

Article

Demystifying DFT-Based Harmonic Phase Estimation, Transformation, and Synthesis

Marco Oliveira ^{1,†} , Vasco Santos ¹ , André Saraiva ^{1,2}  and Aníbal Ferreira ^{1,3,*,†} 

¹ Department of Electrical and Computer Engineering, Faculty of Engineering, University of Porto, 4099-002 Porto, Portugal; marcoantmoliveira@gmail.com (M.O.); vasco.f.santos@gmail.com (V.S.); andre.saraiva@pj.pt (A.S.)

² Forensic Science Laboratory, Judiciary Police, 1169-007 Lisbon, Portugal

³ INESC TEC, 4200-465 Porto, Portugal

* Correspondence: ajf@fe.up.pt

† These authors contributed equally to this work.

Abstract: Many natural signals exhibit quasi-periodic behaviors and are conveniently modeled as combinations of several harmonic sinusoids whose relative frequencies, magnitudes, and phases vary with time. The waveform shapes of those signals reflect important physical phenomena underlying their generation, requiring those parameters to be accurately estimated and modeled. In the literature, accurate phase estimation and modeling have received significantly less attention than frequency or magnitude estimation. This paper first addresses accurate DFT-based phase estimation of individual sinusoids across six scenarios involving two DFT-based filter banks and three different windows. It has been shown that bias in phase estimation is less than 0.001 radians when the SNR is equal to or larger than 2.5 dB. Using the Cramér–Rao lower bound as a reference, it has been demonstrated that one particular window offers performance of practical interest by better approximating the CRLB under favorable signal conditions and minimizing performance deviation under adverse conditions. This paper describes the development of a shift-invariant phase-related feature that characterizes the harmonic phase structure. This feature motivates a new signal processing paradigm that greatly simplifies the parametric modeling, transformation, and synthesis of harmonic signals. It also aids in understanding and reverse engineering the phasegram. The theory and results are discussed from a reproducible perspective, with dedicated experiments supported by code, allowing for the replication of figures and results presented in this paper and facilitating further research.

Keywords: phase estimation; harmonic phase structure; harmonic magnitude and phase parametric modeling; harmonic signal processing



Citation: Oliveira, M.; Santos, V.; Saraiva, A.; Ferreira, A. Demystifying DFT-Based Harmonic Phase Estimation, Transformation, and Synthesis. *Signals* **2024**, *5*, 841–868. <https://doi.org/10.3390/signals5040046>

Academic Editor: Jozef Juhár

Received: 22 August 2024

Revised: 30 September 2024

Accepted: 22 November 2024

Published: 4 December 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

1.1. Motivation

Many natural and synthetic quasi-periodic signals, including speech, singing, physiological signals such as ECG, music, and acoustic waves from mechanical system vibrations, have a harmonic structure of sinusoids whose magnitudes, phases, and underlying fundamental frequencies vary over time.

Harmonic phases are crucial in defining the waveform shape of quasi-periodic signals and are therefore immensely informative about the physical phenomena that generate them. Examples include the periodic glottal excitation signal, which illuminates the physiological processes governing vocal fold vibrations in the larynx; and periodic acoustic signals from mechanical systems, which provide insights into whether these systems are operating correctly within predefined safety margins.

Given that harmonic phases depend explicitly on time, they vary much faster than harmonic magnitudes and fundamental frequencies; this is challenging from the perspectives of signal analysis, estimation, interpretation, modeling, transformation, and synthesis [1–4].

Since accurate frequency and magnitude estimation of sinusoids have been extensively discussed in the literature [5–13], in this paper, we assume that these factors have been addressed and will instead focus on two problems related to phase. The first concerns the practical and accurate DFT-based phase estimation of individual sinusoids. This is instrumental in addressing a second problem: the parametric modeling of sinusoid phases within a harmonic structure in a way that is time-shift invariant, interpretable, insightful, and simplifies harmonic signal processing. To the best of our knowledge, this is the first time such a combined perspective is presented in a manner that is easily apprehensible and extensively illustrated.

This paper demonstrates that these problems can be tackled practically, facilitated by Matlab code that replicates the main results and illustrations presented.

1.2. Problem Statement

This paper focuses on practical methods for representing quasi-periodic signals through the concepts of structure and parametric modeling. By ‘structure’, we mean an identifiable form or organization, and by ‘parametric modeling’, we mean a simple mathematical formulation that captures and models the organization using a limited number of controllable parameters. We acknowledge in this paper that the signals discussed are quasi-periodic and locally stationary. Quasi-periodicity means that the waveform shape in a periodic signal varies slowly among at least three adjacent periods, even though the period length may vary. Local stationarity means that harmonic parameters, such as sinusoidal magnitude and frequency, vary slowly over time, allowing them to be considered approximately constant within a short, windowed region of the observed signal.

In this paper, we focus on phase representation, estimation, and parametric modeling. We first address the concept of structure based on the three aspects that define a stationary periodic signal: (i) frequency, (ii) magnitude, and (iii) phase.

Let us consider a real-valued signal $x(t)$ that consists of L sinusoids, as follows:

$$x(t) = \sum_{\ell=0}^{L-1} A_{\ell} \sin(\Omega_{\ell}t + \phi_{\ell}) + s(t), \quad (1)$$

where A_{ℓ} and Ω_{ℓ} represent, respectively, the magnitude and frequency of the ℓ^{th} sinusoid, ϕ_{ℓ} represents the starting phase of the ℓ^{th} sinusoid, and $s(t)$ represents a random signal without any relevant phase structure. It is well known from the basic Fourier theory that the sinusoids in a periodic signal are harmonically related [14]. Hence, in our context, we have the following:

$$\Omega_{\ell} = (\ell + 1)\Omega_0, \quad (2)$$

i.e., the frequency of each sinusoid is a multiple integer of a fundamental frequency represented by Ω_0 . Thus, this frequency organization represents a frequency-related feature that is intrinsic and, therefore, structural, to any periodic signal. This means that the only information in the frequency structure that is truly unique is the fundamental frequency.

A second structural aspect that defines a periodic signal consists of the organization of the magnitudes of the different harmonics and can be given by the ratio between the magnitude of each harmonic (A_{ℓ}) and that of the fundamental frequency magnitude (A_0). If the latter is looked at as a gain, then a given periodic signal is characterized by a normalized magnitude-related feature vector, where the first value is one, and all other values (A_{ℓ}/A_0) help to define the waveform shape of the periodic signal. This magnitude-related feature vector expresses the magnitude structure of a periodic signal, independently of time, and the fundamental frequency of the signal, provided that the waveform shape is locally preserved.

A third aspect that contributes to defining the particular waveform shape of a periodic signal involves the relationships between the starting phases of the different harmonics, ϕ_{ℓ} . Ideally, it would be interesting to characterize the waveform shape of a given periodic signal based on a normalized phase-related feature vector that, similar to the magnitude-

related feature vector, is independent of time and the fundamental frequency. This paper aims to show that such a time-shift invariant phase-related feature exists and that it can be extracted using fairly conventional spectrum analysis. Moreover, it can be modeled in simple and insightful ways.

To this end, two problems need to be addressed. The first one involves estimating the harmonic starting phase values, ϕ_ℓ , from a short segment of $x(t)$ that is representative of the periodic signal. By ‘representative’, we mean that the short segment contains a few periods with a similar waveform shape, but its duration is not related to the period of the periodic signal. The second aspect involves establishing a model based on the estimated harmonic starting phases, which, when combined with harmonic magnitude information, helps to fully explain a given waveform shape in a way that is time-shift invariant and independent of Ω_0 . The next section further discusses these challenges from a practical perspective and motivates a paradigm in harmonic signal processing oriented toward the harmonic magnitude and phase structure, greatly facilitating signal modeling, transformation, and synthesis.

1.3. A Practical Approach to Harmonic Signal Processing

The problem we address falls within the realm of spectrum estimation using Fourier analysis. Given our aim for practical signal processing that is suitable for real-time operation on low-cost platforms, we highlight simple technical approaches that utilize the discrete Fourier transform (DFT). This choice allows for the benefit of efficient realization algorithms (e.g., the FFT [15]). In addition, we exclude non-causal or iterative processing in order to accommodate real-time operation on low-cost platforms.

Thus, in a practical setting, simple spectrum estimation typically implies three operations: (1) sampling, (2) time–frequency (T-F) transformation, and (3) frequency, magnitude, as well as phase estimation:

1. A discrete-time version of the signal represented by Equation (1) is first obtained using a convenient sampling frequency (F_S), i.e., we have the following:

$$x[n] = x(t)|_{t=nT_S = \frac{n}{F_S}}, \tag{3}$$

where T_S represents the sampling period and corresponds to the reciprocal of the sampling frequency (F_S).

2. A T-F transformation (using e.g., the DFT) is computed on a windowed region of the discrete-time signal containing N samples. We assume that N is a power of two numbers. If the window is represented by $w[n]$ (and we assume is symmetric), this means that, in the case of the DFT, we compute the following:

$$X[k] = \sum_{n=0}^{N-1} x[n]w[n]e^{-j\frac{2\pi}{N}kn}, \quad k = 0, 1, \dots, N - 1, \tag{4}$$

where $j = \sqrt{-1}$.

3. Finally, a suitable estimation procedure is used that takes the spectral coefficients, $X[k]$, as input and delivers robust estimates of all the harmonic frequencies, magnitudes, and phases.

These simple spectrum estimation steps enable the identification of important harmonic parameters such as the normalized fundamental frequency, $\omega_0 = \Omega_0 T_S$. On the other hand, if the parametric representation of harmonic magnitudes and phases is done in such a way that it does not depend on time or the fundamental frequency, it paves the way for a harmonic signal processing paradigm that promotes simplification, flexibility, and even insight in algorithms.

In fact, in many applications, such as speech enhancement and time-scale or pitch-scale modification of speech [3], and special effects in singing and music [2], harmonic sinusoids

are individually modified in terms of their frequency, magnitude, and phase trajectories, as suggested by Figure 1.

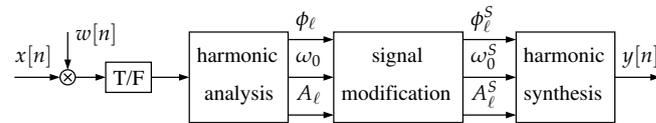


Figure 1. In many harmonic signal processing applications, individual harmonics are processed separately. Signal modification converts the original fundamental frequency (ω_0), the original harmonic magnitudes (A_ℓ), and phases (ϕ_ℓ), into new values ($\omega_0^S, A_\ell^S, \phi_\ell^S$) in the synthesis process.

However, this approach requires careful unwrapping of the phases of all harmonics such that their modifications are correct and do not suffer from the errors that may result from the wrapped phase representation in the interval $[-\pi, \pi]$. In addition, phase unwrapping, which is carried out for all harmonics on an individual basis and in a ‘horizontal’ manner, i.e., along the time axis, is itself prone to estimation errors. The most critical aspect of this approach, however, is that it is not insightful, i.e., it does not capture the overall, or holistic, harmonic phase structure, which means that, most likely, it does not explicitly control it. As a result, although phase coherence may be obtained on an individual sinusoidal basis, ‘vertical’ coherence may not be controlled and artifacts may result, the most common being known as ‘phasiness’ [16].

A more convenient paradigm in harmonic signal processing is represented by the block diagram illustrated in Figure 2.

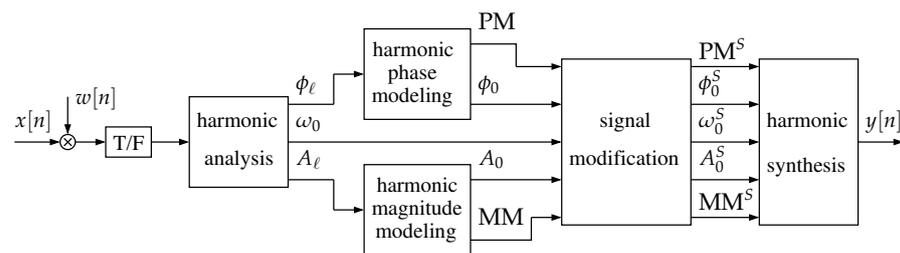


Figure 2. A convenient paradigm in harmonic signal processing: a holistic harmonic magnitude and phase structure represented by a shift-invariant and fundamental frequency-independent magnitude model (MM), and phase model (PM), respectively. Signal transformation just involves modifications to these models.

According to this paradigm, harmonic magnitude and phase models are extracted that represent the holistic harmonic structure (or ‘vertical’ structure) in a way that is time-shift invariant, and independent of the fundamental frequency. This approach not only promotes insight into the harmonic signal structure but also greatly facilitates signal transformation. For example, the magnitude model (MM) and the phase model (PM) may change arbitrarily or may be interpolated in simple ways. Or, if the waveform shape is to be preserved, they may remain unchanged, and only the changes in the synthesis affect the fundamental frequency parameters ($\omega_0^S, A_0^S, \phi_0^S$). In addition, the synthetic phase ϕ_0^S may be decoupled from the original phase ϕ_0 since it can be easily synthesized using the value of ω_0^S .

1.4. Paper Structure

This paper is divided into two major parts. The first part, corresponding to Section 2, is devoted to the estimation of the phases of individual sinusoids using light but robust DFT-based analysis procedures. Two commonly used DFT filter banks and three window functions are considered. Robust phase estimators are found, taking into consideration the specificities of the filter banks and window functions, notably their frequency responses. It has been shown that the relative performance of the six-phase estimation alternatives mainly depends on the main lobe width of the magnitude of those frequency responses and the associated near-end and far-end leakage properties. In particular, using the lower

bound for the error variance of any unbiased estimator, which is set by the Cramér–Rao lower bound (CRLB) [17], it has been demonstrated that the performance of one of the studied windows is not only closer to the CRLB when conditions are favorable but also exhibits a lower performance deviation when signal conditions are adverse.

In the second part of the paper, corresponding to Section 3, we address the computation of a time-shift invariant harmonic phase model (HPM) that implements the holistic PM represented in Figure 2 and validates the harmonic signal processing paradigm that the block diagram of Figure 2 substantiates. Using the concepts of vertical phase coherence, as well as detailed and interpretable test cases, both ground truth and estimated PM are illustrated and compared under various test conditions. This section concludes with an illustration of the Harmonic Phase Model (HPM) of actual natural voice signals and an explanation of its usefulness in demystifying the phasegram of a representative periodic signal. Finally, Section 4 concludes this paper with a summary of the most significant concepts and results presented, in addition to a perspective on further research and application scenarios.

2. Robust DFT-Based Phase Estimation of Individual Sinusoids

The purpose of this section is to estimate the starting phases (ϕ_ℓ) of all sinusoids in a harmonic signal, as in (1), after the signal has been subject to simple uniform sampling, as in (3), multiplied by a window ($w[n]$); the result is transformed into a discrete-frequency Fourier domain, such as in (4). As shown in Section 1.1 and illustrated in Figures 1 and 2, we rely on a pre-existing harmonic analysis framework that can deliver not only an accurate fundamental frequency estimate (ω_0) but also harmonic magnitude estimates (A_ℓ). In our simulations, in this paper, we use ground-truth test signals such that those parameters are known beforehand. This is equivalent to assuming that these parameters are estimated without error.

Given a certain magnitude spectrum describing a harmonic signal, it is tempting to think that the estimated harmonic starting phases can be taken as the phases of the DFT spectral lines (or DFT bins) corresponding to local maxima in the magnitude spectrum. In other words, if $X[k]$, $k = 0, 1, \dots, N - 1$ represents the Fourier spectrum of the windowed harmonic signal, and if $k_\ell = \underset{k}{\operatorname{argmax}} |X[k]|$ with $\ell = 0, 1, \dots, L - 1$ denotes a DFT bin index representing a local maximum in the magnitude spectrum, and corresponding to harmonic ℓ , it is tempting to believe that ϕ_ℓ can be estimated as $\hat{\phi}_\ell = \angle X[k_\ell]$. In fact, this equation is not correct given those estimations must take into consideration the specificity of the time–frequency transformation, the specificity of the window, and the relation between each spectral peak in the magnitude spectrum and fundamental frequency.

As an important contribution of this paper, this section highlights that phase estimation does not depend on accurate frequency estimation of individual sinusoids. In fact, as stated by Rife and Boorstyn, the most practical DFT-based accurate frequency estimators involve a two-step approach: a coarse search followed by a fine search [12]. The first step is almost trivial as it involves only peak picking and the second is what determines the accuracy of the frequency estimator [5]. Usually, sinusoidal magnitude estimation depends on the result of the fine search step in frequency estimation; consequently, the accuracy of the former depends on the accuracy of the latter. In this section, we show that accurate phase estimation does not depend on this fine search step in frequency estimation and, therefore, its accuracy depends mainly on the severity of the signal-to-noise ratio (SNR).

In order to illustrate different possibilities, we consider six cases that result from the combination of two different DFT-based filter banks and three different windows.

In order to simplify results and facilitate their comparison, in all six cases, we consider the analysis of one harmonic sinusoid in (1), i.e., after sampling—according to (3)—our signal is as follows:

$$x[n] = A_\ell \sin(\omega_\ell n + \phi_\ell) + s[n]. \quad (5)$$

In this equation, we have the following:

$$\omega_\ell = \Omega_\ell T_S = \frac{2\pi}{N}(k_\ell + \Delta_\ell), \tag{6}$$

where k_ℓ represents the bin index (or sub-band channel) corresponding to a local maximum in the magnitude spectrum, and Δ_ℓ represents fractional frequency. Depending on the type of T-F transformation and the window that is used, the continuous interval of Δ_ℓ may be either $[-0.5, 0.5[$ or $[0.0, 1.0[$. The rationale for this is explained next.

2.1. DFT and the Rectangular Window

In this subsection, the frequency domain representation of our signal (5) is obtained by computing the DFT according to (4) when the window is the rectangular window:

$$w[n] = w_R[n] = 1, n = 0, 1, \dots, N - 1. \tag{7}$$

Figure 3 represents the magnitude of the frequency response of the rectangular window, i.e., $|W_R(e^{j\omega})| = \left| \sum_{n=0}^{N-1} w_R[n] e^{-j\omega n} \right|$, in the range $-8\pi/N \leq \omega \leq 8\pi/N$, and after the gain is normalized such that the maximum gain is the unity. Two other frequency responses are also represented in Figure 3, which will be discussed in subsequent subsections. The frequency axis in Figure 3 is normalized by $2\pi/N$ such that integer numbers in this normalized axis can be read as bin indices in a DFT filter bank interpretation.

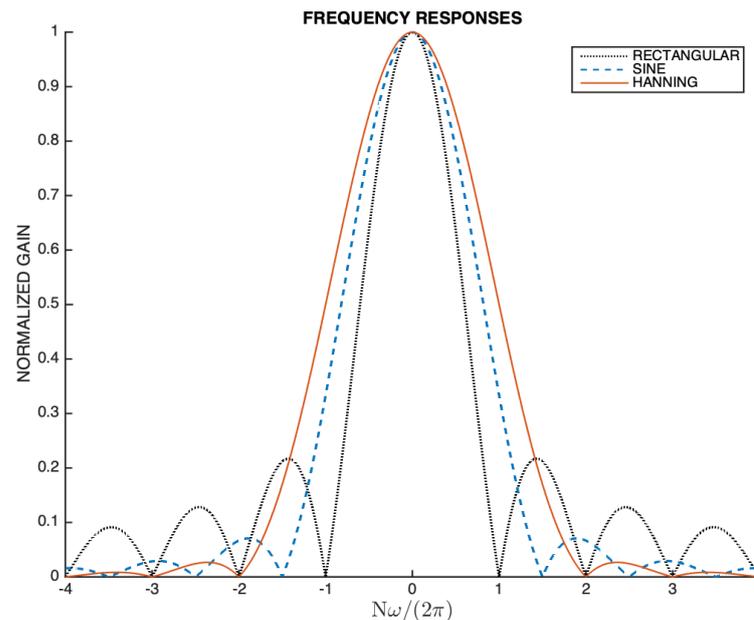


Figure 3. Normalized magnitude of the frequency response of the rectangular, sine, and Hanning windows in the range $-8\pi/N \leq \omega \leq 8\pi/N$. The frequency axis (ω) is normalized by $2\pi/N$.

Given that both $x[n]$ and $w[n]$ in (4) are real-valued, then $X[k]$ is known to be conjugate-symmetric, i.e., $X[k] = X^*[N - k]$, $k = 1, 2, \dots, N/2 - 1$, where $(\cdot)^*$ denotes complex conjugation. This means that unique information provided by $X[k]$ lies in the range $k = 0, 1, \dots, N/2$. Using Euler trigonometric relations and ignoring the Dirichlet kernel (i.e., a function with the form $\sin(\theta) / \sin(\theta/N)$) that is centered outside this range of k and that causes spectral leakage, we are left with the following:

$$X[k] \approx \frac{A}{2} e^{j(\phi_\ell - \frac{\pi}{2} + \pi(k_\ell + \Delta_\ell - k)(1 - \frac{1}{N}))} \frac{\sin \pi(k_\ell + \Delta_\ell - k)}{\sin \frac{\pi}{N}(k_\ell + \Delta_\ell - k)}. \tag{8}$$

Given that the width of the main lobe of the frequency response of the rectangular window is $4\pi/N$, as apparent in Figure 3, then, depending on the value of Δ_ℓ , at most two spectral lines with a non-zero magnitude fall within that main lobe width. If $\Delta_\ell \in [-0.5, 0.5]$, then the spectral line $k = k_\ell$ corresponds to a local maximum in the magnitude spectrum, as illustrated in Figure 4:

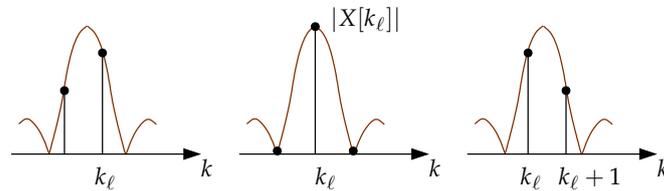


Figure 4. When $\Delta_\ell \in [-0.5, 0.5]$, the magnitude spectrum of the DFT of a rectangular-windowed sinusoid exhibits a local maximum for $k = k_\ell$. The plots illustrate the case when $\Delta_\ell = -0.5 + \epsilon$ (left), when $\Delta_\ell = 0.0$ (center), and when $\Delta_\ell = 0.5 - \epsilon$ (right), where $\epsilon < 0.5$ is a small real positive number.

This means that this spectral line is more likely to be less affected by noise and leakage, which makes it particularly suitable for phase estimation. In fact, if $k = k_\ell$, then from (8), we obtain $\angle X[k_\ell] = \phi_\ell - \frac{\pi}{2} + \pi\Delta_\ell \left(1 - \frac{1}{N}\right)$, which leads to the following:

$$\phi_\ell = \angle X[k_\ell] + \frac{\pi}{2} - \pi\Delta_\ell \left(1 - \frac{1}{N}\right). \tag{9}$$

Despite being usable, this result presents one practical difficulty since it requires that the fractional frequency (Δ_ℓ) be estimated as accurately as possible [5,18]. This means that frequency estimation errors may propagate to the phase estimation. In order to avoid this, a more robust approach is available if, instead of estimating phase with respect to the origin of the time segment, estimation is performed with respect to the group delay of the DFT filter bank. Given that the window is symmetric, the group delay is constant and given by $\tau = (N - 1)/2$. Therefore, using this result and (6) and (9), we obtain the following:

$$\begin{aligned} \hat{\phi}_\ell &= \angle X[k_\ell] + \frac{\pi}{2} - \pi\Delta_\ell \left(1 - \frac{1}{N}\right) + \omega_\ell \tau \\ &= \angle X[k_\ell] + \frac{\pi}{2} + \pi k_\ell \left(1 - \frac{1}{N}\right), \end{aligned} \tag{10}$$

which represents a more robust phase estimator that just depends on the bin index of the spectral line corresponding to the local maximum in the magnitude spectrum, and on the phase of that spectral line.

A perspective on the estimation error that is associated with (10) can be given by a simple metric that evaluates the cumulative distance, or error, between the estimated $\hat{\phi}$, and the corresponding ground-truth value, ϕ , when the latter varies in the range $[-\pi, \pi]$. For illustration purposes, we use the error metric as in (11), we create a ground-truth signal using (5), and configure the noise ($s[n]$) such that the SNR is 30 dB in one case, and 10 dB in another case. In each case, we evaluate the cumulative phase estimation error as a function of the fractional frequency, Δ_ℓ :

$$\text{ERROR} = \int_{-\pi}^{\pi} |e^{j\phi} - e^{j\hat{\phi}}| d\phi \tag{11}$$

Figure 5 represents the results when $N = 128$ and $\ell = 13$.

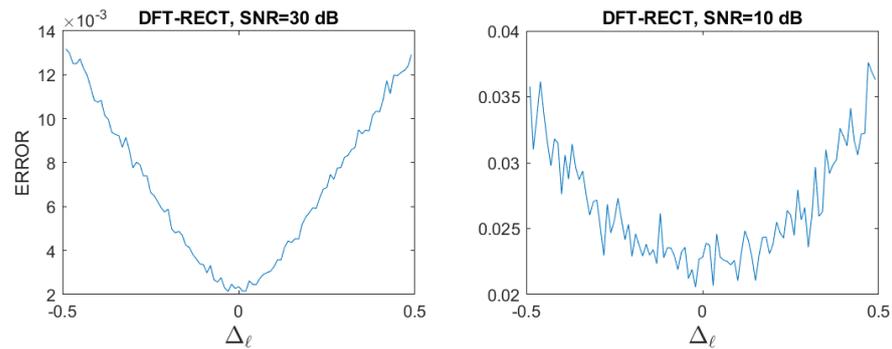


Figure 5. Cumulative phase estimation error as a function of the fractional frequency Δ_ℓ , when the DFT and the rectangular window are used. In the illustrated cases, $N = 128$, $\ell = 13$, SNR = 30 dB (left) and SNR = 10 dB (right).

It can be seen that the estimation error distribution is consistent with the relative magnitudes of adjacent spectral bins as illustrated in Figure 4. In fact, when Δ_ℓ approaches -0.5 , or 0.5 , the cumulative error increases relative to the case when $\Delta_\ell = 0$, which is a consequence of the fact that the magnitudes of two adjacent spectral lines become comparable and significantly lower than the maximum value they can reach (when $\Delta_\ell = 0$), which not only exacerbates leakage effects but also increases vulnerability to the noise influence. Due to the specific frequency response of the rectangular window, as Figures 3 and 4 highlight, then, when $\Delta_\ell = 0$ and the SNR is infinity, the spectral leakage is zero, and the phase is estimated without error. The Matlab code generating Figure 5 is available (`estimatePHASE_DFT_rect.m`), which facilitates experimentation with other values of N , ℓ , or SNR.

2.2. DFT and the Sine and Shifted Hanning Windows

Other windows that provide better main-to-side lobe attenuation compared to the rectangular window are frequently used in spectrum estimation and FIR filter design [15]. Although standard Hamming or Hanning windows could be used in our analysis, we employ two related windows that are particularly significant in perfect reconstruction filter banks [19,20], such as those that are frequently used in audio coding and general analysis synthesis [21]. One window is known as the sine window and is defined as follows:

$$w_S[n] = \sqrt{w_H[n]} = \sin \frac{\pi}{N}(n + 0.5), \quad n = 0, \dots, N - 1, \tag{12}$$

where $w_H[n]$ represents the shifted Hanning window, defined as follows:

$$w_H[n] = \frac{1}{2} \left[1 - \cos \frac{2\pi}{N}(n + 0.5) \right], \quad n = 0, \dots, N - 1. \tag{13}$$

When the sine window is used in the DFT analysis according to (4), where $x[n]$ is a noisy sinusoid given by (5), then, using an approach that is similar to that used in the previous subsection, the spectral line corresponding to the relevant local maximum in the magnitude spectrum ($|X[k_\ell]|$) can be approximated by the following:

$$X[k_\ell] \approx \frac{A}{4} e^{j(\phi_\ell - \frac{\pi}{2} + \pi\Delta_\ell(1 - \frac{1}{N}))} F(\Delta_\ell), \tag{14}$$

where we have the following:

$$F(\Delta_\ell) = \frac{\sin \pi \left(\Delta_\ell + \frac{1}{2} \right)}{\sin \frac{\pi}{N} \left(\Delta_\ell + \frac{1}{2} \right)} + \frac{\sin \pi \left(\Delta_\ell - \frac{1}{2} \right)}{\sin \frac{\pi}{N} \left(\Delta_\ell - \frac{1}{2} \right)}. \tag{15}$$

In the case of the shifted Hanning window, the local maximum in the magnitude spectrum can be approximated by the following:

$$X[k_\ell] \approx \frac{A}{4} e^{j(\phi_\ell - \frac{\pi}{2} + \pi\Delta_\ell(1 - \frac{1}{N}))} G(\Delta_\ell), \tag{16}$$

where we have the following:

$$G(\Delta_\ell) = 2 \frac{\sin \pi\Delta_\ell}{\sin \frac{\pi}{N}\Delta_\ell} + \frac{\sin \pi(\Delta_\ell + 1)}{\sin \frac{\pi}{N}(\Delta_\ell + 1)} + \frac{\sin \pi(\Delta_\ell - 1)}{\sin \frac{\pi}{N}(\Delta_\ell - 1)}. \tag{17}$$

It can be easily concluded that the $F(\Delta_\ell)$ and $G(\Delta_\ell)$ functions have the same polarity when $\Delta_\ell \in [-0.5, 0.5]$ and, thus, they do not affect the phase in this particular range of Δ_ℓ . On the other hand, comparing (14), (16), and (8), when $k = k_\ell$, it can be concluded that the phase is governed by the same function, which is a consequence of the fact that all three windows share the same linear-phase property and the same group delay. This also means that in all three cases, the phase can be estimated using Equation (10), although the quality of the phase estimation depends on each window due to its specific near-end and far-end leakage characteristics.

As apparent in Figure 3, the width of the main lobe of the frequency response is $6\pi/N$ in the case of the sine window, and $8\pi/N$ in the case of the shifted Hanning window [5,11]. This implies that, depending on the value of Δ_ℓ , at most three spectral lines with non-zero magnitude fall within the main lobe width of the sine window frequency response. In the case of the shifted Hanning window, the number of spectral lines is four. As in the previous subsection, if $\Delta_\ell \in [-0.5, 0.5]$, then the spectral line $k = k_\ell$ corresponds to a local maximum in the magnitude spectrum when the sine window or the shifted Hanning window is used. Figure 6 illustrates the case corresponding to the sine window.

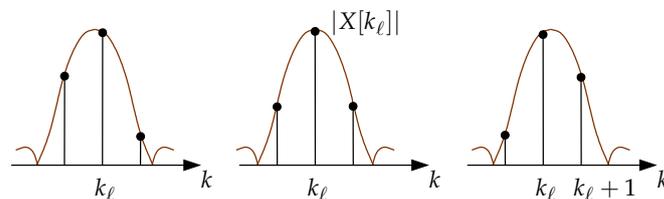


Figure 6. When $\Delta_\ell \in [-0.5, 0.5]$, the magnitude spectrum of the DFT of a sine-windowed sinusoid exhibits a local maximum for $k = k_\ell$. The plots illustrate the case when $\Delta_\ell = -0.5 + \epsilon$ (left), when $\Delta_\ell = 0.0$ (center), and when $\Delta_\ell = 0.5 - \epsilon$ (right), where $\epsilon < 0.5$ is a small real positive number.

Although this figure resembles Figure 4, two important aspects are worth noting. First, the side lobe attenuation decay (not shown in these figures) is stronger for the sine window, when compared to the rectangular window [5]. Second, because the main lobe of the frequency response of the sine window is wider than the main lobe of the rectangular window, this means that when Δ_ℓ is such that the magnitudes of the spectral lines $k = k_\ell - 1$ and $k = k_\ell$ become comparable, or the magnitudes of the spectral lines $k = k_\ell$ and $k = k_\ell + 1$ become comparable, these magnitudes are closer to the maximum value they can reach (which takes place for $\Delta_\ell = 0.0$) than what happens for the rectangular window. These concurrent reasons make it that relative to the case of the rectangular window, phase estimation using the sine window is likely to be more immune to the noise influence and to suffer fewer leakage effects. This can be confirmed by making a simple study on the phase estimation error, as it was described in the previous subsection for the rectangular window. Using the same simulation conditions, and the same phase estimation function (Equation (10)), we obtain the cumulative phase estimation error results that are illustrated in Figure 7.

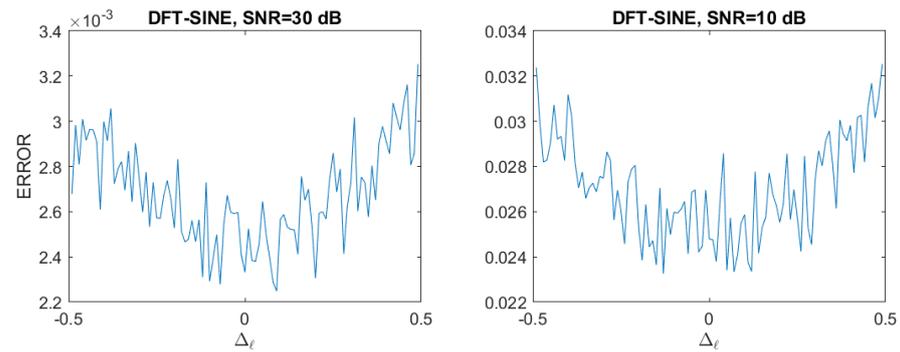


Figure 7. Cumulative phase estimation error as a function of the fractional frequency Δ_ℓ , when the DFT and the sine window are used. In the illustrated cases, $N = 128$, $\ell = 13$, SNR = 30 dB (left) and SNR = 10 dB (right).

When compared to the results in Figure 5, it can be concluded that phase estimation appears to be more accurate when the sine window is used, especially when the SNR is high, which is a natural consequence of the smaller leakage caused by this window. Results are similar if the shifted Hanning window is considered instead. As we shall see in Section 2.5, more informative conclusions will emerge from a study of the phase estimation error variance.

2.3. ODFT and the Rectangular Window

An N -point DFT corresponds to a filter bank whose N sub-bands are ‘evenly stacked’, i.e., their center frequencies correspond to $\frac{\pi}{N}(2k)$, $k = 0, 1, \dots, N - 1$. An alternative option involves ‘oddly stacked’ sub-bands, i.e., their center frequencies correspond to $\frac{\pi}{N}(2k + 1)$, $k = 0, 1, \dots, N - 1$. In this case, the time–frequency transformation results as

$$X[k] = \sum_{n=0}^{N-1} x[n]w[n]e^{-j\frac{2\pi}{N}n(k+\frac{1}{2})}, \quad k = 0, 1, \dots, N - 1, \tag{18}$$

and is known as Odd-frequency DFT, or just Odd-DFT [22]. We also abbreviate this designation to ODFT. Although it is not too much different from the basic DFT, it has several advantages that make it preferable in certain contexts. For example, assuming that the window is real-valued and symmetric, and if $x[n]$ is also real-valued, then the conjugate-symmetric property is expressed as $X[k] = X^*[N - 1 - k]$, $k = 0, 1, \dots, N/2 - 1$, i.e., only $N/2$ values of $X[k]$ are unique, instead of $N/2 + 1$, as in the case of the DFT. This subtle difference facilitates, for example, signal modification by avoiding the special cases of the DFT corresponding to $k = 0$ and $k = N/2$. In particular, the ODFT facilitates signal integration in the frequency domain because none of the sampled frequencies is zero, which avoids the singularity that occurs in the case of the DFT when $k = 0$ [23].

When the sinusoidal signal (5) is multiplied by the rectangular window (7) and the result is ODFT-transformed, then the relevant maximum that exists in the magnitude spectrum and that occurs for $k = k_\ell$, is given by the following:

$$X[k_\ell] \approx \frac{A}{2} e^{j(\phi_\ell - \frac{\pi}{2} + \pi(\Delta_\ell - \frac{1}{2})(1 - \frac{1}{N}))} \frac{\sin \pi \left(\Delta_\ell - \frac{1}{2} \right)}{\sin \frac{\pi}{N} \left(\Delta_\ell - \frac{1}{2} \right)}. \tag{19}$$

However, in this case, the fact that the sampled frequencies of the ODFT rotate by π/N relative to those of the DFT, makes that the phase can be estimated from the same local maximum as long as $\Delta_\ell \in [0.0, 1.0]$. This means that the illustration of the spectral magnitudes surrounding the local maximum in Figure 4 is also valid in this context, ex-

cept that it applies to this new Δ_ℓ range. Under this assumption, from (19) we obtain $\angle X[k_\ell] = \phi_\ell - \frac{\pi}{2} + \pi\left(\Delta_\ell - \frac{1}{2}\right)\left(1 - \frac{1}{N}\right)$, which leads to the following:

$$\phi_\ell = \angle X[k_\ell] + \frac{\pi}{2} - \pi\left(\Delta_\ell - \frac{1}{2}\right)\left(1 - \frac{1}{N}\right). \tag{20}$$

As in the previous two subsections, we can avoid the dependence on the prior estimation of Δ_ℓ by considering a time reference corresponding to the group delay of the window, which implies adding $\omega_\ell \tau$ to (20), as follows:

$$\begin{aligned} \hat{\phi}_\ell &= \angle X[k_\ell] + \frac{\pi}{2} - \pi\left(\Delta_\ell - \frac{1}{2}\right)\left(1 - \frac{1}{N}\right) + \omega_\ell \tau \\ &= \angle X[k_\ell] + \pi\left(1 - \frac{1}{2N}\right) + \pi k_\ell \left(1 - \frac{1}{N}\right). \end{aligned} \tag{21}$$

Although this result is different from the one discussed in the previous two subsections, it reveals the same important property that phase estimation depends on the spectral bin index of the local maximum in the magnitude spectrum, instead of the fractional frequency Δ_ℓ , which adds robustness to the phase estimation process. Figure 8 shows the cumulative phase estimation error that results from (21) when the test conditions are the same as those that are considered in Sections 2.1 and 2.2.

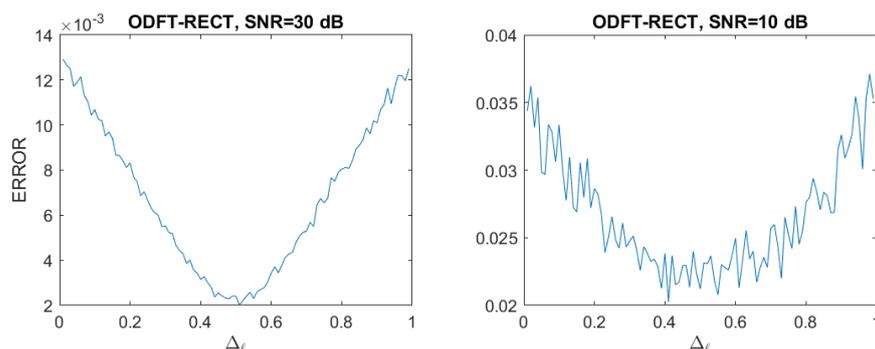


Figure 8. Cumulative phase estimation error as a function of the fractional frequency Δ_ℓ , when the ODFT and the rectangular window are used. In the illustrated cases, $N = 128$, $\ell = 13$, SNR = 30 dB (left) and SNR = 10 dB (right).

When the results in Figure 8 are compared to the results in Figure 5, it can be seen that no major differences exist, as expected (essentially because the same window is used), although the results in Figure 8 are slightly better. This is explained by the fact that the ODFT implicitly performs a small frequency modulation (i.e., an upshift in frequency by π/N) of the input signal, which improves the frequency separation between the two Dirichlet kernels, and this slightly reduces the mutual interference due to leakage.

2.4. ODFT and the Sine and Shifted Hanning Windows

When we compute the N-point ODFT (Equation (18)) of the sinusoid (5) and take for $w[n]$ the sine window (Equation (12)), a relevant local maximum exists in the magnitude spectrum $|X[k_\ell]|$ at $k = k_\ell$ whose spectral coefficient can be approximated by the following:

$$X[k_\ell] \approx \frac{A}{4} e^{j\left(\phi_\ell - \frac{\pi}{2} + \pi\left(\Delta_\ell - \frac{1}{2}\right)\left(1 - \frac{1}{N}\right)\right)} R(\Delta_\ell), \tag{22}$$

where we have the following:

$$R(\Delta_\ell) = \frac{\sin \pi \Delta_\ell}{\sin \frac{\pi}{N} \Delta_\ell} + \frac{\sin \pi (\Delta_\ell - 1)}{\sin \frac{\pi}{N} (\Delta_\ell - 1)}. \tag{23}$$

In the case of the shifted Hanning window, the local maximum in the magnitude spectrum can be approximated by the following:

$$X[k_\ell] \approx \frac{A}{4} e^{j(\phi_\ell - \frac{\pi}{2} + \pi(\Delta_\ell - \frac{1}{2})(1 - \frac{1}{N}))} S(\Delta_\ell), \tag{24}$$

where we have the following:

$$S(\Delta_\ell) = 2 \frac{\sin \pi \left(\Delta_\ell - \frac{1}{2} \right)}{\sin \frac{\pi}{N} \left(\Delta_\ell - \frac{1}{2} \right)} + \frac{\sin \pi \left(\Delta_\ell + \frac{1}{2} \right)}{\sin \frac{\pi}{N} \left(\Delta_\ell + \frac{1}{2} \right)} + \frac{\sin \pi \left(\Delta_\ell - \frac{3}{2} \right)}{\sin \frac{\pi}{N} \left(\Delta_\ell - \frac{3}{2} \right)}. \tag{25}$$

Given that both (23) and (25) are always positive when $\Delta_\ell \in [0.0, 1.0[$, they do not affect phase in this range of Δ_ℓ .

Equations (22) and (24) show that the phase of the ODFT spectral line corresponding to the local maximum in the magnitude spectrum is the same as that already found for the ODFT and rectangular window combination. As noted previously, this is expected given that all three windows share the same linear-phase property and the same group delay. As a consequence, the phase can be estimated using Equation (21) although it remains true that the quality of the phase estimation depends on the specific near-end and far-end leakage characteristics of each window.

Taking the sine window as an example, Figure 9 shows the cumulative phase estimation error arising from phase estimation according to (21) and taking into consideration the processing conditions in this section.

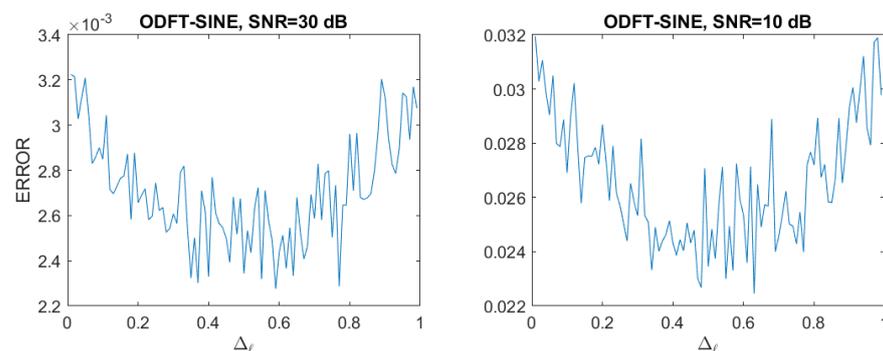


Figure 9. Cumulative phase estimation error as a function of the fractional frequency Δ_ℓ , when the ODFT and the sine window are used. In the illustrated cases, $N = 128$, $\ell = 13$, SNR = 30 dB (left) and SNR = 10 dB (right).

When the results in Figure 9 are compared to the results in Figure 7, our results here are slightly better despite the fact that both relate to the same window. As pointed out in the previous section, this is due to the fact that the ODFT promotes a wider separation between Dirichlet kernels, which means that leakage effects become weaker. Results for the shifted Hanning window do not differ significantly.

Matlab code is available (`estimatePHASE_ODFT_sine.m`), which generates results and displays Figure 9.

2.5. Bias and Variance of the Phase Estimation Error

We complete this section (Section 2) with an assessment of the bias, and of the variance of the phase estimation error that results from the estimators that are explained in Section 2.1 (Equation (10)), and in Section 2.4 (Equation (21)). Bias is computed using $\text{mean}\{\phi - \hat{\phi}\}$. Variance is computed using $\text{var}\{\phi - \hat{\phi}\}$, and we take as a reference the Cramér–Rao lower bound (CRLB) for the error variance of an unbiased phase estimator [17] (page 33).

Assuming that both magnitude and frequency of the sinusoid are known—which is true in our simulations, this CRLB is given by the following:

$$\text{var}\{\phi - \hat{\phi}\} = \frac{2\sigma^2}{NA^2}, \tag{26}$$

where σ^2 is the variance of real-valued white Gaussian noise.

As the results in Figures 5 and 9 make clear, the phase estimation error is influenced by—although it does not directly depend on—the fractional frequency Δ_ℓ , which we represent in this context simply as Δ given that we are considering just one sinusoid. Hence, we evaluate the bias and variance of the phase estimation error in two almost extreme situations: when $\Delta = -0.49$ or when $\Delta = 0.00$ in the case of the estimator given by Equation (10), and when $\Delta = 0.01$ or when $\Delta = 0.5$ in the case of the estimator defined by Equation (21). In our simulations, $N = 128$ and $\ell = 13$. Results are obtained when the SNR varies from 0 dB to 30 dB in steps of 2.5 dB. For each particular SNR value, the bias and variance of the phase estimation error are computed when ϕ varies between $-\pi$ and $+\pi$ in 200 steps and, at each step, stable statistics are reached after 100 Monte Carlo runs accounting for noise contamination.

Figure 10 displays representative results regarding bias. Results reveal that in all six cases, bias reduces as the SNR increases, as expected. However, there is not a specific combination of DFT-based filter bank and window that stands out. This remains true even after multiple runs of the simulations, although a relative degradation can be observed that persists for the tested rectangular window context (i.e., DFT filter bank and rectangular window) under more adverse Δ conditions (i.e., when Δ approaches -0.5 to 0.5), which is easily explained by the poor leakage characteristics of the rectangular window.

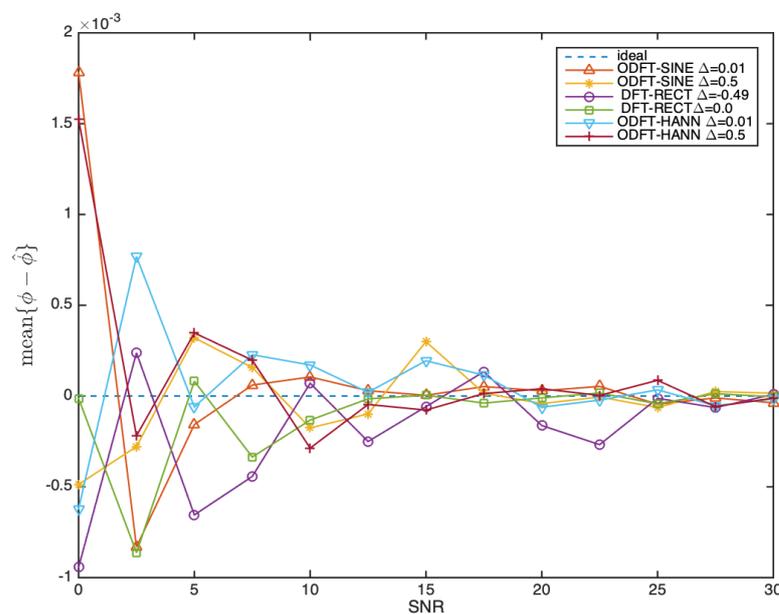


Figure 10. Mean over 100 Monte Carlo runs of the phase estimation error when $N = 128$, $\ell = 13$, and when Δ takes on two extreme values depending on the estimator.

In general, it can be concluded that bias is fairly low, in the order or less than 0.001 radians for SNR equal to or larger than 2.5 dB. This represents less than 0.016% of the 2π dynamic range.

The most significant results regard estimation error variance and are shown in Figure 11.

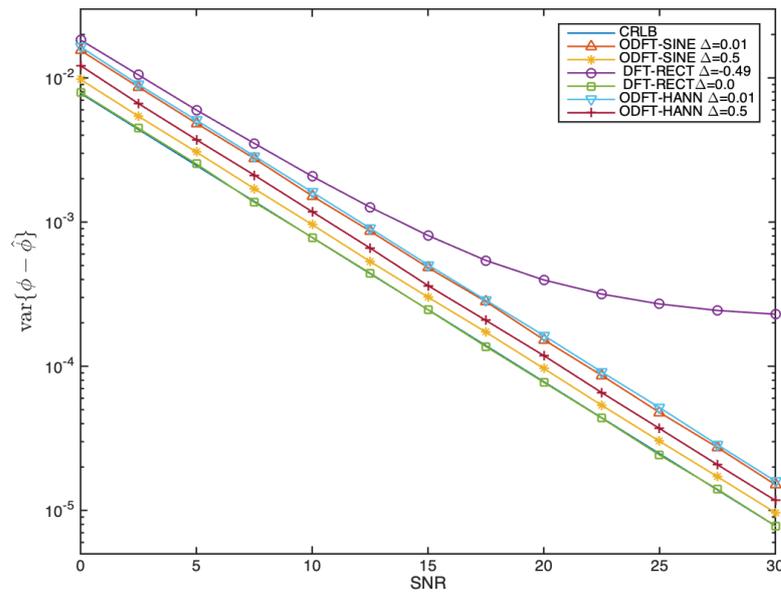


Figure 11. Variance of the phase estimation error when $N = 128$, $\ell = 13$, and when Δ takes on two extreme values depending on the estimator. The Cramér–Rao lower bound is also represented although it is not too visible since it is overlapped by the DFT-RECT results when $\Delta = 0.0$.

It is an interesting and somewhat unexpected outcome that the rectangular window gives rise to the best results when the tested Δ conditions are more favorable (i.e., when $\Delta = 0.0$), and to the worst results when the tested Δ conditions are more adverse (i.e., when $\Delta = -0.49$).

In the former case, the performance reaches the CRLB because in that ideal case, there is no leakage, as already noted at the end of Section 2.1, which means that the error variance is entirely due to noise contamination. It should be noted that in practice, this rarely happens with real-world, natural, signals as it is quite unlikely that the analyzed frequencies are exactly aligned with the center frequencies of the sub-bands of the DFT-based filter bank.

In the latter case, the performance shown by the same estimator is quite poor, which reveals that, in that case, leakage effects due to the rectangular window are quite strong, as the results in Figures 5 and 8 easily anticipate. In particular, the performance becomes asymptotic when the SNR exceeds 10 dB, which is commensurate with the known main-to-side lobe attenuation of the rectangular window, in the order of 13 dB [15].

The performance of the sine and shifted Hanning windows fall in between the two extreme cases due to the rectangular window. In particular, for the same DFT-based filter bank and Δ conditions, the error variance performance of the sine window clearly exceeds that of the shifted Hanning window in the sense that a closer approximation to the CRLB is reached. This is more evident under the more favorable Δ test conditions (e.g., when $\Delta = 0.5$) than under the more adverse Δ test conditions (e.g., when $\Delta = 0.01$).

A possible explanation may be linked to the relationship between the main lobe width of the magnitude of the frequency responses of those two windows, the relative prominence of the spectral coefficients inside that main lobe, the association with the discrete frequencies defining the different DFT channels (or sub-bands), and the near-end and far-end leakage characteristics of each window.

The most impactful implication of these results is that under more general test conditions, the phase estimation error performance of the sine window is not only closer to the CRLB but also offers a lower deviation when signal conditions are more adverse. For these reasons, it can be considered that the performance of the sine window is better behaved and, thus, it will be used in the remainder of this paper.

As a final note, it should be highlighted that changing ℓ in the simulations produces a marginal effect on the results. In particular, when $\ell = N/4 - 1$, which corresponds to the ‘sweet spot’ given that leakage effects are minimized, results do not differ appreciably from those in Figure 11. The Matlab file that generates and displays the results in Figures 10 and 11 is available (CRLBphiivar.m), such that other combinations of parameters can be tried.

3. An Interpretable Time-Shift Invariant Harmonic Phase Model

In this section, we take advantage of phase estimation of individual sinusoids using the signal representation in a discrete-time and discrete-frequency Fourier domain, as it was discussed in the previous section, especially Section 2.4. The objective is to identify a vertical harmonic phase model that is time-shift invariant, and that can be approximated by a simple parametric model.

In our development in this section, we use the deterministic part of (1) taking into consideration the harmonic relationship given by (2). Thus, in our analysis, we use the following:

$$\begin{aligned}
 x(t) &= \sum_{\ell=0}^{L-1} A_{\ell} \sin((\ell + 1)\Omega_0 t + \phi_{\ell}) \\
 &= \sum_{\ell=0}^{L-1} A_{\ell} \sin((\ell + 1)\Omega_0(t + t_{\ell})) \\
 &= A_0 \sin(\Omega_0(t + t_0)) + v(t),
 \end{aligned}
 \tag{27}$$

where t_{ℓ} represents the starting delay of harmonic ℓ , and $v(t)$ represents all harmonics above the fundamental frequency. If $\Omega_0 = 2\pi/T_0$, where T_0 is the fundamental period, i.e., the reciprocal of the fundamental frequency (F_0), a convenient manipulation of $v(t)$ leads to the following:

$$\begin{aligned}
 v(t) &= \sum_{\ell=1}^{L-1} A_{\ell} \sin((\ell + 1)\Omega_0(t + t_{\ell})) \\
 &= \sum_{\ell=1}^{L-1} A_{\ell} \sin((\ell + 1)\Omega_0(t + t_0 - t_0 + t_{\ell})) \\
 &= \sum_{\ell=1}^{L-1} A_{\ell} \sin\left((\ell + 1)\Omega_0(t + t_0) + 2\pi \frac{t_{\ell} - t_0}{T_0/(\ell + 1)}\right) \\
 &= \sum_{\ell=1}^{L-1} A_{\ell} \sin((\ell + 1)\Omega_0(t + t_0) + 2\pi \text{NRD}_{\ell}),
 \end{aligned}
 \tag{28}$$

where NRD_{ℓ} denotes the normalized relative delay of harmonic ℓ , and is defined as follows:

$$\text{NRD}_{\ell} = \frac{t_{\ell} - t_0}{T_0/(\ell + 1)} = \frac{\phi_{\ell} - (\ell + 1)\phi_0}{2\pi}.
 \tag{29}$$

NRD_{ℓ} expresses the difference between the starting delay of harmonic ℓ , and the starting delay (i.e., the onset) of the fundamental frequency, which is further normalized by the period of harmonic ℓ [24]. Therefore, in its most intuitive interpretation, NRD can be taken modulo 1, which means that $\text{NRD}_{\ell} \in [0.0, 1.0]$. Equation (29) allows two other important conclusions. First, by definition, $\text{NRD}_0 = 0$. This means that, by definition, the NRD phase-related feature is intrinsically time-shift invariant. Second, as the second part of (29) highlights, the NRD does not depend on the fundamental frequency. Therefore, we may rewrite (27) as follows:

$$x(t) = \sum_{\ell=0}^{L-1} A_{\ell} \sin((\ell + 1)\Omega_0(t + t_0) + 2\pi \text{NRD}_{\ell}),
 \tag{30}$$

which highlights the fact that, in terms of phase, all harmonics can be expressed as a part that depends on the time-varying phase of the fundamental frequency, and on its starting phase, and another part, the NRD, which is time-shift invariant. Thus, the NRD acts as a holistic harmonic phase model that is identified in Figure 2 as PM.

Other harmonic phase descriptors that are similar to NRD were proposed by Stylianou in 1996 (phase envelope [25] (p. 44)), Di Federico in 1998 [26], and Saratxaga in 2009 (relative phase Shift-RPS [27]).

An inspiring metaphor from nature, elucidating the meaning of NRD, is depicted in Figure 12.



Figure 12. An inspiring metaphor from nature: NRD can be regarded as being similar to the space-invariant bird formation structure in a flock that is represented by the line connecting the birds in the upper branch of the illustrated flock.

Each bird in the flock formation is like a harmonic in a harmonic structure, and the dynamics of the wings of the former are like the dynamics of the phase rotation of the latter. The NRD feature represents a shift-invariant harmonic phase structure, similar to the space-invariant bird formation structure of the flock that is represented by the line connecting the birds in the upper branch of the flock. Ultimately, the whole flock's flying dynamics are governed by the position and wing evolution of the leading bird in the flock, since the whole harmonic dynamics are governed by the phase evolution of the fundamental frequency in the harmonic structure.

On the other hand, since the NRD inherits the properties of phase, then phase wrapping and phase unwrapping also apply to NRD feature vectors. It should be noted that, in this case, phase (un)wrapping is vertical, meaning that it is performed along the frequency axis, as opposed to the time axis, which is by far the most common use of phase (un)wrapping operations.

As a summary, the NRD captures the vertical phase structure (i.e., along the frequency axis) of all the harmonic sinusoids, which just depends on the waveform shape of the periodic signal they define. It should be stressed that when the (vertical) harmonic magnitude and (vertical) harmonic phase structure are combined, they uniquely define the waveform shape of a given periodic signal, independently of its time shift, and its fundamental frequency.

In the next three subsections, we take inspiration from [23] in order to demonstrate the practical estimation of three different NRD feature vectors that characterize three different waveforms, all of them emerging from the sawtooth wave.

3.1. NRD Estimation Example Based on the Sawtooth Wave

In this subsection, and throughout the rest of this paper, we use a convenient ground-truth signal that consists of the sawtooth waveform. In order to work with realistic NRD

estimation conditions, the fundamental frequency (F_0) is time-varying, and two levels of noise contamination are considered.

Regarding the time-varying fundamental frequency, we take the average pitch frequency between female and male human speakers, which is around 150 Hz, and subject it to frequency-modulation (FM) as specified in Equation (31).

$$F_0(t) = 150.0 + FM \cos(2\pi \cdot FM \cdot t) \tag{31}$$

In our simulations, we consider two FM values: $FM = 2.5$ Hz, and $FM = 0.25$ Hz. These cases of maximum frequency deviation around the mean are illustrated in Figure 13.

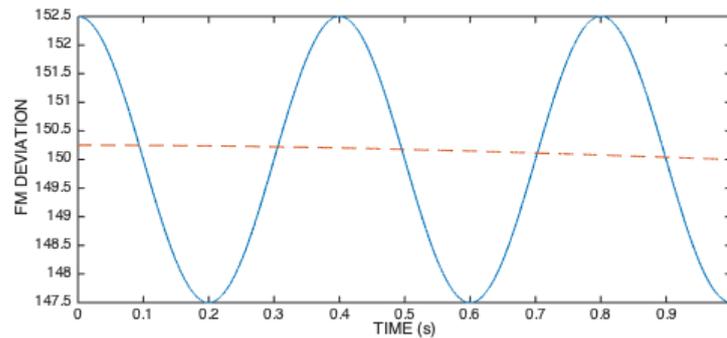


Figure 13. Two FM deviation cases characterizing test signals: $FM = 2.5$ Hz (solid line) and $FM = 0.25$ Hz (dashed line).

The sawtooth waveform is illustrated in Figure 14 for two noise contamination scenarios that we consider in our simulations: when it is mild ($SNR = 30$ dB), and when it is strong ($SNR = 10$ dB).

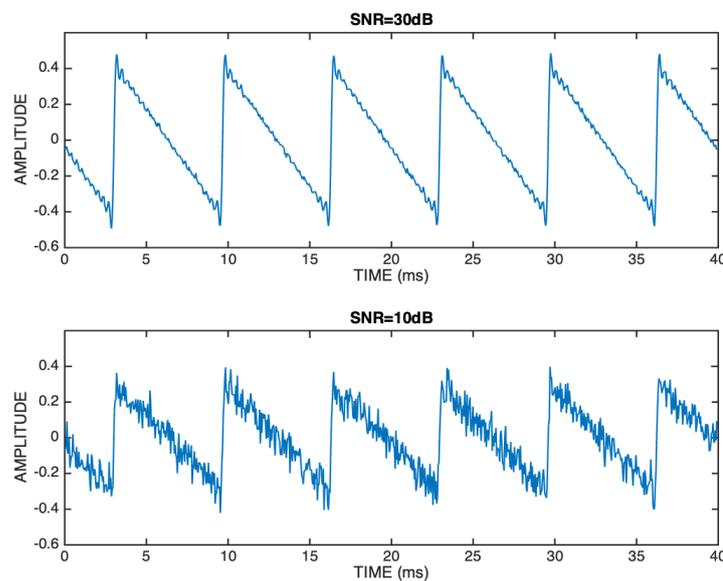


Figure 14. Illustration of the influence of noise on the sawtooth test signal when $SNR = 30$ dB (top), and when $SNR = 10$ dB (bottom).

In our simulations, the sampling frequency is 22,050 Hz, the duration of the test signals is 1 s, and the number of harmonics is $L = 20$. The time–frequency transformation is performed as discussed in Section 2.4, with $N = 1024$, and 50% overlap between adjacent frames.

Using standard Fourier analysis [14], it is easy to show that the starting phases of the sawtooth waveform in (1) are all $\phi_\ell = 0$, $\ell = 0, 1, \dots, L - 1$, which means that the harmonic

phase structure is simply given by $\text{NRD}_\ell = 0, \ell = 0, 1, \dots, L - 1$. On the other hand, the normalized harmonic magnitude ratios that express the harmonic magnitude structure of the sawtooth waveform are simply given by the following:

$$A_\ell / A_0 = \frac{1}{\ell + 1}, \ell = 0, 1, \dots, L - 1. \tag{32}$$

Given that we have access to the ground-truth value of the instantaneous fundamental frequency, as specified by (31), and by adopting the frequency front-end that is assumed in Section 2.4 (i.e., ODFT and sine window), we may easily obtain the ground-truth spectral indices (k_ℓ) of all 20 local maxima in the ODFT magnitude spectrum. These indices allow us to estimate the individual harmonic phases as specified by (21) and, from these, we estimate the NRD coefficients for all harmonics. By vertically unwrapping these coefficients, we obtain interpretable representations that are amenable to simple parametric modeling.

Figure 15 represents an overlay of the magnitude spectra and unwrapped NRD vectors of our test signal when the maximum frequency deviation around the mean is 0.25 Hz, and the SNR is 30 dB.

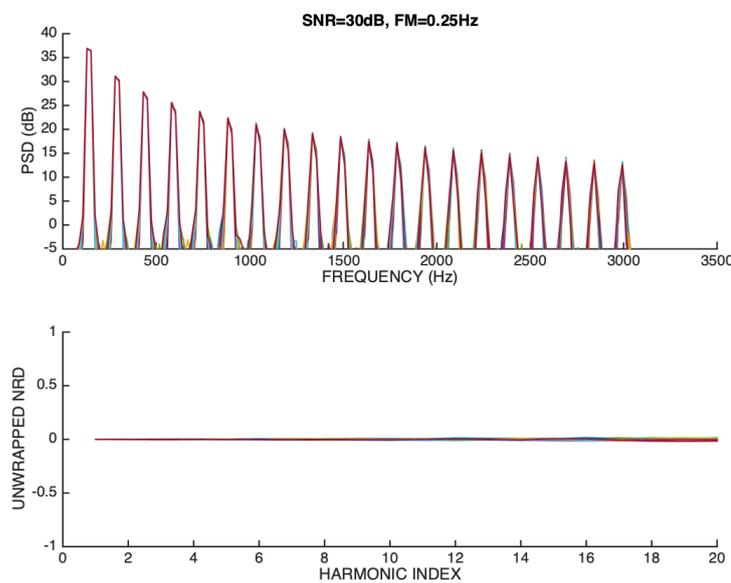


Figure 15. Overlay representation of magnitude spectra (top), and unwrapped NRD vectors (bottom) of a sawtooth signal that is FM modulated (0.25 Hz deviation around the mean) and whose SNR is 30 dB.

It can be seen that, as expected, the magnitude spectra are aligned in a very consistent way because the FM modulation is small, and the SNR is high. On the other hand, it can also be seen that the unwrapped NRD vectors are also very closely aligned to the ideal ground-truth model ($\text{NRD}_\ell = 0, \ell + 1 = 1, 2, \dots, 20$). In Figure 15 and subsequent figures the horizontal axis labels $\ell + 1 = 1, 2, \dots, 20$ as the harmonic index.

Figure 16 repeats the overlay representation when the SNR is 10 dB. It can be seen that higher-order harmonics are strongly contaminated by noise, especially when their magnitude approaches the noise floor, which adversely impacts the accuracy of phase estimation.

This is clear in Figure 16 because the unwrapped NRD vectors deviate more from the ideal ground-truth model as the harmonic order increases. Still, their average follows the correct model.

Figure 17 represents the results when $\text{SNR} = 30 \text{ dB}$ and $\text{FM} = 2.5 \text{ Hz}$.

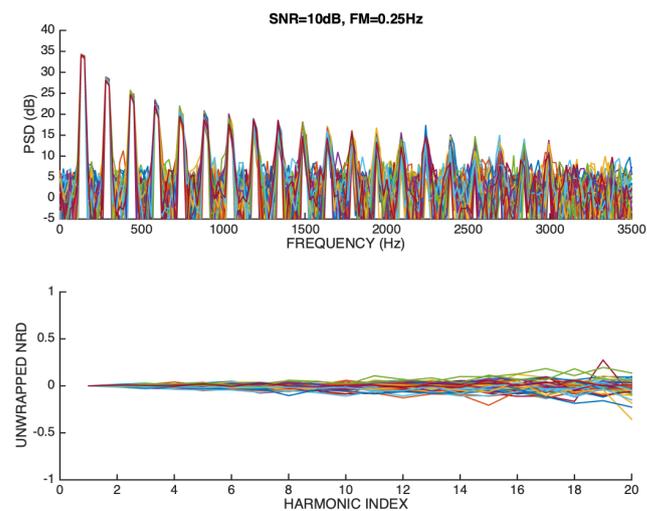


Figure 16. Overlay representation of magnitude spectra (**top**), and unwrapped NRD vectors (**bottom**) of a sawtooth signal that is FM modulated (0.25 Hz deviation around the mean) and whose SNR is 10 dB.

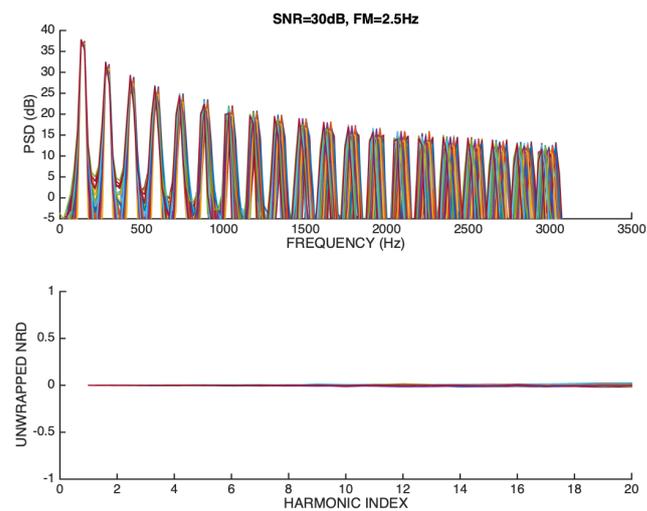


Figure 17. Overlay representation of magnitude spectra (**top**), and unwrapped NRD vectors (**bottom**) of a sawtooth signal that is FM modulated (2.5 Hz deviation around the mean) and whose SNR is 30 dB.

It can be seen that, in this case, the blurred aspect of higher-order harmonics in the magnitude spectrum is a natural consequence of the fact that the frequency deviation is proportional to the harmonic order. In particular, for $\ell = 19$, the maximum frequency deviation around the mean is 50 Hz, which represents $1/3$ of the average fundamental frequency. It can be confirmed that despite the frequency modulation, the dispersion on the unwrapped NRD vectors is quite small and quite comparable with the one that is observed in Figure 15, in which the frequency modulation is negligible. This relevant experimental outcome unequivocally confirms that the NRD phase-related feature is intrinsically independent of the fundamental frequency. In other words, the stability of the estimated unwrapped NRD vectors reflects mainly the fact that the waveform shape of the periodic signal that is observed in each frame of the signal does not change appreciably.

When $FM = 2.5$ Hz and $SNR = 10$ dB, new results are obtained that are represented in Figure 18.

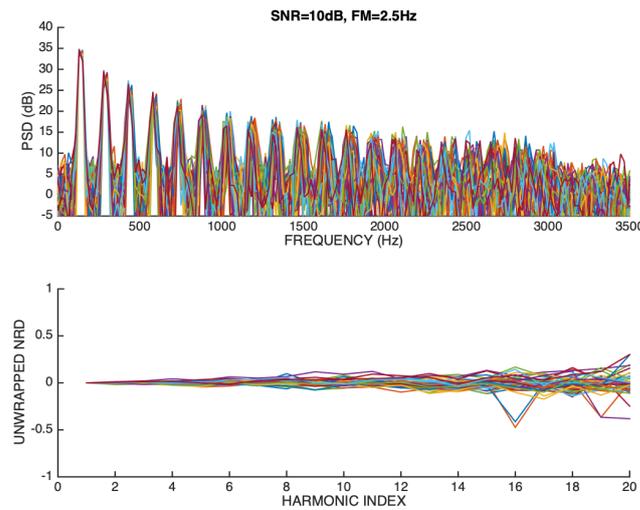


Figure 18. Overlay representation of magnitude spectra (**top**), and unwrapped NRD vectors (**bottom**) of a sawtooth signal that is FM modulated (2.5 Hz deviation around the mean) and whose SNR is 10 dB.

It is easy to anticipate from the spectral magnitude representation that the strong noise influence, when it is allied to the significant frequency deviation of high-order harmonics, paves the way for serious difficulties in phase estimation, which is reflected in the substantial deviation of the unwrapped NRD vectors relative to the ideal ground-truth model. In particular, phase estimation errors cause the phase unwrapping algorithm to introduce sudden jumps of 0.5 in certain unwrapped NRD vectors. However, mitigating strategies could be easily adopted that could detect and avoid such jumps.

The Matlab code that generates signals and creates the overlay representations, as illustrated in previous figures, is available (*gerawav_v4.m*). This facilitates obtaining results for other simulation parameters.

3.2. NRD Estimation Examples Based on the Sawtooth Wave Derivative

The sawtooth wave is defined by (27) with $\phi_\ell = 0, \ell = 0, 1, \dots, L - 1$, i.e.,

$$x(t) = \sum_{\ell=0}^{L-1} A_\ell \sin((\ell + 1)\Omega_0 t), \tag{33}$$

where A_ℓ / A_0 is defined by (32). Its derivative is obtained as follows:

$$\begin{aligned} \frac{d}{dt}x(t) &= \sum_{\ell=0}^{L-1} A_\ell (\ell + 1) \frac{2\pi}{T_0} \cos((\ell + 1)\Omega_0 t) \\ &= \frac{2\pi A_0}{T_0} \sum_{\ell=0}^{L-1} \sin\left((\ell + 1)\Omega_0 t + \frac{\pi}{2}\right), \end{aligned} \tag{34}$$

and its negative derivative is as follows:

$$-\frac{d}{dt}x(t) = \frac{2\pi A_0}{T_0} \sum_{\ell=0}^{L-1} \sin\left((\ell + 1)\Omega_0 t - \frac{\pi}{2}\right). \tag{35}$$

The derivative of the sawtooth wave is illustrated in Figure 19 after magnitude normalization and after it has been contaminated by white Gaussian noise at 30 dB and 10 dB SNR.

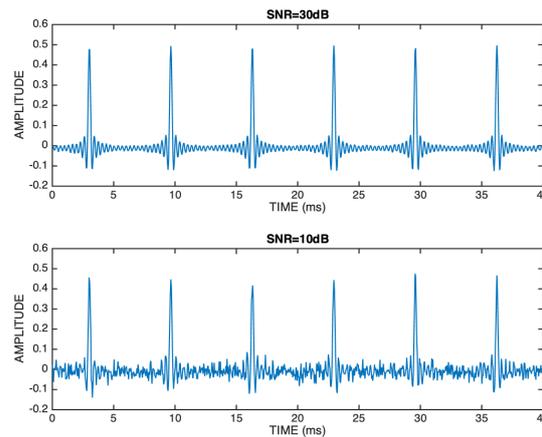


Figure 19. Illustration of the derivative of the sawtooth test signal when it is affected by noise at SNR = 30 dB (top), and at SNR = 10 dB (bottom).

In the case of the sawtooth wave derivative, according to (34), it results that $\phi_\ell = \pi/2$, $\ell = 0, 1, \dots, L - 1$ and, in the case of the negative of the sawtooth wave derivative, $\phi_\ell = -\pi/2$, $\ell = 0, 1, \dots, L - 1$. As a consequence, using (29), in the first case we obtain the following:

$$\text{NRD}_\ell = \frac{\frac{\pi}{2} - (\ell + 1)\frac{\pi}{2}}{2\pi} = -\frac{\ell}{4}, \ell = 0, 1, \dots, L - 1, \tag{36}$$

and, in the second case, we obtain the following:

$$\text{NRD}_\ell = \frac{-\frac{\pi}{2} + (\ell + 1)\frac{\pi}{2}}{2\pi} = \frac{\ell}{4}, \ell = 0, 1, \dots, L - 1. \tag{37}$$

We illustrate in Figure 20 the experimental results regarding the sawtooth wave derivative when FM = 2.5 Hz and SNR = 30 dB.

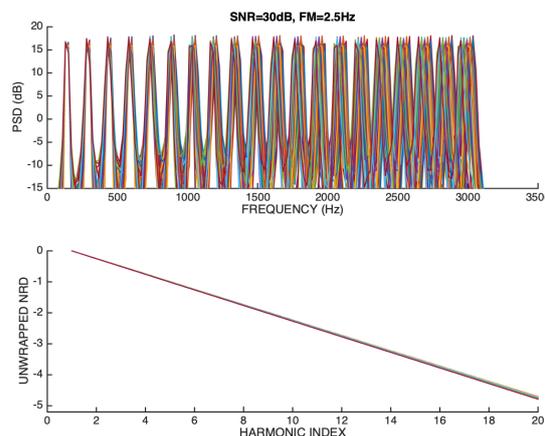


Figure 20. Overlay representation of magnitude spectra (top), and unwrapped NRD vectors (bottom) of a sawtooth derivative signal that is FM-modulated (2.5 Hz deviation around the mean) and whose SNR is 30 dB.

It can be confirmed that the unwrapped NRD vectors match quite precisely the ground-truth NRD model according to (36). This model is a simple first-order equation whose value is $\text{NRD} = -4.75$ when $\ell = 19$. This good match happens despite the fact that FM modulation is substantial, especially for higher-order harmonics. As noted in the previous subsection, this is a consequence of the fact that NRD is intrinsically independent of the fundamental frequency.

Similar conclusions are obtained in the practical estimation of the negative of the sawtooth derivative signal that is subject to FM modulation and noise contamination, as in the previous case. The results, when FM = 2.5 Hz and SNR = 30 dB, are displayed in Figure 21.

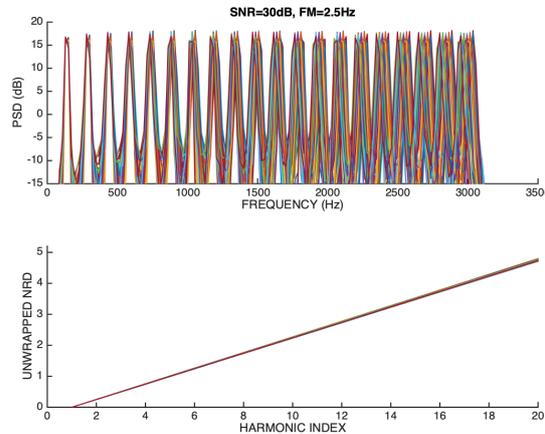


Figure 21. Overlay representation of magnitude spectra (**top**), and unwrapped NRD vectors (**bottom**) of the negative of a sawtooth derivative signal that is FM modulated (2.5 Hz deviation around the mean) and whose SNR is 30 dB.

It can be confirmed that the estimated unwrapped NRD vectors closely follow the ground-truth NRD model as defined in (37) and reach an NRD of $NRD = 4.75$ when $\ell = 19$.

The simulation results for the sawtooth derivative signal, and its negative, when FM = 2.5 Hz and SNR = 10 dB, are shown in Figure 22 and in Figure 23, respectively.

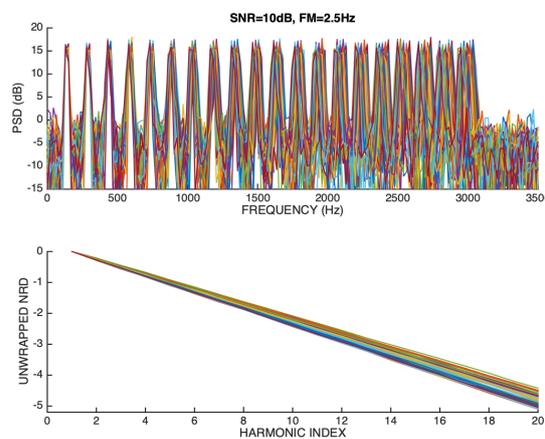


Figure 22. Overlay representation of magnitude spectra (**top**), and unwrapped NRD vectors (**bottom**) of a sawtooth derivative signal that is FM modulated (2.5 Hz deviation around the mean) and whose SNR is 10 dB.

It can be seen that, in both cases, the unwrapped NRD vectors show some dispersion which is mainly caused by the increased noise floor, which affects phase estimation more severely. In any case, even for this poor SNR level, there are no drastic mismatches relative to the ground-truth NRD model that applies to each case.

These examples suggest that the unwrapped NRD can also be used to detect the polarity of harmonic signals.

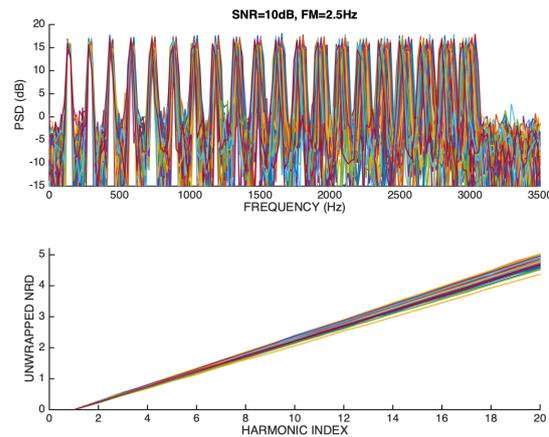
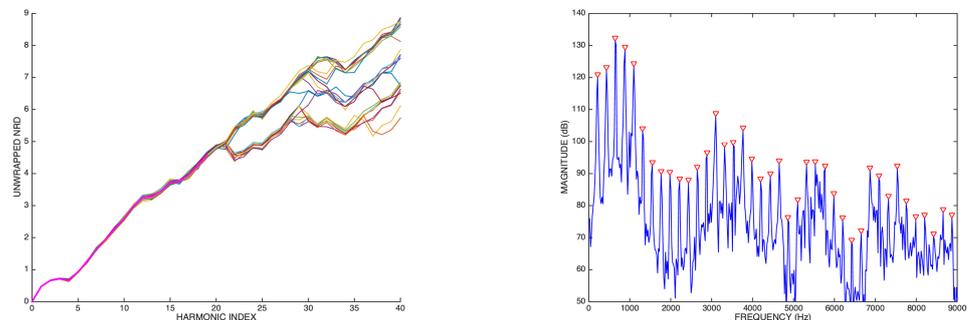


Figure 23. Overlay representation of magnitude spectra (**top**), and unwrapped NRD vectors (**bottom**) of the negative of a sawtooth derivative signal that is FM modulated (2.5 Hz deviation around the mean) and whose SNR is 10 dB.

3.3. NRD Estimation Examples Using Natural Voice Signals

An example of NRD-based analysis with real-world signals involves voiced sounds, i.e., those natural voice sounds that possess a periodic nature as a consequence of the vibration of the vocal folds in the larynx. Figure 24a illustrates an overlay of NRD vectors resulting from a sustained vowel (22,050 Hz sampling rate and 1 second long) that was uttered by a female speaker.



(a) Overlay representation of unwrapped NRD vectors resulting from a female sustained vowel utterance. The magenta line represents the mean of the first 20 NRD coefficients.

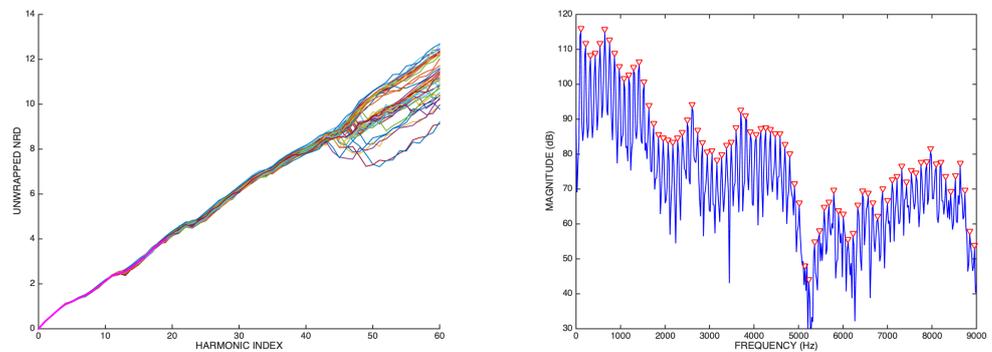
(b) Representative magnitude spectrum of the sustained vowel signal whose NRD vectors are represented in (a). Harmonics are signaled by red triangles.

Figure 24. Example of a harmonic phase model (a) and harmonic magnitude model (b) pertaining to a female sustained vowel signal.

It can be seen that NRD vectors are very consistent up to harmonic 22, beyond which, dispersion in the NRD vectors is apparent. This reflects the fact that phase estimation is disturbed after harmonic 22, which can be easily understood by observing one representative magnitude spectrum of the signal, as represented in Figure 24b. In fact, it can be confirmed that harmonics 22 and 23, at around 5 kHz, have a very small magnitude, which makes them particularly vulnerable to the noise influence, and leakage effects, which obviously adversely affect phase estimation. In any case, the most important harmonics are correctly characterized in their phase structure, which means reliable NRD models can be built that can be used in parametric-oriented speech processing, as illustrated in Figure 2. Our recent research, which assumes this framework, has revealed that when simple NRD

modeling is used in the synthesis process, signal reconstruction with transparent audio quality can be achieved, even if the NRD region that exhibits dispersion is replaced by a simple line resulting from extrapolation based on the reliable NRD region (i.e., for which the unwrapped NRD coefficients are consistent) [28].

A similar NRD analysis for a sustained vowel that was uttered by a male speaker is illustrated in Figure 25a.



(a) Overlay representation of unwrapped NRD vectors resulting from a male sustained vowel utterance. The magenta line represents the mean of the first 20 NRD coefficients. (b) Representative magnitude spectrum of the sustained vowel signal whose NRD vectors are represented in (a). Harmonics are signaled by red triangles.

Figure 25. Example of a harmonic phase model (a) and harmonic magnitude model (b) pertaining to a male sustained vowel signal.

It can be seen that the NRD vectors are quite consistent up to harmonics 43–44, which are located around 5 kHz. Figure 25b depicts one representative magnitude spectrum of the vowel signal and it can be confirmed that, similarly to the previous case, the magnitudes of higher-order harmonics approach the noise floor, which makes phase estimation essentially unreliable.

However, the region of the unwrapped NRD vectors that span the most important voice resonances can be reliably modeled. When this is combined with simple modeling of higher-order NRD coefficients through interpolation, very high-quality synthesis is achieved which sounds essentially indistinguishable from the original signal [28].

Our recent research in parametric voice re-synthesis, wideband speech coding, and speaker identification, highlighted that NRD patterns are idiosyncratic, and that they reflect more strongly the influence of the glottal excitation (i.e., the glottal pulse) than the influence of the vocal tract filter [23,28,29]. The fact that “the vocal tract phase during voicing is not important in achieving naturalness” has been understood already in previous studies ([30], p. 208).

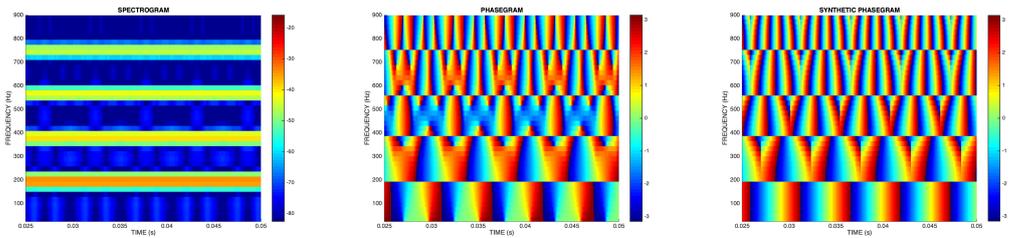
3.4. Demystifying the Phasegram

DFT-based non-parametric spectrum estimation techniques are commonly used to analyze, represent, and interpret the spectral properties of a given signal for which no strong assumptions are made. Given that those spectral properties can be characterized in terms of magnitude and phase, typically, 2D plots and 3D plots can be used in each case, although only seldom in the case of phase representation.

In the case of the magnitude spectrum, also known as a periodogram, a 2D plot is simply obtained by taking the absolute value of the short-time DFT of a windowed region of the signal as a function of frequency. This is illustrated, for example, in Figure 15. The graphical representation can be linear or logarithmic, either on the horizontal or vertical axis, or on both axes.

When the goal is to observe how the magnitude spectrum evolves through time, a 3D representation is created by abutting several magnitude spectra, next to each other, and

where a colormap is used to represent power spectral density (PSD). A new magnitude spectrum is obtained by sliding the short-time DFT window over the signal by a certain hop size that, typically, is less than the window length. Such a 3D representation is known as a spectrogram. Usually, the horizontal axis represents time, the vertical axis represents frequency, and the third dimension, which is perpendicular to the time–frequency plane, is represented by a specific color of a colormap denoting a suitable PSD range. A simple example is illustrated in Figure 26a that corresponds to the first four harmonics of a sawtooth signal and whose fundamental frequency is $F_0 = 187.34$ Hz.



(a) Spectrogram of a sawtooth signal. Only the first four harmonics are visible in the displayed frequency range. (b) Phasegram of a sawtooth signal whose spectrogram is represented in (a). The displayed frequency range encompasses the first four harmonics only. (c) Synthetic phasegram representing the same net phase information as that in (b).

Figure 26. Spectrogram (a), phasegram (b), and synthetic phasegram (c) characterizing the first 4 harmonics of a sawtooth waveform.

This particular example allows us to conclude that the signal is stationary because the PSD is steady through time. However, since spectrograms are blind to phase, the waveform could change its shape with time and may be quite different from the expected sawtooth waveform, albeit the magnitude spectrum remains exactly the same. In other words, the phase structure of the harmonics of the periodic wave may evolve with time. This can be detected by means of another three-dimensional representation called phasegram. The idea of the phasegram is the same as that of the spectrogram, except that the third dimension represents phase instead of power spectral density.

Although colorful and interesting to look at due to the repetitive patterns at different scales, the phasegram is difficult to interpret and, therefore, is not very useful. Evidence of this is that common signal processing-related numerical and graphical environments offer a callable ‘spectrogram’ function but do not offer a ‘phasegram’ function.

Here, we want to demystify the phasegram in Figure 26b (and that is valid for a stationary harmonic signal that is not too contaminated by noise) by showing that it is simply an eye-catching representation of the result of phase interpolation at different scales, and where the unique information is just the starting phase of the fundamental frequency, and a shift-invariant phase structure such as NRD.

In fact, the different scales are simply caused by the phase rotation of the different harmonics. As expected, the time evolution of the phase is periodic and, relative to the phase periodicity of the fundamental frequency, the time evolution for the second harmonic is two times as fast, three times as fast for the third harmonic, and so on. Thus, according to Equation (30), the phase evolution in time regarding the fundamental frequency is governed by $\Omega_0(t + t_0)$, and the phase evolution in time for any harmonic ℓ is governed by $(\ell + 1)\Omega_0(t + t_0) + 2\pi\text{NRD}_\ell$, which highlights that the only phase-related parameters that are unique are the starting phase of the fundamental frequency and the NRD coefficients. This explains the horizontal phase evolution for a given harmonic whose local maximum in the magnitude spectra has index k_ℓ . The fact that phase changes abruptly at harmonic frequencies between spectral index $k_\ell - 1$ and k_ℓ is explained, for example, with the help of Equation (8). In fact, such phase change is $\pi(1 - 1/N)$ which, in practical terms (i.e., for large N), consists of a phase inversion. Such a phase inversion does not occur for other

spectral indices because, in those cases, the polarity of the real-valued Dirichlet kernel function also inverts, thereby canceling out the phase inversion.

The remaining spectral lines are ‘passive’ as far as phase is concerned, which means that their phases are the result of a simple vertical interpolation between the phases of the ‘active’ spectral lines (i.e., those corresponding to local maxima in the magnitude spectrum). In order to illustrate this, we show in Figure 26c a simple replica of the phasegram of Figure 26b, which was built by defining first the horizontal phase evolution of the harmonics, by performing phase rotation by $\pi(1 - 1/N)$ for $k_\ell - 1$ spectral indices, and by performing piecewise vertical phase interpolation for the remaining spectral lines. This helps to reinforce the point—and conclusion—that in the case of stationary periodic signals, the phasegram is a highly redundant graphical representation whose unique information is just the starting phase of the fundamental frequency, and the NRD feature vector.

4. Conclusions

This paper focused on a harmonic signal analysis, modeling, and processing paradigm that eases significantly the representation, transformation, and synthesis of harmonic signals, especially from the point of view of the phase information.

In the first part of the paper, practical DFT-based approaches that build on a filter-bank perspective were discussed for estimating the starting phases of individual sinusoids. Their performance was characterized by considering the CRLB for the variance of an unbiased phase estimator. In particular, it was shown that contrary to harmonic frequency and magnitude estimation, accurate phase estimation depends only on ‘coarse search’ and not on ‘fine search’, which makes the estimation more robust. Six phase estimation alternatives were studied by combining two DFT-based filter banks and three different window functions. Results were explained in a reproducible manner.

In the second part of the paper, it was shown that the starting phases of individual sinusoids that are harmonically related may be converted into a phase-related feature (NRD) that expresses the holistic phase structure of a harmonic signal, has the advantage of being time-shift-invariant, helps to explain the waveform shape of a quasi-periodic signal, and helps to provide insight into the physical process that generates it.

Finally, it was shown that the unique information that exists in a phasegram resulting from a stationary harmonic signal consists of the starting phase of the fundamental frequency and the NRD feature vector.

Matlab code (https://github.com/Anibal-Ferreira/demo_AccPhaseEst, accessed on 1 January 2024) is provided that illustrates the most relevant concepts and results that are discussed in the paper.

Many application scenarios may benefit from the results in this paper paving the way for new research results, namely speech coding, pitch, and time-scale modification of speech, singing, audio, and music; speech enhancement; whispered-speech to voiced-speech conversion and voice rehabilitation; audio forensics; physiological signal analysis and diagnosis (e.g., using ECG signals); and monitoring of the operation of mechanical systems using sound.

Author Contributions: Conceptualization, A.F.; methodology, M.O., V.S. and A.S.; software, M.O., V.S. and A.F.; validation, M.O., V.S., A.S. and A.F.; formal analysis, V.S. and A.F.; investigation, M.O., V.S. and A.S.; writing—original draft preparation, A.F.; writing—review and editing, M.O., V.S., A.S. and A.F.; visualization, M.O., V.S. and A.F.; supervision, A.F.; project administration, A.F.; funding acquisition, A.F. All authors have read and agreed to the published version of the manuscript.

Funding: This work was developed within the context of Research Project COMPETE2030-FEDER-00923600 supported by the Portuguese Foundation for Science and Technology (FCT). Support was also received from the FCT-POCH Program under the SFRH/BSAB/150372/2019 Sabbatical Leave Fellowship.

Data Availability Statement: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Acknowledgments: Parts of this paper were written while the corresponding author (A.F.) was with the Digital Signal Processing Group at MIT during a sabbatical leave. The corresponding author would like to thank Prof. Alan Oppenheim for allowing magic to happen. The four co-authors would like to thank the anonymous reviewers for their comments and suggestions, which helped to improve the quality of the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

CRLB	Cramér–Rao lower bound
DFT	discrete Fourier transform
ECG	electrocardiogram
FFT	fast Fourier transform
FM	frequency modulation
HPM	harmonic phase model
MM	magnitude model
NRD	normalized relative delay
ODFT	odd-frequency discrete Fourier transform
PM	phase model
SNR	signal-to-noise ratio
T-F	time-to-frequency

References

1. Mowlae, P.; Kulmer, J.; Stahl, J.; Mayer, F. *Single Channel Phase-Aware Signal Processing in Speech Communication*; John Wiley and Sons Ltd.: Hoboken, NJ, USA, 2017.
2. Bonada, J.; Serra, X.; Amatriain, X.; Loscos, A. Spectral processing. In *DAFX: Digital Audio Effects*; Zölzer, U., Ed.; John Wiley & Sons Ltd.: Hoboken, NJ, USA, 2011; Chapter 10, pp. 393–445.
3. Quatieri, T.F.; McAulay, R.J. Audio Signal Processing Based on Sinusoidal Analysis/Synthesis. In *Applications of Digital Signal Processing to Audio and Acoustics*; Kahrs, M., Brandenburg, K., Eds.; Kluwer Academic Publishers: London, UK, 2002; Chapter 9, pp. 343–416.
4. Oppenheim, A.V.; Lim, J.S. The importance of phase in signals. *Proc. IEEE* **1981**, *69*, 529–541. [[CrossRef](#)]
5. Silva, J.M.; Oliveira, M.A.; Saraiva, A.F.; Ferreira, A.J.S. One-Step Discrete Fourier Transform-Based Sinusoid Frequency Estimation under Full-Bandwidth Quasi-Harmonic Interference. *Acoustics* **2023**, *5*, 845–869. [[CrossRef](#)]
6. Matusiak, A.; Borkowski, J.; Mroczka, J. Noniterative method for frequency estimation based on interpolated DFT with low-order harmonics elimination. *Measurement* **2022**, *196*, 111241. [[CrossRef](#)]
7. Belega, D.; Petri, D. Effect of noise and harmonics on sine-wave frequency estimation by interpolated DFT algorithms based on few observed cycles. *Signal Process.* **2017**, *140*, 207–218. [[CrossRef](#)]
8. Jacobsen, E.; Kootsookos, P. Fast, Accurate Frequency estimators. *IEEE Signal Process Mag.* **2007**, *24*, 123–125. [[CrossRef](#)]
9. Schoukens, J.; Pintelon, R.; Hamme, H.V. The Interpolated Fast Fourier Transform: A comparative study. *IEEE Trans. Instrum. Meas.* **1992**, *41*, 226–232. [[CrossRef](#)]
10. Keiler, F.; Marchand, S. Survey on extraction of sinusoids in stationary sounds. In Proceedings of the 5th International Conference on Digital Audio Effects (DAFx-02), Hamburg, Germany, 26–28 September 2002; pp. 51–58.
11. Ferreira, A.J.S. Accurate Estimation in the ODFT Domain of the Frequency, Phase and Magnitude of Stationary Sinusoids. In Proceedings of the 2001 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, USA, 21–24 October 2001; pp. 47–50.
12. Rife, D.C.; Boorstyn, R.R. Single-Tone Parameter Estimation from Discrete-Time Observations. *IEEE Trans. Inf. Theory* **1974**, *20*, 591–598. [[CrossRef](#)]
13. Quinn, B.G. Estimation of Frequency, Amplitude, and Phase from the DFT of a Time Series. *IEEE Trans. Signal Process.* **1997**, *45*, 814–817. [[CrossRef](#)]
14. Oppenheim, A.V.; Willsky, A.S.; Hamid, S. *Signals and Systems*, 2nd ed.; Pearson Education Limited: Upper Saddle River, NJ, USA, 1996.

15. Oppenheim, A.V.; Schafer, R.W. *Discrete-Time Signal Processing*; Pearson Higher Education Inc.: Upper Saddle River, NJ, USA, 2010.
16. Laroche, J.; Dolson, M. Phase-vocoder: About this phasiness business. In Proceedings of the Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, USA, 19–22 October 1997.
17. M.Kay, S. *Fundamentals of Statistical Signal Processing Estimation Theory*; Prentice Hall, Inc.: Saddle River, NJ, USA, 1993.
18. Liguori, C.; Paolillo, A.; Pignotti, A. Estimation of Signal Parameters in the Frequency Domain in the Presence of Harmonic Interference: A Comparative Analysis. *IEEE Trans. Instrum. Meas.* **2006**, *55*, 562–569. [[CrossRef](#)]
19. Vaidyanathan, P.P. *Multirate Systems and Filter Banks*; Prentice-Hall: Saddle River, NJ, USA, 1993.
20. Malvar, H. *Signal Processing with Lapped Transforms*; Artech House, Inc.: Norwood, MA, USA, 1992.
21. Painter, T.; Spanias, A. Perceptual Coding of Digital Audio. *Proc. IEEE* **2000**, *88*, 451–513. [[CrossRef](#)]
22. Bellanger, M. *Digital Processing of Signals*; John Willey & Sons: Hoboken, NJ, USA, 1989.
23. Ferreira, A.J.; Tribolet, J.M. A holistic glottal phase related feature. In Proceedings of the 21st International Conference on Digital Audio Effects (DAFx-18), Aveiro, Portugal, 4–8 September 2018.
24. Sousa, R.; Ferreira, A. Importance of the Relative Delay of Glottal Source Harmonics. In Proceedings of the 39th AES International Conference on Audio Forensics—Practices and Challenges, Hillerød, Denmark, 17–19 June 2010; pp. 59–69.
25. Stylianou, I. Harmonic Plus Noise Models for Speech, Combined with Statistical Methods, for Speech and Speaker Modification. Ph.D Thesis, École Nationale Supérieure des Télécommunications, Paris, France, 1996.
26. Federico, R.D. Waveform Preserving Time Stretching and Pitch Shifting for Sinusoidal Models of Sound. In Proceedings of the COST-G6 Digital Audio Effects Workshop, Limerick, Ireland, 6–8 December 1998; pp. 44–48.
27. Saratxaga, I.; Hernaez, I.; Erro, D.; Navas, E.; Sanchez, J. Simple representation of signal phase for harmonic speech models. *Electron. Lett.* **2009**, *45*, 381–383. [[CrossRef](#)]
28. Ferreira, A.; Oliveira, M.; Santos, V. On the mismatch between the phase structure of all-pole-based synthetic vowels and natural vowels. In Proceedings of the IEEE Workshop on Signal Processing Systems (SiPS), Cambridge, MA, USA, 4–6 November 2024; pp. 1–6.
29. Ferreira, A. Phonetic-oriented identification of twin speakers using 4-second vowel sounds and a combination of a shift-invariant phase feature (NRD), MFCCs and F0 information. In Proceedings of the 2019 AES International Conference on Audio Forensics, Porto, Portugal, 18–20 June 2019.
30. Quatieri, T.F.; McAulay, R.J. Phase Coherence in Speech Reconstruction for Enhancement and Coding Applications. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Glasgow, Scotland, 23–26 May 1989; pp. 207–210.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.