

Research Article

On the Characterization of Slowly Varying Sinusoids

Xue Wen (EURASIP Member) and Mark Sandler

Centre for Digital Music, Queen Mary University of London, Mile End Road, London E1 4NS, UK

Correspondence should be addressed to Xue Wen, xue.wen@elec.qmul.ac.uk

Received 4 March 2010; Revised 29 April 2010; Accepted 6 July 2010

Academic Editor: Roger Dannenberg

Copyright © 2010 X. Wen and M. Sandler. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We give a brief discussion on the amplitude and frequency variation rates of the sinusoid representation of signals. In particular, we derive three inequalities that show that these rates are upper bounded by the 2nd and 4th spectral moments, which, in a loose sense, indicates that every complex signal with narrow short-time bandwidths is a slowly varying sinusoid. Further discussions are given to show how this result helps providing extra insights into relevant signal processing techniques.

1. Introduction

Sinusoid representations of signals have been widely used in various signal processing areas including speech [1], music [2, 3], telecommunications [4]. Given any complex variable $x = x(t)$, its sinusoid representation is

$$x(t) = a(t)e^{j\varphi(t)}, \quad a(t) = \pm|x(t)|, \quad \varphi(t) = \text{Arg} \frac{x(t)}{a(t)}, \quad (1)$$

where the real variables $a(t)$ and $\varphi(t)$ are the *amplitude* and *phase angle* (or *phase* for short) of x . By this definition every nonzero complex variable has a unique sinusoid representation, up to the polarization of a and $2k\pi$ shift of φ , $k \in \mathbb{Z}$. In practice, these ambiguities are relieved by assuming that a and φ be continuous and smooth [5].

The parameter variations considered in this paper are the first and second derivatives of a , and the second derivative of φ , which, respectively, characterize amplitude and frequency modulations. We say that a sinusoid representation is slowly varying if these derivatives have small absolute values. Slowly varying sinusoids include many signals in speech, music, and telecommunications that we technically handle as sinusoids [4, 6, 7], and are the central elements of sinusoid modelling.

Spectral properties of slowly varying sinusoids in terms of parameter variation have been well studied, and have

been compared to those of stationary sinusoids [8–10]. One result [8] relates the 2nd moment of the energy spectrum of a time-varying sinusoid to its amplitude and phase variations by

$$\begin{aligned} & \int (\omega - \omega_0)^2 |X(\omega)|^2 d\omega \\ &= \int \left(\left(\frac{a'(t)}{a(t)} \right)^2 + (\varphi'(t) - \omega_0)^2 \right) a^2(t) dt, \end{aligned} \quad (2)$$

where $X(\omega) = (2\pi)^{-1/2} \int x(t)e^{-j\omega t} dt$ is the Fourier transform of x , and ω_0 is a real constant. Given the Parseval's equation $\int |X(\omega)|^2 d\omega = \int a^2(t) dt \equiv \|x\|^2$, we divide both sides of (2) by $\|x\|^2$ and get

$$\begin{aligned} & \frac{\int (\omega - \omega_0)^2 |X(\omega)|^2 d\omega}{\int |X(\omega)|^2 d\omega} \\ &= \frac{\int \left(\left(\frac{a'(t)}{a(t)} \right)^2 + (\varphi'(t) - \omega_0)^2 \right) a^2(t) dt}{\int a^2(t) dt}. \end{aligned} \quad (3)$$

We introduce an operator $\overline{\bullet}_x$, where \bullet stands for the operand, to simply if expressions like (3) that will appear all through this paper. Given $x(t) = a(t)e^{j\varphi(t)} \in \mathbb{C}^2(-\infty, +\infty)$

and its Fourier transform $X(\omega)$, we define the operator $\overline{\bullet}_x$, which maps a real function of t or ω to a real number, by

$$\overline{f(t)}_x = \frac{\int f(t) \cdot a^2(t) dt}{\|x\|^2}, \quad \overline{F(\omega)}_x = \frac{\int F(\omega) \cdot |X(\omega)|^2 d\omega}{\|x\|^2},$$

where $\|x\|^2 = \int a^2(t) dt = \int |X(\omega)|^2 d\omega$.

(4)

$\overline{\bullet}_x$ is interpreted as linear averaging weighted by the energy density of x in time or frequency domain, depending on the operand. For example, \bar{t}_x and $\bar{\omega}_x$ are the time and spectral centroids of x , respectively. Using this notation, (3) can be written as

$$\overline{(\omega - \omega_0)^2} = \overline{\left(\frac{a'(t)}{a(t)}\right)^2} + \overline{(\varphi'(t) - \omega_0)^2}, \quad (5)$$

where we have omitted the subscript x from $\overline{\bullet}_x$. At $\omega_0 = \bar{\omega}_x$, the term $\overline{(\omega - \omega_0)^2}_x$ reaches its minimum whose square root defines the L^2 -bandwidth of x (likewise we call $(|\omega - \bar{\omega}_x|^{p_x})^{1/p}$ the L^p -bandwidth of x at ω_0 , as it is the $|X|^2$ -weighted L^p -norm of $\omega - \bar{\omega}_x$). According to (5), the L^2 -bandwidth of x is upper-bounded if a'/a and the range of φ' are. However, as the latter may grow very large over a long time span, (5) does not imply that slowly varying sinusoids have concentrated spectra, as stationary sinusoids do.

On the other hand, if φ'' is small, then φ' does not vary much over a short time. In this case, (2) guarantees that x has concentrated short-time Fourier transform (STFT), provided that this is calculated with a low-pass window function $w(t)$. In fact, using (2), it is trivial to show that

$$\int (\omega - \omega_0)^2 |X_w(\omega)|^2 d\omega \leq \sup(\varphi''^2) \int (t - t_0)^2 (w(t)a(t))^2 dt + \int (w(t)a(t))^2 dt, \quad (6)$$

for any fixed t_0 , where $X_w(\omega)$ is the Fourier transform of xw (i.e., STFT of x with window w) and $\omega_0 = \varphi'(t_0)$. We call the bandwidth of xw the *short-time bandwidth* of x with window w . Figure 1 illustrates the concept of short-time bandwidth applied to a linear chirp in the time-frequency plane. The solid line depicts the angular frequency of the chirp as a function of time. Its short-time spectra are evaluated using two rectangular windows, w_1 and w_2 , whose durations are marked along the time axis. Although the linear chirp is not band limited, each window captures a band-limited portion of it. The frequency content captured by window w_1 distributes uniformly over $(\omega_{11}, \omega_{12})$ while that by window w_2 distributes uniformly over $(\omega_{21}, \omega_{22})$. If both windows contain plenty periods of the sinusoid, then the bandwidths of the two spectra, bw_1 and bw_2 , are roughly proportional to $\omega_{12} - \omega_{11}$ and $\omega_{22} - \omega_{21}$, which are in turn proportional to the length of w_1 and w_2 and the chirp rate of the sinusoid.

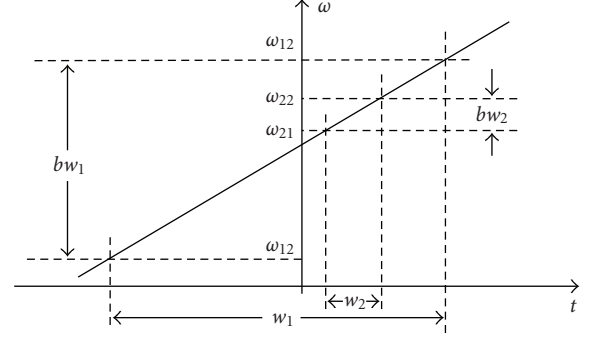


FIGURE 1: Measuring short-time bandwidth with different windows.

A loosely related result is given in [9] by direct comparison of the Fourier transforms of xw and yw where $y = a_0 e^{j(\varphi_0 + \omega_0 t)}$ is a stationary sinusoid, with $a_0 = a(t_0)$, $\omega_0 = \varphi'(t_0)$ and $\varphi_0 = \varphi(t_0)$, and w is centred at t_0 . It is shown that their difference is upper bounded by

$$|X_w(\omega) - Y_w(\omega)| \leq \sup|\varphi''| \cdot \frac{a_0}{2} \int |(t - t_0)^2 w| dt + \sup|a'| \cdot \int |(t - t_0)w| dt, \quad (7)$$

where $Y_w(\omega)$ is the Fourier transform of yw . Since Y_w is concentrated at ω_0 , (7) shows that X_w is also concentrated at ω_0 as long as $|\varphi''|$ and $|a'|$ remain small. Mallet [9] used the term *windowed Fourier ridge* to describe this time-frequency distribution as it involves a spectral peak that evolves in time. We call it *spectral ridge* for short.

Despite all these studies on sinusoid representations, one question has been overlooked: what type of signals can be modelled as slowly varying sinusoids? From the results above, it is obvious that signals with wide short-time bandwidths, such as wide-band noise, *cannot* be slowly varying sinusoids. In this paper, we consider the inverse: do narrow short-time bandwidths always imply a slowly varying sinusoid? In other words, does a concentrated short-time spectrum necessarily set certain upper bounds on a' and φ'' ? The concentration of a spectrum is measured by the moments of the spectral energy distribution (i.e., normalized energy spectrum), or *spectral moments* for short. The n th spectral moment of x with centre ω_0 is given by $\overline{(\omega - \omega_0)^n}_x$ and can be interpreted as the *biased L^n -bandwidth*, as it becomes the L^n -bandwidth if $\omega_0 = \bar{\omega}_x$. From (5), it is obvious that $\overline{(\omega - \omega_0)^2}_x$ upper bounds the average amplitude derivative and average frequency departure from ω_0 . However, the 2nd moment is not enough to set an upper bound on φ'' . In what follows, we provide a new result that employs higher spectral moments to upper bound φ'' as well as a' and a'' .

2. Parameter Variation Rate Upper Bounds in Terms of 2nd and 4th Spectral Moments

High spectral moments are less often used than the 2nd moment. Notably, [10] employs an operator approach that relates an arbitrary spectral moment to the derivatives of amplitude and phase. In particular, regarding the 4th moment, we have

$$\begin{aligned} \overline{(\omega - \omega_0)^4} &= \left((\varphi'(t) - \omega_0)^2 - \frac{a''(t)}{a(t)} \right)^2 \\ &\quad + \left(2(\varphi'(t) - \omega_0) \frac{a'(t)}{a(t)} + \varphi''(t) \right)^2, \end{aligned} \quad (8)$$

where ω_0 can take arbitrary value. From (5) and (8) we can prove that

$$\begin{aligned} \left| \frac{a''(t)}{a(t)} \right| + \left(\frac{a'(t)}{a(t)} \right)^2 + |\varphi''(t)| \\ \leq \sqrt{2(\omega - \omega_0)^4} + 2(\omega - \omega_0)^2 \end{aligned} \quad (9)$$

Equation (9) states that the average parameter variation rates are upper bounded by the 2nd and 4th spectral moments. Two bounds respectively regarding amplitude and phase can also be obtained as

$$\left| \frac{a''(t)}{a(t)} \right| + \left(\frac{a'(t)}{a(t)} \right)^2 \leq \sqrt{(\omega - \omega_0)^4} + (\omega - \omega_0)^2, \quad (10)$$

$$|\varphi''(t)| \leq \sqrt{(\omega - \omega_0)^4} + (\omega - \omega_0)^2. \quad (11)$$

A detailed proof of (9)~(11) is given in the appendix. Identities hold in these inequalities if specific couplings exist between amplitude and phase variations (see the appendix). From physics' point of view, (9)~(11) states that finite 2nd and 4th spectral moments can only "contain" limited amount of modulation of amplitude and frequency; to allow faster modulations one has to increase the spectral moments. The 2nd and 4th spectral moments are further connected through the *biased kurtosis* at ω_0 , defined as

$$k_u(\omega_0)_x = \frac{\overline{(\omega - \omega_0)^4}_x \cdot \|x\|^2}{(\overline{(\omega - \omega_0)^2}_x)^2}. \quad (12)$$

We call it "biased" because it is evaluated using an arbitrary centre ω_0 instead of the true centroid $\bar{\omega}_x$. (12) gives

$$\sqrt{\overline{(\omega - \omega_0)^4}_x} = \overline{(\omega - \omega_0)^2}_x \cdot \|x\|^{-1} \cdot k_u^{1/2}(\omega_0)_x. \quad (13)$$

The kurtosis is generally understood to represent the "peakedness" of $|X(\omega)|$: a small kurtosis indicates bulky peak and sharp tails; a large kurtosis indicates narrow peak and heavy tails. In the context of (13), (9)~(11) states that the for the same 2nd moment, more modulation is allowed by larger kurtoses.

Inequalities (9)~(11) can also be directly applied to windowed Fourier transforms by replacing a with wa where w is the window function. As $(wa)' / wa = w' / w + a' / a$, if $(wa)'$ and w' are upper bounded, then so is a' . (9)~(11) indicate that the sinusoid representation of a signal whose STFT forms a spectral ridge is necessarily slowly varying in terms of short-time average of parameter variation rates. This, together with our comments in the introduction, completes the following statement.

A complex signal has slowly varying sinusoid representation if and only if it has narrow short-time bandwidths. ()*

Here we have changed the term "spectral moment" to "bandwidth" considering that the L^p -bandwidth is simply the p th spectral moment computed with $\omega_0 = \bar{\omega}_x$. In (*) the "only if" part comes from the previous studies we summarized by (6) in the introduction; the "if" part comes from our results (9)~(11). The plural form in "bandwidths" refers to the values evaluated in both L^2 - and L^4 -norms at different points over the whole duration. A quantitative presentation of (*) is given by rewriting (6) and (10), (11) employing a sliding window

$$\begin{aligned} \overline{(\omega - \varphi'(\tau))^2}_{w_\tau x} \\ \leq \sup(\varphi''^2)(t - \tau)^2_{w_\tau x} + \left(\frac{(w_\tau(t)a(t))'}{w_\tau(t)a(t)} \right)^2_{w_\tau x}, \end{aligned} \quad (14)$$

$$\begin{aligned} \left| \frac{(w_\tau(t)a(t))''}{w_\tau(t)a(t)} \right|_{w_\tau x} + \left(\frac{(w_\tau(t)a(t))'}{w_\tau(t)a(t)} \right)^2_{w_\tau x} \\ \leq \sqrt{(\omega - \varphi'(\tau))^4_{w_\tau x}} + (\omega - \varphi'(\tau))^2_{w_\tau x}, \end{aligned} \quad (15)$$

$$|\varphi''(t)|_{w_\tau x} \leq \sqrt{(\omega - \varphi'(\tau))^4_{w_\tau x}} + (\omega - \varphi'(\tau))^2_{w_\tau x}, \quad (16)$$

where $w_\tau(t) = w(t - \tau)$ is the window function centred at τ .

We notice that φ'' is measured differently in (14) and (16), giving a double meaning to "slowly varying frequency" in (*). For this reason, (*) does not actually give a pair of strictly converse statements, and the equivalence between slowly varying sinusoids and spectral ridges, as established by (14)~(16), should only be considered qualitatively. Nevertheless, by these results we have partially answered the question of what kind of signals can be modelled as slowly varying sinusoids. In the rest of this paper, we focus on (*) as a guideline and see how it relates to various sinusoid modelling practices.

3. Discussions and Conclusion

3.1. Combining Sinusoids with Close Frequencies. Beating [11] is a well known effect observed from adding two sinusoids with similar frequencies, in which they "melt" into a single tone with additional modulation. This phenomenon can be easily explained by (*): as slowly varying sinusoids have short-time spectral energy concentrated near their

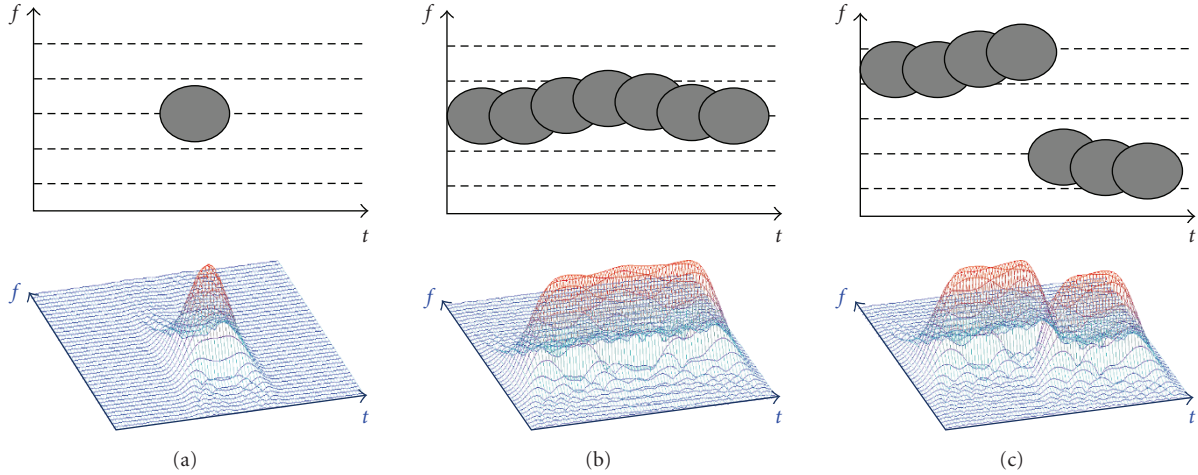


FIGURE 2: Spectrograms of atomic synthesis.

angular frequencies, if the frequencies are close, then their sum also has concentrated short-time spectral energy, therefore is also a slowly varying sinusoid. Additional modulation may be introduced as the result of a wider bandwidth contributed by the small interval between the participant frequencies. A quantitative proof of this argument is given in [12], which leads to an additive re-estimation algorithm for measuring parameters of slowly varying sinusoids.

Statement (*) also reveals the difficulty in separating close sinusoids by the slow variation criterion alone. Since there are infinite number of ways to divide a spectral ridge into 2 or more subridges, and since all narrow ridges are necessarily slowly varying sinusoids, there are infinite number of separations that are slowly varying.

3.2. Atomic Decomposition. In time-frequency analysis, the term *atom* refers to basic waveforms with concentrated time and frequency localization into which a signal is decomposed. Windowed sinusoid atoms have been used in short-time Fourier and Gabor transforms [13], matching pursuits [14], auditory scene analysis [15], and methods for approximating time-varying sinusoids [16, 17]. An *overlap-add sinusoidal model* was proposed [16] in the typical form of atomic decomposition

$$x(t) = \sum_i w^i(t) a_i e^{j(\varphi_i + 2\pi f_i t)}, \quad (17)$$

where x is a slowly varying sinusoid, a_i , f_i and φ_i are constants for given i , and w^i is the overlap-add window centred at the i th reference point, say t_i . Adjacent windows are arranged to have considerable overlap.

The use of overlapping stationary sinusoids to approximate time-varying sinusoids is partially justified by (*). Since windowed sinusoid atoms have concentrated spectral energy at the sinusoid frequencies, their sum will form a narrow spectral ridge as long as frequencies of adjacent atoms are close enough so that the result represents a slowly varying sinusoid. It is also apparent that if there is a large frequency jump between any adjacent atoms, then the sum

is no longer a slowly varying sinusoid, indicating that (17) is not a suitable representation of x .

Figure 2 illustrates atomic synthesis with 3D spectrogram, in which the sinusoids are directly visible as ridges. Images in the top row show the atoms in the time-frequency plane, in which frequency bins are marked out by dashed lines; images in the bottom row are the corresponding spectrograms. Figure 2(a) shows a single atom whose spectrogram consists of a single peak. Figure 2(b) shows a signal constructed from seven overlapping atoms without a frequency jump between adjacent atoms. A spectral ridge is clearly observed from its spectrogram. In Figure 2(c), we include a frequency jump of three bins between the 4th and 5th atoms, which is enough to break the ridge in Figure 2(b) into two separate ridges. Representing this signal as *two* sinusoids will allow much slower modulation rates than a single-sinusoid representation.

3.3. Real Sinusoids and Analytic Signals. A real slowly varying sinusoid can always be written as the sum of two conjugate complex slowly varying sinusoids. According to (*), its spectrogram is made up of two spectral ridges. To find a slowly varying *double-sinusoid* representation for a real sinusoid, one only needs to separate the spectrogram into two parts, each containing one ridge. This separation is generally not unique. If the two parts are conjugate to each other, then the real part of the corresponding complex sinusoids equals half of the real sinusoid.

Most of the real sinusoids we encounter in practice have always-positive frequencies so that each spectral ridge lies in a half plane on either side of the time axis. In this case, the most natural separation is obtained by splitting the spectrum along time axis, which leads to analytic complex sinusoids [5]. We observe by (*) that the analytic representation is slowly varying by design, if the concerned sinusoid does have a slowly varying representation at all.

As an example, we illustrate the spectrogram of a real linear chirp in Figure 3(a) and that of the corresponding complex linear chirp in Figure 3(b). Amplitude values are

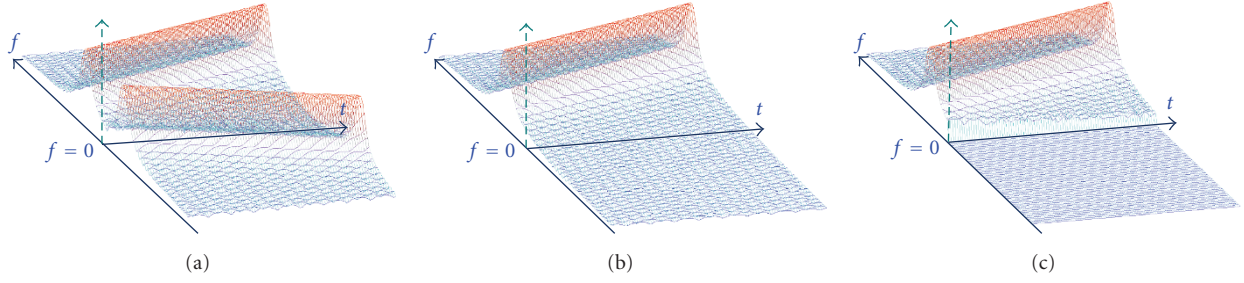


FIGURE 3: Two conjugate separations of a linear chirp.

warped to the 5th root to make small amplitudes visible. Figure 3(c) shows a *nearly* analytic sinusoid obtained by setting the spectrogram in Figure 3(a), rather than the spectrum, to zero over negative frequencies. Although Figure 3(b) and Figure 3(c) look different, both are slowly varying complex sinusoids with the real part equalling half the linear chirp in Figure 3(a).

3.4. More on Slowly Varying Real Sinusoids. Although the analytic representation has almost guaranteed slow variation, it is generally not the slowest varying, judged by the values of a' and φ'' . In [18], we have measured the parameter variation rates by

$$I = \eta \int \frac{a'(t)^2}{a^2(t)} dt + (1 - \eta) \int \varphi''(t)^2 dt, \quad (18)$$

where $\eta \in (0, 1)$ is a balancing factor. A necessary condition for a and φ to give the slowest varying representation, in the sense of minimizing I in (18), is given as

$$-\frac{\eta}{1 - \eta} \frac{a(t)a''(t) - a'(t)^2}{a^2(t)} \tan \varphi(t) + \varphi^{(4)}(t) = 0, \quad (19)$$

where $\varphi^{(4)}$ is the 4th-order derivative of φ . This condition is automatically satisfied regardless of η if a is exponential and φ is trinomial, but can be more constraining in other cases.

Nonunique representations of real sinusoids may cause problems in evaluating sinusoid estimators. For example, while a complex linear chirp defines a linear frequency for its corresponding real chirp, the latter's analytic counterpart defines a nonlinear frequency which is no less convincing. Fortunately, in [18] we have shown that the difference between various sinusoid representations of the same real signal is bounded by their parameter variation rates. Consequently, if a signal has multiple slowly varying sinusoid representations, then they are close to each other.

3.5. Conclusion. In this paper, we have given three inequalities that bound the parameter variation rates of the sinusoid representation of a complex signal by its 2nd and 4th spectral moments, indicating that every complex signal with narrow short-time bandwidths is necessarily a slowly varying sinusoid. This, together with several previous results, serves to argue towards the equivalence between slowly varying sinusoids and signals with narrow short-time bandwidths, which, in return, provides extra insights into various aspects of sinusoid modelling.

Appendix

A. Proof of (9)~(11)

All variables in the appendix, with the exception of ω , are functions of t , and all $\overline{\bullet}_x$ operators are defined for the same x . For simplicity, we omit t from all function notations and subscript x from operator $\overline{\bullet}_x$.

We first summarize without proof a few properties of the $\overline{\bullet}$ operator defined by (4) as follows:

- (1) (linearity) $\overline{f + g} = \overline{f} + \overline{g}$, $\overline{af} = a\overline{f}$, $a \in \mathbb{R}$,
- (2) (monotonicity) $f \leq g \implies \overline{f} \leq \overline{g}$,
- (3) (Cauchy-Schwarz) $|\overline{fg}|^2 \leq \overline{f^2} \cdot \overline{g^2}$, $|\overline{f}|^2 \leq \overline{f^2}$,

and immediate results of (1) and (2):

- (4) (triangular inequality) $|\overline{f}| - |\overline{g}| \leq \overline{|f + g|} \leq |\overline{f}| + |\overline{g}|$,
- (5) $2\overline{fg} \leq \overline{f^2} + \overline{g^2}$.

We quote (5) and (8) here as (A.1) and (A.2)

$$\overline{(\omega - \omega_0)^2} = \overline{\left(\frac{a'}{a}\right)^2} + \overline{(\varphi' - \omega_0)^2}, \quad (\text{A.1})$$

$$\overline{(\omega - \omega_0)^4} = \overline{\left((\varphi' - \omega_0)^2 - \frac{a''}{a}\right)^2} + \overline{\left(\frac{2(\varphi' - \omega_0)a'}{a} + \varphi''\right)^2}. \quad (\text{A.2})$$

From (A.2) using (3), we get

$$\overline{(\omega - \omega_0)^4} \geq \overline{\left((\varphi' - \omega_0)^2 - \frac{a''}{a}\right)^2} + \overline{\left(\frac{2(\varphi' - \omega_0)a'}{a} + \varphi''\right)^2}, \quad (\text{A.3})$$

in which identity holds if there exist C_1, C_2 , $|(\varphi' - \omega_0)^2 - a''/a| = C_1$, $|2(\varphi' - \omega_0)a'/a + \varphi''| = C_2$. (A.3) yields three inequalities

$$\sqrt{2\overline{(\omega - \omega_0)^4}} \geq \overline{\left((\varphi' - \omega_0)^2 - \frac{a''}{a}\right)} + \overline{\left(\frac{2(\varphi' - \omega_0)a'}{a} + \varphi''\right)}, \quad (\text{A.4})$$

in which identity holds if there exist C so that

$$\left| (\varphi' - \omega_0)^2 - \frac{a''}{a} \right| = C, \quad (\text{A.5a})$$

$$\left| \frac{2(\varphi' - \omega_0)a'}{a} + \varphi'' \right| = C; \quad (\text{A.5b})$$

$$\sqrt{(\omega - \omega_0)^4} \geq \left| (\varphi' - \omega_0)^2 - \frac{a''}{a} \right|, \quad (\text{A.6})$$

in which identity holds if there exist C so that

$$\frac{2(\varphi' - \omega_0)a'}{a} + \varphi'' = 0, \quad (\text{A.7a})$$

$$\left| (\varphi' - \omega_0)^2 - \frac{a''}{a} \right| = C, \quad (\text{A.7b})$$

$$\sqrt{(\omega - \omega_0)^4} \geq \left| \frac{2(\varphi' - \omega_0)a'}{a} + \varphi'' \right|, \quad (\text{A.8})$$

in which identity holds if there exist C so that

$$(\varphi' - \omega_0)^2 - \frac{a''}{a} = 0, \quad (\text{A.9a})$$

$$\left| \frac{2(\varphi' - \omega_0)a'}{a} + \varphi'' \right| = C. \quad (\text{A.9b})$$

Using (4) and (A.1), we get

$$\begin{aligned} \left| (\varphi' - \omega_0)^2 - \frac{a''}{a} \right| &\geq \left| \frac{a''}{a} \right| - \overline{(\varphi' - \omega_0)^2} \\ &= \left| \frac{a''}{a} \right| + \left(\frac{a'}{a} \right)^2 - \overline{(\omega - \omega_0)^2}, \end{aligned} \quad (\text{A.10})$$

in which identity holds if

$$\frac{a''}{a} \geq (\varphi' - \omega_0)^2, \quad (\text{A.11a})$$

using (4), (5), and (A.1), we get

$$\begin{aligned} \left| \frac{2(\varphi' - \omega_0)a'}{a} + \varphi'' \right| &\geq |\varphi''| - \left| \frac{2(\varphi' - \omega_0)a'}{a} \right| \\ &\geq |\varphi''| - \overline{(\varphi' - \omega_0)^2} - \left(\frac{a'}{a} \right)^2 \\ &= |\varphi''| - \overline{(\omega - \omega_0)^2}, \end{aligned} \quad (\text{A.12})$$

in which identity holds if

$$\frac{a'}{a} = \pm(\varphi' - \omega_0), \quad (\text{A.13a})$$

$$\mp \varphi'' \geq \pm \frac{2(\varphi' - \omega_0)a'}{a} \geq 0. \quad (\text{A.13b})$$

Substituting (A.10) into (A.6), we get (10), (A.12) into (A.8), we get (11), and both into (A.4), we get (9). This concludes the proof.

To find out when identity holds in (9), we jointly solve (A.5a), (A.5b), (A.11a), (A.13a), and (A.13b). From (A.5a) and (A.10), we have

$$\frac{a''}{a} - (\varphi' - \omega_0)^2 = C \geq 0, \quad (\text{A.5a\&A.11a})$$

and from (A.13a), we have

$$\frac{a''}{a} = \pm \left(\frac{a'}{a} \right) (\varphi' - \omega_0) \pm \varphi'' = (\varphi' - \omega_0)^2 \pm \varphi''. \quad (\text{A.13c})$$

These two together give

$$\pm \varphi'' = C \geq 0, \quad (\text{A.5a, A.11a\&A.13c})$$

which, together with (A.13b), gives $C = 0$, and consequently $\varphi'' = 0$, $a'/a = \varphi' - \omega_0 = 0$; that is, the sinusoid is of constant amplitude and angular frequency at ω_0 . Notice that this condition implies $\int a^2 dt$ being infinitely large so that $\overline{\bullet}$ is not properly defined by (4). However, identity still holds in (9) for stationary sinusoids if we accept that $0/\infty = 0$.

Similarly, we derive that identity holds in (11) if the sinusoid has constant amplitude and angular frequency at ω_0 , and in (10) if $2(\varphi' - \omega_0)a'/a + \varphi'' = 0$ and there exist $C \geq 0$, $a''/a - (\varphi' - \omega_0)^2 = C$.

Acknowledgment

This paper was supported by the EPSRC EP/E017614/1 project OMRAS2 (Online Music Recognition and Searching).

References

- [1] R. J. McAulay and T. F. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, no. 4, pp. 744–754, 1986.
- [2] X. Serra, "Musical sound modeling with sinusoids plus noise," *Musical Signal Processing*, pp. 91–122, 1997.
- [3] G. Peeters and X. Rodet, "SINOLA: a new analysis/synthesis method using spectrum peak shape distortion, phase and reassigned spectrum," in *Proceedings of the International Computer Music Conference (ICMC '99)*, pp. 153–156, Beijing, China, 1999.
- [4] A. B. Carlson, *Communication Systems*, McGraw-Hill, New York, NY, USA, 2nd edition, 1981.
- [5] L. Cohen, P. Loughlin, and D. Vakman, "On an ambiguity in the definition of the amplitude and phase of a signal," *Signal Processing*, vol. 79, no. 3, pp. 301–307, 1999.
- [6] G. Fant, "The acoustics of speech," in *Proceedings of the 3rd International Conference Solar Air-Conditioning*, pp. 188–201, Stuttgart, Germany, 1959.
- [7] N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments*, Springer, New York, NY, USA, 2nd edition, 1998.

- [8] L. Cohen and C. Lee, "Standard deviation of instantaneous frequency," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP '89)*, vol. 4, pp. 2238–2241, May 1989.
- [9] S. Mallat, *A Wavelet Tour of Signal Processing*, chapter 4, Academic Press, 2nd edition, 1999.
- [10] K. L. Davidson and P. J. Loughlin, "Instantaneous spectral moments," *Journal of the Franklin Institute*, vol. 337, no. 4, pp. 421–436, 2000.
- [11] L. A. Jeffress, "Beating sinusoids and pitch changes," *Journal of the Acoustical Society of America*, vol. 43, no. 6, p. 1464, 1968.
- [12] X. Wen and M. Sandler, "Additive and multiplicative reestimation schemes for the sinusoid modeling of audio," in *Proceedings of 17th European Signal Processing Conference (EUSIPCO '09)*, Glasgow, UK, 2009.
- [13] D. Gabor, "Theory of communication," *Journal of the Institute of Electronics Engineers*, vol. 3, pp. 429–459, 1946.
- [14] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.
- [15] G. J. Brown and M. Cooke, "Computational auditory scene analysis," *Computer Speech and Language*, vol. 8, no. 4, pp. 297–336, 1994.
- [16] E. B. George and M. J. T. Smith, "Analysis-by-synthesis/overlap-add sinusoidal modeling applied to the analysis and synthesis of musical tones," *Journal of the Audio Engineering Society*, vol. 40, no. 6, pp. 497–516, 1992.
- [17] M. Davy and S. J. Godsill, "Bayesian harmonic models for musical signal analysis," in *Bayesian Statistics 7*, Oxford University Press, Oxford, UK, 2003.
- [18] X. Wen, *Harmonic sinusoid modelling of tonal music events*, Ph.D. thesis, University of London, London, UK, 2007.