approx 131$ Hz using a window that spans approximately $3T_0$ to get enough frequency resolution, requires $M > 1000$ samples when $f_s = 44.1$ kHz, which, in turn, requires a DFT with $N > 6000$ for ZP-Log-PI with the Hann window to achieve performance similar to Pow-PI with $M = N = 1024$.

Finally, we note that the errors ε_a and ε_ν have different orders of magnitude depending on the chosen window. Table 4 shows ε_a and ε_ν for Pow-PI when $N = M = 512$. Firstly, notice that the Blackman-Harris and the Dolph-Chebychev windows yield estimation errors below of those for the Hann and Gaussian windows. Consequently, using ZP-Log-PI to match ε_ν with Pow-PI for the Blackman-Harris and Dolph-Chebychev windows requires a larger L to decrease ε_ν sufficiently. This observation is consistent with

the pattern of minimum L in Table 3 because, in general, L for ZP-Log-PI must be higher to match a lower $\tilde{\varepsilon}$ for Pow-PI. A practical strategy to achieve lower estimation errors is then to select the windows appropriately. According to Table 4, the Blackman-Harris window is the best choice and the Hann window should be avoided. However, zero-padding may also be combined with power-scaling parabolic estimation (ZP-Pow-PI). The next section investigates the impact of zero-padding in parabolic interpolation over power-scaled samples of the magnitude spectrum.

Table 4: *Maximum amplitude and frequency estimation errors for power scaling parabolic interpolation when* $N = M = 512$.

Window	ε_a	ε_ν
Hann	9.6417×10^{-4}	3.2654×10^{-4}
Gaussian	2.1980×10^{-4}	9.9713×10^{-5}
Blackman-Harris	1.0921×10^{-5}	8.1425×10^{-6}
Dolph-Chebychev	1.6889×10^{-5}	1.2056×10^{-5}

4.2. Zero-Padding and Power-Scaled Parabolic Interpolation

In this section, we will investigate the impact of zero padding on the power-scaled parabolic interpolation. As before, we measure ε_a and ε_ν as a function of L for the four windows, as shown in Fig. 8. The motivation is to visualize the behavior of ZP-Pow-PI and understand whether zero-padding is also beneficial for parabolic estimation with power-scaling. Otherwise, we are constrained to the condition $N = M$ of Pow-PI. Additionally, we aim to compare the estimation performance using the model $p(M)$ of eq. (12) against the original optimal values of p from Table 1. Fig. 8 shows $\varepsilon_a(L)$ and $\varepsilon_\nu(L)$ for three values of M corresponding to different conditions, namely *extrapolation* ($M = 500$) and *interpolation* ($M = 600$) using eq. (12), as well as *tabulated* p ($M = 512$). Firstly, we note that the errors have different orders of magnitude depending on the window as shown in Table 4. Additionally, we note that the curves for *tabulated* ($M = 512$), *extrapolation* ($M = 500$) and *interpolation* ($M = 600$) present comparable values across L for all windows except for *interpolation* ($M = 600$) for the Gaussian window, which deviates slightly from the others (around the maximum between $L = 1$ and $L = 1.5$).

More importantly, the curves $\varepsilon_a(L)$ and $\varepsilon_\nu(L)$ do indicate that zero-padding decreases further the estimation error for parabolic estimation with power scaling. However, comparing Fig. 7 with Fig. 8 reveals a different behavior of ZP-Log-PI and ZP-Pow-PI. The error decreases monotonically for ZP-Log-PI as L increases, whereas ZP-Pow-PI has a range of values of L for which the error *increases* after an initial steep decline. For the Hann and Gaussian windows, this increase remains below the initial value of ε, so zero-padding is always beneficial (even if not monotonically decreasing with L). However, zero-padding can worsen the estimation performance for the Blackman-Harris and Dolph-Chebychev windows depending on L. For these windows, the condition $L > 3$ always results in beneficial estimation performance, so a suggested rule of thumb is $4 \leq L \leq 5$.

The implementation of the sinusoidal model available at `https://github.com/marcelo-caetano/sinusoidal-model` includes power-scaling estimation of the sinusoidal parameters with the model $p(M)$ described here.

5. CONCLUSIONS

The estimation of the amplitudes and frequencies of sinusoids embedded in a signal mixture is an important topic in several applications of signal processing, and more specifically in audio processing. Sinusoidal parameter estimation classically requires estimating both the position and the height of local maxima of a scaled version of the magnitude spectrum. This typically relies on a parabolic interpolation. In this paper, we propose a few improvements and further characterization of a method recently developed by Werner and Germain [1], who worked on a power-scaled magnitude spectrum. They heuristically determined the power scale p for each window type and a few window sizes. In this article, we propose a simple model to get the power-scaling factor p for any window size M. Parameters of the model of p are derived from the original optimal tabulated values of p available in [1].

We further re-visited the role of zero-padding in this context to see how much zero-padding a parabolic estimation performed on log-magnitude peaks requires presenting performance comparable to parabolic estimation of power-scaled peaks without zero-padding. We found that the Hann window requires the lowest zero-padding factor ($L = 2$ for amplitude and $L = 6$ for frequency) while both Blackman-Harris and Dolph-Chebychev windows require $L > 16$ getting comparable results. Additionally, we verified that zero-padding also decreases the estimation error for parabolic estimation of power-scaled peaks. However, $L > 3$ must be satisfied to ensure that zero-padding is beneficial for any window.

We also found two groups of windows with different behavior. The Hann and Gaussian windows consistently behave comparably, so do the Blackman-Harris and Dolph-Chebychev windows. This different behavior seems to be linked to their main features in terms of width of the main lobe, and height of the side-lobes. It appears that the wider the main lobe and the lower the side lobes, the more linear the evolution of $p(M)$ is (e.g. for Blackman-Harris and Dolph-Chebychev). Conversely, the narrower the main lobe and the higher the side lobes, the better $p(M)$ fits the shape of a unit step response to a first order filter (for Hann and Gaussian).

Finally, there are conditions that change the shape of the main lobe of the window, such as additive noise, adjacent sinusoidal peaks, and non-stationary sinusoids. Among those, future work includes investigating the robustness of the method to non-stationary sinusoids.

6. REFERENCES

[1] K. J. Werner and F. G. Germain, "Sinusoidal parameter estimation using quadratic interpolation around power-scaled magnitude spectrum peaks," *Applied Sciences*, vol. 6, no. 10, pp. 306, 2016.

[2] X. Serra and J. O. Smith, "Spectral Modeling Synthesis: A Sound Analysis/Synthesis Based on a Deterministic plus Stochastic Decomposition," *Computer Music Journal*, vol. 14, pp. 12–24, 1990.

[3] R. McAulay and T. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Trans Acoust, Speech, Sig Proc*, vol. 34, no. 4, pp. 744–754, 1986.

[4] D. C. Rife and G. A. Vincent, "Use of the discrete Fourier transform in the measurement of frequencies and levels of tones," *Bell Sys Tech J*, vol. 49, no. 2, pp. 197–228, 1970.

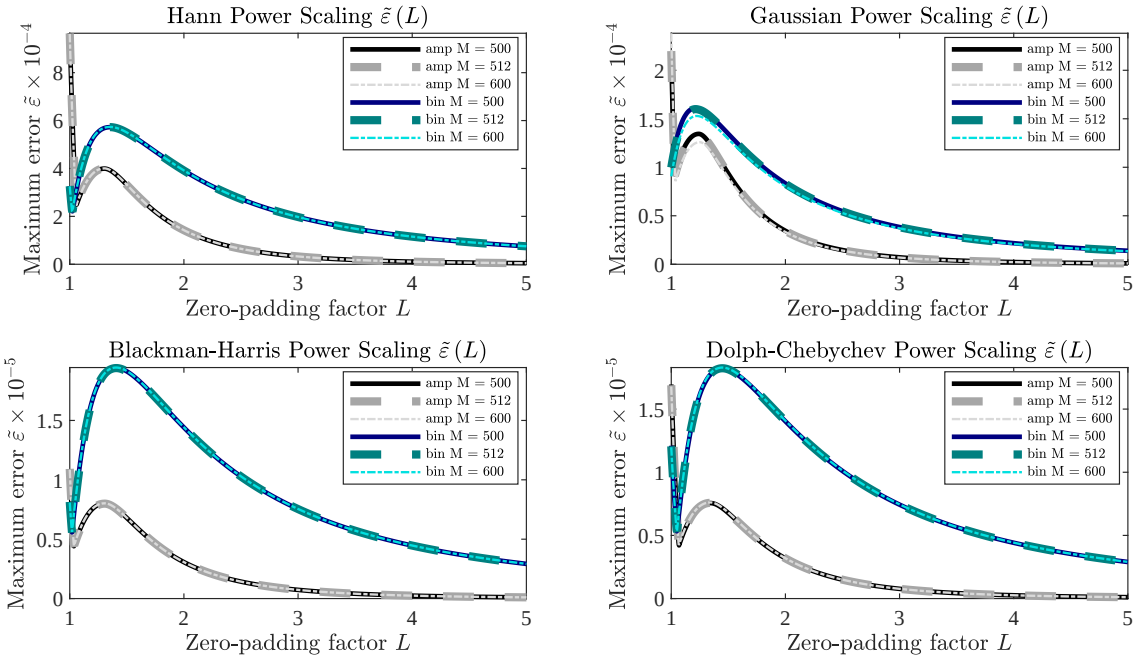


Figure 8: *The impact of zero-padding in power-scaling parabolic interpolation for the windows under investigation.*

[5] S. Kay, "A fast and accurate single frequency estimator," *IEEE Trans Acoust, Speech, Sig Proc*, vol. 37, no. 12, pp. 1987–1990, 1989.

[6] B. G. Quinn, "Estimation of frequency, amplitude, and phase from the DFT of a time series," *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 814–817, 1997.

[7] M. D. MacLeod, "Fast nearly ML estimation of the parameters of real or complex single tones or resolved multiple tones," *IEEE Transactions on Signal Processing*, vol. 46, no. 1, pp. 141–148, 1998.

[8] Y. C. Xiao, P. Wei, X. C. Xiao, and H. M. Tai, "Fast and accurate single frequency estimator," *Electronics Letters*, vol. 40, no. 14, pp. 910–911, 2004.

[9] E. Jacobsen and P. Kootsookos, "Fast, accurate frequency estimators [DSP tips & tricks]," *IEEE Signal Processing Magazine*, vol. 24, no. 3, pp. 123–125, 2007.

[10] S. Provencher, "Estimation of complex single-tone parameters in the DFT domain," *IEEE Transactions on Signal Processing*, vol. 58, no. 7, pp. 3879–3883, 2010.

[11] S. Provencher, "Parameters estimation of complex multitone signal in the DFT domain," *IEEE Transactions on Signal Processing*, vol. 59, no. 7, pp. 3001–3012, 2011.

[12] K. Duda, "DFT interpolation algorithm for Kaiser-Bessel and Dolph-Chebyshev windows," *IEEE Trans Instrumentation and Measurement*, vol. 60, no. 3, pp. 784–790, 2011.

[13] D. Belega and D. Petri, "Frequency estimation by two- or three-point interpolated Fourier algorithms based on cosine windows," *Signal Processing*, vol. 117, pp. 115–125, 2015.

[14] Ç. Candan, "Fine resolution frequency estimation from three DFT samples: Case of windowed data," *Signal Processing*, vol. 114, pp. 245–250, 2015.

[15] L. Fan and G. Qi, "Frequency estimator of sinusoid based on interpolation of three DFT spectral lines," *Signal Processing*, vol. 144, pp. 52–60, 2018.

[16] J. Smith and X. Serra, "PARSHL: An analysis/synthesis program for non- harmonic sounds based on a sinusoidal representation," in *Proc of Int Computer Music Conf (ICMC)*, 1987, pp. 290–297.

[17] M. Abe and J. O. Smith, "Design criteria for simple sinusoidal parameter estimation based on quadratic interpolation of FFT magnitude peaks," in *Proc of Audio Engineering Society Convention (AES) 117*, 2004.

[18] M. Abe and J. O. Smith, "AM/FM rate estimation for time-varying sinusoidal modeling," in *Proc of IEEE Int Conf Acoustics, Speech, and Signal Processing (ICASSP)*, 2005, vol. 3, pp. 201–204.

[19] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[20] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Trans Acoust, Speech, Sig Proc*, vol. 37, no. 7, pp. 984–995, 1989.

[21] Y. Pantazis, O. Rosec, and Y. Stylianou, "Iterative estimation of sinusoidal signal parameters," *IEEE Signal Processing Letters*, vol. 17, no. 5, pp. 461–464, 2010.

[22] F. Keiler and S. Marchand, "Survey on Extraction of Sinusoids in Stationary Sounds," in *Proc of Int Conf Digital Audio Effects (DAFx)*, 2002, pp. 51–58.

[23] F. J. Harris, "On the use of windows for harmonic analysis with the discrete Fourier transform," *Proceedings of the IEEE*, vol. 66, no. 1, pp. 51–83, 1978.