ROBUST CHIRP PARAMETER ESTIMATION FOR HANN WINDOWED SIGNALS

Aaron S. Master[*] *and Yi-Wen Liu*

Stanford University,
Center for Computer Research in Music and Acoustics,
Stanford, CA 94305-8180 USA

ABSTRACT

The sinusoidal model has been a fundamentally important signal representation for coding and analysis of audio. We present an enhancement to sinusoidal modeling in the form of a linear frequency chirp parameter estimator applicable to Hann-windowed quasi-sinusoidal signals. The estimator relies on models of the phase curvature and peak width of a given chirp signal's FFT magnitude domain peak. We show that different models are applicable for smaller and larger values of the chirp parameter, derived respectively from Taylor series and Fresnel integral analysis of the signal. We construct an estimator for the transition region between the two models via a neural net. Results indicate that the estimator is robust to noise and outperforms any known chirp parameter estimators for Hann windowed signals.

1. INTRODUCTION

The sinusoidal model [1, 2] has been a fundamentally important signal representation for coding and analysis of audio. In conventional sinusoidal modeling, the parameters used are frequency, amplitude and phase of each frequency component. We presently consider the enhancement of sinusoidal modeling by using an additional parameter for each such component: the rate of linear increase or decrease in instantaneous frequency.

Though estimators for this parameter have been investigated for Gaussian-windowed signals [3, 4], estimators for Hann windowed signals have proven elusive. In audio applications such as coding, the Hann window is often more desirable than the Gaussian window, especially due to its constant-overlap-and-add (COLA) property. Hence, our estimators provide a valuable tool for such applications. A previous attempt at a chirp estimator for Hann windowed signals [5] relied on a model that did not apply when the chirp parameter became small (as small as often seen in speech and music signals). Though the error manifested itself in a predictable way incorporated into the estimator via a least squares solution, the analysis lacked a rigorous definition of the point at which the basic model required modification. The current models, however, are valid for specific ranges of the chirp parameter.

We apply different models for larger and smaller values of the chirp parameter alpha (α); both ranges are seen in audio signals [6]. The first model corresponds to a Fresnel integral analysis (for larger α) and the second to a Taylor series analysis of the signal's DFT (for smaller α). The Fresnel model yields an expression for the DFT magnitude domain peak shape, and both models yield expressions for the curvature of the DFT phase corresponding to the FFT magnitude peak. The Taylor series model has the novel property that it is valid for DFT phase curvature modeling when any zero phase symmetric time windowing function is used.

In section 2, we derive expressions for relevant portions of a chirp signal's DFT, based on our models. We use a highly simplified, but completely generalizeable chirp signal, and explicitly state the range of validity of each model in terms of α. In section 3, we solve the expressions given in section 2 for the chirp parameter α, yielding estimators. We also make important notes about the application of each such solution, and describe a neural net based estimator to cover the α values for which neither model is fully applicable. In section 4, we evaluate the estimators' performance.

2. THEORY

To develop our models, we first consider a simple discrete time linear frequency chirp signal with center frequency zero:

$$y(n) = \exp(j\alpha n^2). \tag{1}$$

where α is the one-half the chirp rate in radians per sample squared. (It may be shown that the results obtained for such a signal may be generalized to arbitrary center frequency and amplitude scaling.) We may write the DFT of the rectangle-windowed signal as

$$Y(k) \quad = \quad \sum_{n=-(N-1)/2}^{(N-1)/2} \exp(j(\alpha n^2 - 2\pi kn/K)). \tag{2}$$

where we implicitly apply a zero-phase window of odd sample length N, and where K is the length of the optionally zero padded transform.

2.1. Fresnel Analysis Based Model

It may be shown [6] that for sufficiently large α and N,

$$Y(k) \approx Y(\omega) = \int_{-T}^{T} \exp(j(\alpha t^2 - \omega t))dt, \tag{3}$$

where we have applied the midpoint approximation to the definite integral in the Fourier transform of the analogous continuous time chirp, $y(t) = e^{j\alpha t^2}$.

Applying this approximation, it may be shown [6] that the real and imaginary parts of $Y(k)$ are given by:

$$\Re Y(k) \approx \pm\sqrt{\frac{\pi}{2\alpha}}\Big(\cos(\phi)\int_{l_1}^{l_2}\cos\left(\frac{\pi}{2}u^2\right)du + \sin(\phi)\int_{l_1}^{l_2}\sin\left(\frac{\pi}{2}u^2\right)du\Big) \tag{4}$$

[*] Supported by the National Science Foundation

$$\Im Y(k) \approx \pm\sqrt{\frac{\pi}{2\alpha}}\Big(\cos(\phi)\int_{l_1}^{l_2}\sin\left(\frac{\pi}{2}u^2\right)du - \sin(\phi)\int_{l_1}^{l_2}\cos\left(\frac{\pi}{2}u^2\right)du\Big), \quad (5)$$

$\phi = \frac{\pi^2k^2}{K^2\alpha}$; $l_1 = \sqrt{\frac{2\alpha}{\pi}}\left(-\frac{N}{2} - \frac{\pi k}{\alpha K}\right)$; $l_2 = \sqrt{\frac{2\alpha}{\pi}}\left(\frac{N}{2} - \frac{\pi k}{\alpha K}\right)$.

We recognize the integrals in the above expressions as Fresnel integrals. When α, N, k,and K are such that $l_1 \ll -1$ and $l_2 \gg 1$, we may apply what we call the large limits approximations,

$$\int_0^{l_2}\sin\left(\frac{\pi}{2}u^2\right)du \approx \frac{1}{2} - \frac{1}{\pi l_2}\cos\left(\frac{\pi}{2}l_2^2\right) \quad (6)$$

$$\int_0^{l_2}\cos\left(\frac{\pi}{2}u^2\right)du \approx \frac{1}{2} + \frac{1}{\pi l_2}\sin\left(\frac{\pi}{2}l_2^2\right) \quad (7)$$

and where the odd symmetry of the Fresnel integrals leads to analogous negative results when the limits are $[l_1, 0]$ rather than $[0, l_2]$ as above. These approximations will allow us to create an invertible model of the signal's DFT, and incur a modeling error of less than 1% when $\{-l_1, l_2\} \geq 3$. (When $\{-l_1, l_2\} \geq 2$ or $\{-l_1, l_2\} \geq 1$ the error bounds are 2.3% or 14% respectively.)

The condition on the limits merits discussion. In an estimation context, N and K are in practice fixed and only certain small magnitude values of k will be of interest. In our estimator, we will implicitly use $|k| \leq \frac{KN\hat{\alpha}}{4\pi}$ where $\hat{\alpha}$ is at the same time estimated. We see then that as α becomes very small, no range of k will be valid. Requiring $\{-l_1, l_2\} \geq 3$ and solving for α we find a constraint of $\alpha N^2 \geq 72\pi$. (For $\{-l_1, l_2\} \geq 2.5$, we require $\alpha N^2 \geq 50\pi$.) Figure 1 should help clarify the relationship of the limits to α and k. In the top subplot, we see l_1 and l_2 plotted versus α, for each of $k = -1, 0,$ and 1, with N fixed at 641 and K fixed at 2048. The middle subplot is similar, but considers values of k from -10 to 10. We see that the larger magnitude values of k cause one of the limits to be too small for a given value of α. We also see that in general, the model is valid for smaller α when smaller k values are used. The bottom subplot is similar to the top two, but uses k values "dynamically" selected by $|k| \leq \frac{KN\hat{\alpha}}{4\pi}$ (see Fresnel estimator discussion below).

Returning to the model, it may be shown that substituting the approximations above in the present case, and applying a time domain Hann window via a frequency domain convolution [6] leads to the following expressions for the magnitude and phase of $Y(k)$, valid for k as noted above:

$$\angle Y(k) \approx \pm\frac{\pi}{4} - \frac{\pi^2k^2}{K^2\alpha} \quad (8)$$

$$|Y(k)| \approx \sqrt{\frac{\pi}{|\alpha|}}\left(\frac{1}{2} + \frac{1}{2}\cos\left(\frac{2\pi^2k}{KN\alpha}\right)\right) \quad (9)$$

where the $\frac{\pi}{4}$ term in the phase expression is positive for $\alpha > 0$ and negative for $\alpha < 0$. As we will see in the next section, these models are invertible.

2.2. Taylor Series Based Model

We must also consider the case where the above model does not apply: when α is too small for the large limits approximations to be valid. In that case, we may use the Taylor series approximation for an exponential signal with a small argument, namely:

$$y(n) = \exp(j\alpha n^2) \approx 1 + j\alpha n^2, \quad (10)$$

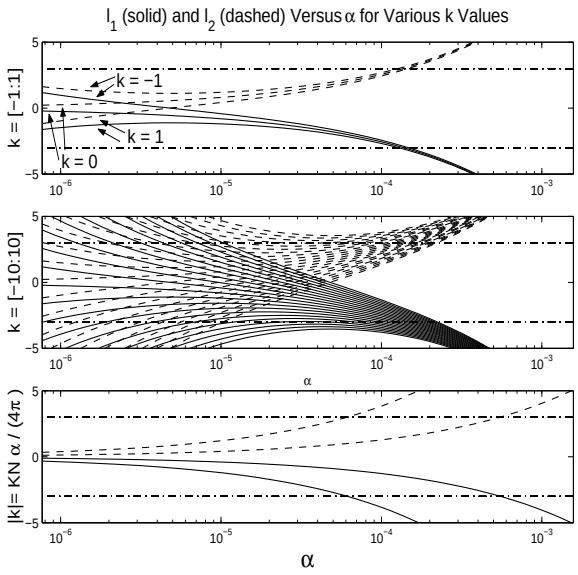

Fig. 1. The choice of k and the value of α affect the limits l_1 and l_2, which in turn dictate Fresnel model validity. The values of l_1 and l_2 giving 1% model error bounds (± 3) are shown with dash-dot lines.

where the Taylor series error introduced is $O(\alpha^2((N-1)/2)^4)$, as the argument of the exponential will be at most $\alpha(\frac{N-1}{2})^2$. Hence, the model loses accuracy as the argument value strays from zero, though we can ensure that our model introduces less than 2.1% error, for example, if we require that $\alpha(\frac{N-1}{2})^2 \leq \pi/16$. The accuracy of the model versus α will be apparent in section 4. Applying the Taylor series approximation allows us to write the DFT of the windowed chirp signal as

$$Y(k) = \sum_{n=-(N-1)/2}^{(N-1)/2}(1 + j\alpha n^2)w(n)e^{-j2\pi kn/K}. \quad (11)$$

where $w(n)$ is an arbitrary zero-phase symmetric windowing function whose DFT is $W(k)$. Applying Fourier theory and approximating frequency domain differentiation with DFT domain differencing (with large K), we may write

$$Y(k) = W(k) + i\alpha W''(k) \quad (12)$$

$$Y''(k) = W''(k) + i\alpha W''''(k). \quad (13)$$

where the number of $'$ superscripts indicates the order of differencing applied. Though the result in equation 12 may suggest an estimator in itself, it is not tolerant of phase offset in $y(n)$. An estimator robust to this may be found by analyzing the phase curvature of $Y(k)$ at $k = 0$. Specifically, it may be shown that, for zero-phase symmetric $w(n)$

$$\left.\frac{\Delta^2\angle Y(k)}{\Delta k^2}\right|_{k=0} = \frac{\Im Y_0 \Re Y_0'' - \Re Y_0 \Im Y_0''}{(\Re Y_0)^2 + (\Im Y_0)^2} \quad (14)$$

where the subscript 0 indicates $k = 0$. By substituting equations 12 and 13 into the above expression, we obtain that

$$\left.\frac{\Delta^2\angle Y(k)}{\Delta k^2}\right|_{k=0} = \frac{\alpha((W_0'')^2 - W_0W_0'''')}{W_0^2 + \alpha^2(W_0'')^2}. \quad (15)$$

For future convenience, we will denote the observed value of phase curvature at zero as p. The expression in equation 15 may be solved for α to obtain an estimator for small α.

3. MODEL APPLICATION: ESTIMATORS

We now solve the equations presented above for α to obtain two Fresnel-based and one Taylor-based estimators. Since a region may exist where neither model's validity is guaranteed, a neural net may be used to obtain a fourth estimator – a combination of the other three – to be used when the three estimators give data suggesting that α is in this transition region.

3.1. Fresnel Model Estimators

We present two estimators based on Fresnel analysis, one each from equations 8 and 9. Our magnitude-based (equation 9) estimator will rely on the fact that the cosine function in the chirp magnitude model has an argument valued at $\frac{\pi}{2}$ at the k value where the magnitude is at half the maximum height on a linear scale. Thus, we may solve for α via

$$|\alpha| \approx \frac{4\pi k_{hh}}{KN}, \tag{16}$$

where k_{hh} corresponds to the k value at half the peak's height in linear amplitude. To get very accurate k_{hh} estimates, we linearly interpolate between the appropriate nearest k values, recalling that a point of inflection exists as a cosine argument passes through $\frac{\pi}{2}$. A rough phase curvature estimate (below) may be used to determine the sign of α.

Other similar magnitude estimators relying on a pre-specified or smaller range of k (even as small as $k \in \{-1, 0, 1\}$) are indeed possible, but are less accurate both with and without noise. In applications where interference at higher (analogous) magnitude k values is likely, such estimators may prove useful, or may at least highlight such interference. This is left as an area for future exploration.

Our phase-based (equation 8) estimator ensures neutrality to phase shift in the input by analyzing the curvature of the phase via evaluation of a second order difference (again with sufficiently large K). Doing so and solving for α yields

$$\alpha \approx \frac{-2\pi^2}{K^2}\left(\frac{\Delta^2 \angle(Y(k))}{\Delta k^2}\right)^{-1}. \tag{17}$$

where we estimate the second order phase difference by averaging the second order difference values obtained when considering some $k \in k_{hh}$. The exact portion of k examined by our estimator may be chosen as a function of the desired Fresnel model accuracy. Choosing a greater range of k increases the amount of data included in the average, but strains the Fresnel approximations by requiring limits to be smaller (recall the discussion in section 2.1). It may be shown [6] that to ensure the limits in the Fresnel integrals are at least c, the range of k used must be

$$|k| \leq \left(2 - \frac{2c}{M}\sqrt{\frac{2\pi}{\hat{\alpha}}}\right) \cdot k_{hh}, \tag{18}$$

where $\hat{\alpha}$ in this case is the estimate given by the magnitude model estimator above, and preliminary experiments show that the ideal choice of the Fresnel limits parameter is around $c = 2.5$. This limits our estimator to cases where $|\alpha|N^2 \geq 50\pi$. For $|\hat{\alpha}|$ smaller than this, equation 18 will lead to a range of k entirely within the k_{hh} for a stationary sinusoid (the window transform), or to no k at all. In those cases, our estimator will use only $k \in \{-1, 0, 1\}$, generating estimates that are not accurate but that provide useful information for our neural net. The neural net is used in the transition region between our large α Fresnel model, and small α Taylor series model.

3.2. Taylor Series Model Estimator

We may cast the result in equation 15 as a simple quadratic in α, $(0 = a\alpha^2 + b\alpha + c)$, with $a = (W_0'')^2 p$, $b = (W_0 W_0'''' - (W_0'')^2)$, and $c = W_0^2 p$. We may then solve for alpha by using the familiar quadratic formula,

$$\alpha = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \tag{19}$$

where we note that only the positive sign is used in the numerator. This is because we require that as α goes to zero (as required by our Taylor series model), our estimate does so also; it may be shown that only by choosing the positive sign will this occur.

3.3. Model Transition Region: Neural Net

The constraints on the above models ensure that there is a region in which neither model is accurate. Specifically, the Fresnel model requirement that $|\alpha|N^2 \geq 50\pi$ and the Taylor model requirement that $|\alpha|(\frac{N-1}{2})^2 \leq \frac{\pi}{16}$ leave what may appear a large transition region. (In reality, however, these constraints are conservative, as can be seen in section 4.)

To obtain an estimator in the transition region, we use the outputs of each of the three estimators as the inputs to a 3-4-1 neural net with 'bias' (in neural net terms). We train the net with normalized data, taken from estimator outputs in the transition region. When choosing momentum of 0.9 for the input-to-hidden and hidden-to-output weights and a learning rate of 0.04, the net converges after approximately 1000 iterations.

4. RESULTS

In figure 2, we see the results generated by each estimator when noise is not present. In the plot, we use 31 positive α values evenly

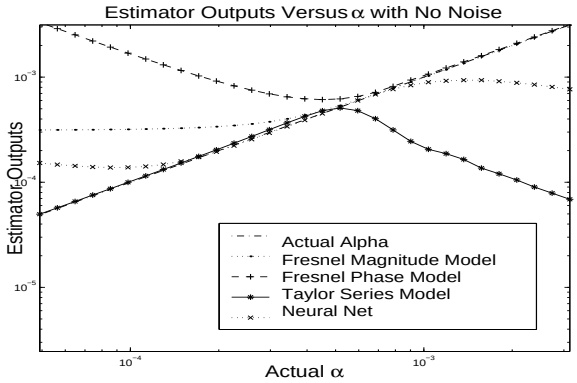

Fig. 2. Estimators without noise.

spaced on an logarithmic scale from $2\pi/F \cdot 2^{-3}$ to $2\pi/F \cdot 2^{3}$, where F is the sampling frequency (here 16000 Hz), $K = 1024$, and $N = 201$. We see how each estimator – including the neural net estimator – functions reliably only for its specific intended range of α. The figure reveals that the Fresnel and Taylor model estimators behave consistently when given data with α outside their intended regions; this will make the decision of which estimator to use for a given signal near-trivial.

We presently evaluate the performance of our estimator relative to the Cramér-Rao lower bound. To obtain this bound, we first consider the Fisher information in the signal, defined as:

$$J(\alpha) = E\left[\left(\frac{\partial \ln(f(\mathbf{y};\alpha))}{\partial \alpha}\right)^2\right] \tag{20}$$

where $f(\mathbf{y};\alpha)$ is the PDF of our chirp signal (here, we write $y(n)$ as $\mathbf{y} = [y(-(N-1)/2)...y(N-1)/2])$ in the presence of complex additive white Gaussian noise, namely:

$$f(\mathbf{y};\alpha) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-1}{2\sigma^2}\sum_n \left[(u_n - \mu_n)^2 + (v_n - \nu_n)^2\right]\right)$$

where μ_n and ν_n represent the real and imaginary parts of the signal, respectively, and u_n and v_n represent the real and imaginary parts of the signal including noise. We assume that the noise has independent real and imaginary parts each with variance σ^2. Given the above, it may be shown that

$$J(\alpha) = \frac{1}{\sigma^2}\left(\frac{N^5}{80} - \frac{N^3}{24} + \frac{7N}{240}\right), \tag{21}$$

where we note that in this case the Fisher information is independent of α. Since the Cramér-Rao lower bound on the variance of an estimator is given by $(1/J(\alpha))$, we plot this value along with the variance $\mathrm{var}(\hat{\alpha} - \alpha)$ of each set of estimates $\hat{\alpha}$ produced by each estimator. Figures 3 and 4 show plots for SNRs of 30dB and 15db, respectively. One thousand runs were used to determine the shown variance. Though the approximations invoked in creating the models used here lead to observable error variances, we note that the relative errors appear on the order of one to two percent. In applications where multiple sinusoidal signals and their parameters are detected [3], estimators such as those presented here are used as the first step in an iterative process, in general rendering such small errors inconsequential.

5. SUMMARY AND FUTURE DIRECTIONS

We demonstrated four estimators of the linear frequency chirp parameter α that may be used to enhance sinusoidal modeling. The estimators apply to Hann windowed signals, which are often found in COLA systems (as opposed to Gaussian windowed signals). The estimators yield α estimates accurate to within 2%, though the inexact nature of our models still leads to observable distance from the Cramèr-Rao lower bound.

Future work will compare the current estimators to estimators for Gaussian windowed signals, and will detail implementation of the current estimators as a single-output estimator.

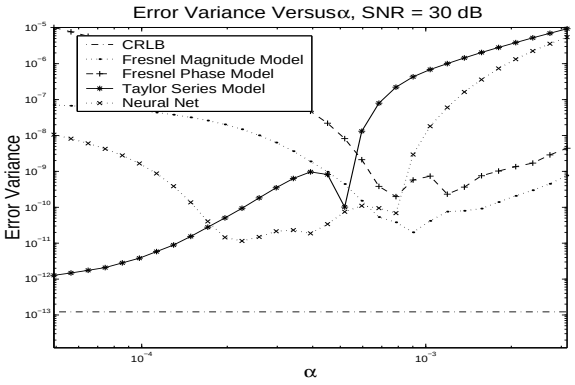

Fig. 3. Each estimator is accurate in its intended range.

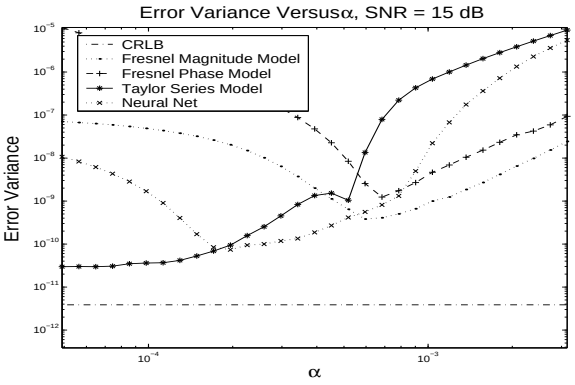

Fig. 4. Relative estimator performance is similar at lower SNRs.

6. REFERENCES

[1] Julius O. Smith and Xavier Serra, "PARSHL: A program for the analysis/synthesis of inharmonic sounds based on a sinusoidal representation," in *Int Computer Music Conf.* 1987, Computer Music Association, Also available as Stanford Music Department Technical Report STAN–M–43.

[2] R. J. McAulay and T. F. Quatieri, "Speech analysis-synthesis based on a sinusoidal representation," Tech. Rep. 693, Lincoln Laboratory, MIT, 1985.

[3] J. S. Marques and L. B. Almeida, "A background for sinusoid based representation of voiced speech," in *ICASSP*, 1986.

[4] Julius O. Smith III, "Mus420/ee367a supplemental lecture: Gaussian windows and transforms," Available from http://www-ccrma.stanford.edu/ jos/Gauss/, 2002.

[5] Aaron S. Master and Yi-Wen Liu, "Nonstationary sinusoidal modeling with efficient estimation of linear frequency chirp parameters," Submitted to ICASSP, 2003.

[6] Aaron S. Master, "Nonstationary sinusoidal model frequency parameter estimation via fresnel integral analysis," Tech. Rep., CCRMA, Stanford University, 2002, Available from http://www-ccrma.stanford.edu/˜asmaster/.