NONSTATIONARY SINUSOIDAL MODELING WITH EFFICIENT ESTIMATION OF LINEAR FREQUENCY CHIRP PARAMETERS

Aaron S. Master and Yi-Wen Liu*

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

ABSTRACT

We present an enhanced sinusoidal modeling system that efficiently parameterizes spectral peaks by linear frequency chirp rate, in addition to the standard parameters of amplitude, center frequency, and phase. Similar to a conventional sinusoidal modeling systems, the current system operates in a frame-by-frame manner, but also efficiently obtains a chirp parameter estimate for each peak in a given frame. We demonstrate an application to voice coding where our model captures most of the significant signal information in a spectrogram operating at double the current system's frame rate.

1. INTRODUCTION

In the field of audio signal processing, sinusoidal modeling is a fundamentally important signal representation. It finds applications in audio coding, speech enhancement, sound source separation, and pitch tracking. All these applications benefit from high quality sinusoidal parameters.

In conventional sinusoidal modeling [1, 2], these parameters are defined as frequency, amplitude and phase of each frequency component. Other efforts have focused on non-sinusoidal parameters such as bandlimited noise [3] or transients [4].

Efforts have also been made to expand or improve sinusoidal parameters. Matching-pursuit [5] seeks to identify exponentially decaying envelopes of quasistationary sinusoids. Linear frequency chirp rates of frequency components have been estimated [6], when constant- or Gaussian-amplitude time envelopes are present. A dynamic vocoder model [7] has been used to estimate very accurate sinusoidal trajectories within each frame, but at great computational expense.

We present a new technique that efficiently and robustly obtains the linear chirp parameter. The current estimation technique is robust to phase shift, and to time domain amplitude modulation.

*Supported by the National Science Foundation

2. THEORY

To develop our model, we first consider a simple discrete time linear frequency chirp signal:

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

where α is the one-half the chirp rate in radians per sample. We may write the DFT of the rectangle-windowed signal as

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

where N is the odd-sample zero phase window length and K is the length of the optionally zero padded transform. It may be shown [8] that for sufficiently large α and N,

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

where we have applied the midpoint approximation to the definite integral of the analogous continuous time chirp,

$$y(t) = e^{j\alpha t^2}.$$

Considering the above integral, it may be shown [9] that

$$\frac{dY_W^{\mathrm{Rec}}(\omega)}{d\omega} = \frac{-1}{2\alpha}\big(- e^{j\alpha T^2} 2j \sin \omega T + j\omega Y_W^{\mathrm{Rec}}(\omega)\big)$$

and we see by inspection that this expression becomes zero when $\omega = 0$, indicating a stationary point at the center frequency. When using a Hann window, it may be shown [9] that:

$$\begin{aligned}\frac{dY_W^{\mathrm{Hann}}(\omega)}{d\omega} &= \frac{dY_W^{\mathrm{Rec}}(\omega)}{d\omega} + \frac{1}{2}\frac{dY_W^{\mathrm{Rec}}(\omega - \omega_T)}{d\omega} \\ &\quad + \frac{1}{2}\frac{dY_W^{\mathrm{Rec}}(\omega + \omega_T)}{d\omega} \\ &= \frac{-1}{2\alpha} j\omega Y_W^{\mathrm{Hann}}(\omega)\end{aligned}$$

and the second derivative evaluated at $\omega = 0$ becomes

$$\left.\frac{d^2 Y_W^{\text{Hann}}(\omega)}{d\omega^2}\right|_{\omega=0} = \frac{-j}{2\alpha}(Y_W^{\text{Hann}}(0)). \tag{4}$$

By choosing K sufficiently large, we may approximate the second order derivative with a second order difference. Doing so and solving for α yields

$$\hat{\alpha} \approx \frac{-jY^{\text{Hann}}(0)}{2}\left(\left(\frac{K}{2\pi}\right)^2 \left.\frac{\Delta^2 Y^{\text{Hann}}(k)}{\Delta k^2}\right|_{k=0}\right)^{-1}. \tag{5}$$

where we note that second order differencing operation with respect to frequency bin k has been normalized by twice multiplying by $\frac{K}{2\pi}$. This expression will serve as the main part of the chirp parameter estimator.

3. MODEL APPLICATION

We now may consider a more general and practical discrete time signal model for a given spectral peak:

$$g(n) = \exp(j(\alpha n^2 + \beta_0 n + \phi)), \tag{6}$$

where α is one-half the chirp rate in radians (increased) per sample, β_0 is the center frequency in radians per sample, and ϕ is the phase offset. The DFT of $g(n)$ is $G(k)$.

To estimate α from $G(k)$, it may be shown [8] that we may use our estimator in eqn. 5, provided that we consider k corresponding to β_0 rather than $k = 0$. It may also be shown [9] that the estimator is valid when $\phi \neq 0$.

We may also use the estimator when P order polynomial amplitude modulation occurs in the signal. To prove that this is valid, we first recall that taking an order 3 or higher derivative (approximated by differencing) of $Y(k)$ yields zero. Now, recalling that order P polynomial amplitude modulation in the time domain corresponds to order P differentiation (approximated by differencing) and multiplication by j in the frequency domain, we see that such AM cannot influence our estimator.

Although our estimator survives these conditions, it cannot function normally when α or N is very small. (Because we will treat N as the fixed frame length of the system, we may view this as a constraint on α.) The error occurs because we cannot achieve the spectral resolution needed to accurately estimate the very large second order derivative in eqn. 4 with a second order difference, barring a $1/\alpha$ increase in K. The error manifests itself in a predictable way, however, with the estimated α value from eqn. 5 showing an imaginary part when in fact the estimate should be purely real. This imaginary part mimics the ratio of the real part to the correct α value, allowing us to solve the equation

$$x_1 + x_2 \Im\{\hat{\alpha}\} \approx \frac{\Re\{\hat{\alpha}\}}{\alpha} \tag{7}$$

as a least squares problem to obtain optimal coefficients x_1 and x_2. We may then substitute these and the real and imaginary parts of the eqn. 5 estimate and solve for α in cases where the initial estimator's guess is below our smallness threshold of 0.08 radians per N-sample window. (This was found to be reasonable when using zero padding to $K \leq 5N$.) Doing so creates a small-α estimator that is as accurate for small α as eqn. 5 is for larger α. Results showing the estimator applied to a range of α values are shown in figure 1. In that figure, $\beta_0 = 0$, $N = 201$ and $K = 1024$. An arbitrary affine amplitude modulation of $0.5 + 0.001n$ was applied as was a phase shift of $\phi = \frac{\pi}{16}$.

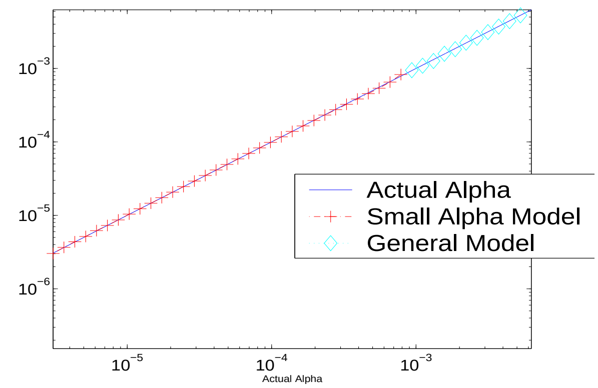

Fig. 1. Results achieved by the small-α and general estimator.

4. SYSTEM OPERATION

The overall structure of our system is similar to conventional sinusoidal modeling. First, each frame is read in to a buffer, which is FFT'd and then analyzed for peaks. A parabola is fit to the log of each magnitude peak, yielding accurate magnitude and frequency estimates. A phase estimate is also recorded, and is normalized by the chirp parameter after the latter is estimated. (This is necessary because of the nonlinear mapping of phase offsets to detected peak phase for chirp signals.)

Next, we apply the estimator in eqns. 5 and 7 as appropriate to each peak, assigning a chirp parameter to each. When calculating the second order difference in eqn. 5, we use the $Y(k)$ values at the nearest frequency bin k to the detected peak (and its neighbors $k \pm 1$) in the calculation. In the future, we may try to use linear or other interpolation of $Y(k)$ to use values corresponding more exactly to the interpolated peak center frequency of the component.

We leave the task of frequency trajectory alignment across frames to future work. We note that since our chirp and phase parameters are detected with great accuracy, it may no longer be necessary to align such trajectories. Direct synthesis of the detected frequency components via IFFT or a bank of oscillators will be explored as an alternative to trajectory alignment.

5. RESULTS

We provide two examples of the application of our system. For each example, we provide both a spectrogram, which hints at the capabilities of a conventional SMS system, and trajectory plot. The trajectory plot includes the center frequency for each detected peak (shown as a bold dot) and a line showing the linear frequency trajectory estimated by our chirp detector. Because 50% overlap of windows was used in our system, each linear trajectory is extended only half way to the adjacent frame. Hence, the trajectories should meet up between frames if our estimator is accurate. Indeed, this can in general be seen.

We see that most of the information gleaned from the spectrogram can be observed in our trajectory plot, and we recall that our system is operating at one half the frame rate of the spectrogram. Hence, we may consider our system an improvement to conventional SMS in that we have both reduced the frame rate by a factor of 2 and captured the essential features of the signal.

In the first example figures, we see the female speech utterance "you always" and in in the second example figures the female speech utterance "example" [10].

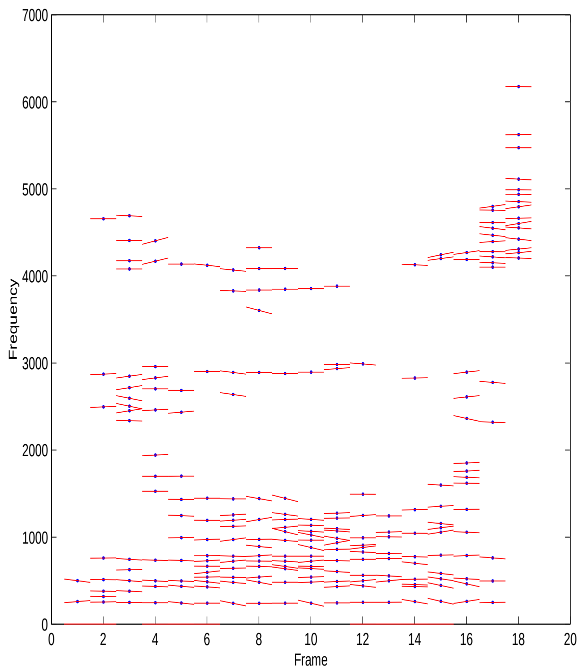

Fig. 2. "You always" Speech Trajectory Plot.

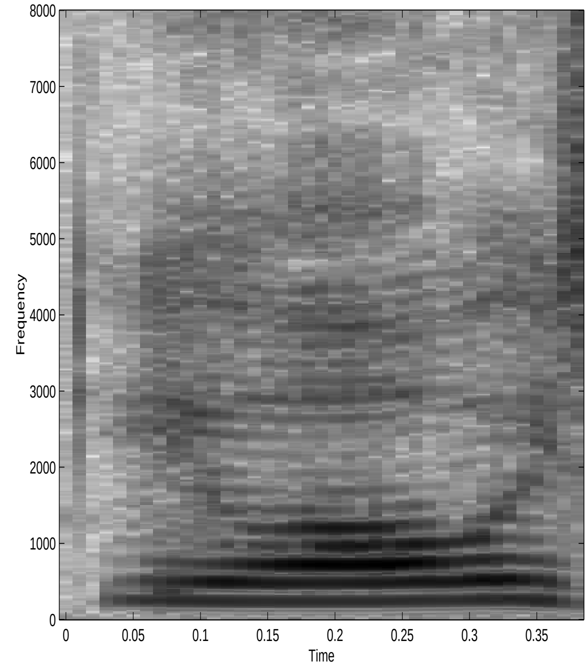

Fig. 3. "You always" Speech Spectrogram.

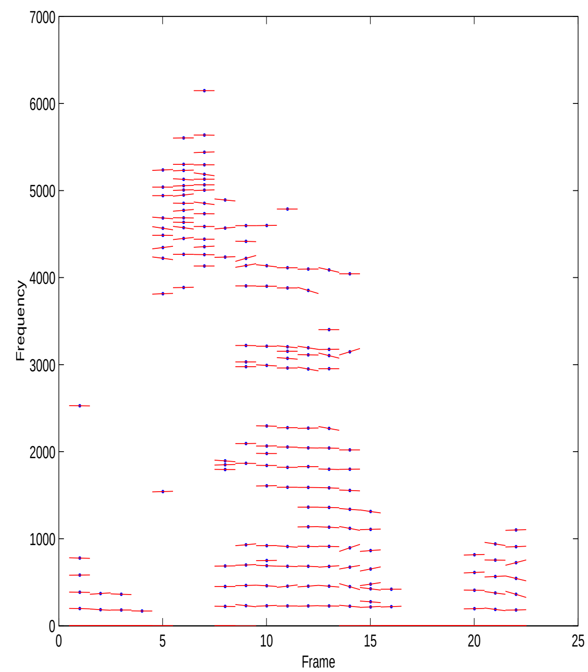

Fig. 4. "Example" Speech Trajectory Plot.

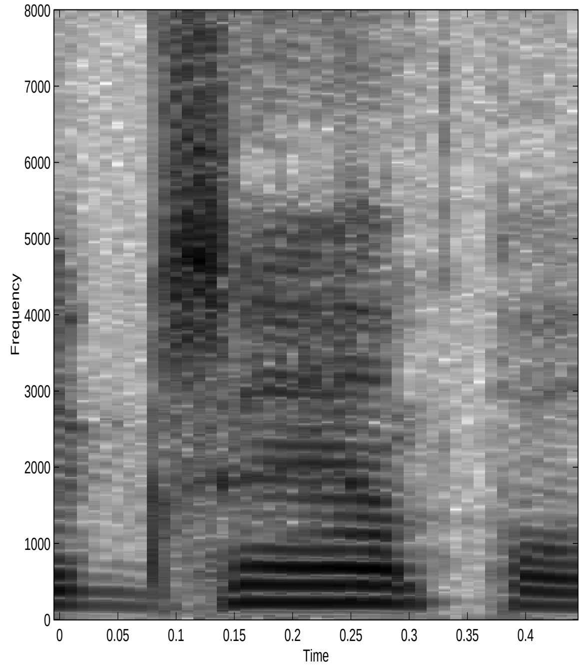

Fig. 5. "Example" Speech Spectrogram.

6. SUMMARY AND FUTURE DIRECTIONS

We have presented a system that obtains linear frequency chirp parameters efficiently and accurately. The system is significant because it reduces the frame rate needed to capture the essential features of input signals, while incurring little computational expense at the frame level.

Future work will focus on making the chirp parameter estimation more consistent. We will attempt spectral interpolation for the derivative estimates in eqn. 5 and will investigate combining estimates from multiple chirp estimators into a more robust estimate.

The task of intra-frame trajectory matching is yet to be addressed. This could provide for a significant area of exploration, with trajectory matching incorporating a confidence measure in the chirp parameter or a history of a given trajectory. Advances in this area will also have a significant impact on the implementation of chirps in synthesis.

7. 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] Xavier Serra and Julius O. Smith, "Spectral modeling synthesis: A sound analysis/synthesis system based on a deterministic plus stochastic decomposition," *Computer Music Journal*, vol. 14, no. 4, pp. 12–24, Winter 1990.

[4] Scott Nathan Levine, *Audio Representations for Data Compression and Compressed Domain Processing*, Ph.D. thesis, Electrical Engineering Dept., Stanford University (CCRMA), December 1998.

[5] Michael Goodwin, "Matching pursuit with damped sinusoids," in *ICASSP*, 1997.

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

[7] Aaron S. Master, "Sinusoidal modeling parameter estimation via a dynamic channel vocoder model," in *ICASSP*, 2001.

[8] 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/.

[9] Yi-Wen Liu and Aaron S. Master, "Phase of a continuous time linear-frequency chirp signal: Analysis and application," Tech. Rep., CCRMA, Stanford University, 2002, Available from http://www-ccrma.stanford.edu/˜asmaster/.

[10] E. Wan, A. Nelson, and Rick Peterson, *Speech Enhancement Assessment Resource (SpEAR) Database*, Oregon Graduate Institute of Science and Technology CSLU, Beta Release v1.0, avaliable online at http://ee.ogi.edu/NSEL/.