ADDITIVE ANALYSIS/SYNTHESIS USING ANALYTICALLY DERIVED WINDOWS

Stefan Borum

Br el & Kj¾
Skodsborgvej 307
DK-2850 N¾um
Denmark
sborum@bk.dk

Kristoffer Jensen

Copenhagen University
Department of Computer Science
Universitetsparken 1,
DK-2100 Copenhagen fl, Denmark
krist@diku.dk

ABSTRACT

This paper presents a simple and easy way to obtain good estimates of the frequencies of spectral peaks in voiced sounds. The instantaneous frequency throughout the spectrum is calculated using an analytically derived window. Peaks are detected by examining the difference between the frequencies of spectral lines and their corresponding instantaneous frequencies. The instantaneous frequencies near peaks are used as estimates of their actual frequencies. The corresponding amplitude values are calculated from the Gabor transform at the estimated frequencies.

The performance of the estimation of the frequencies using analytically derived windows has been verified using synthetic signals and musical sounds. This method gives a better estimate of the instantaneous frequency than commonly used methods and the resynthesis of sound from the additive parameters gives good results.

1. INTRODUCTION

This work presents a method of finding and estimating the frequencies and amplitudes of the sinusoidal components of musical signal, so called additive analysis. In additive analysis, frequencies are generally estimated by interpolating peaks [1], or by time difference. In this work, the frequencies are found by calculating the time derivative of the phase in one sample. This method gives numerically more stable results, while keeping the same order of complexity as commonly used methods.

The results obtained here are part of ongoing research [2], [3], in time-frequency analysis, and particularly the methods presented in [4].

This paper first presents the signal model used, then the instantaneous frequency calculation using analytically derived windows is presented. Some results and numerical comparisons with peak interpolation and time difference are presented, and finally there is a conclusion.

2. SIGNAL MODEL

The model of the sounds analyzed in this paper is

$$s(t) = \sum_{k=1}^{N} a_k(t) \cdot \sin(2\pi \int_{\tau=0}^{t} f_k(\tau) d\tau) \qquad (1)$$

where a_k and f_k are the time-varying amplitudes and frequencies of N partials. The equation (1) is both an analysis and a synthesis model.

The sound is thus the sum of a number of sinusoidals with time-varying amplitude and frequency. The problem stated is the estimation of the a_k and f_k of the sinusoidals.

3. ANALYTIC FREQUENCY CALCULATION

In order to calculate the instantaneous frequency over the time-frequency plane we define the continuous Gabor transform G_w as follows [4], [5],

$$G_w(s)(t,f) = \int s(\tau)\overline{w}(\tau - t)e^{-i2\pi f(\tau - t)} d\tau = \langle s, g_{t,f} \rangle \qquad (2)$$

- where s is the signal to be analyzed, w is a windowing function, t means time, f means frequency, <Æ,Æ> denotes the inner product, and the so-called Gabor functions $g_{t,f}$ are defined by

$$g_{t,f}(x) = w(x-t)e^{i2\pi f(x-t)} \qquad (3)$$

In the following formulas s is omitted for clarity. G_w is a complex-valued function. It thus defines a non-negative real amplitude function A_w and a real phase function φ_w over the time-frequency plane (at time-frequency coordinates where $A_w(t,f) \neq 0$):

$$G_w(t,f) = A_w(t,f)e^{i\varphi_w(t,f)} \qquad (4)$$

G_w is a phase modulated variant of the short-time Fourier transform $STFT_w$ [6],

$$G_w(t,f) = e^{i2\pi ft} STFT_w(t,f) \qquad (5)$$

This phase modulation gives the Gabor transform the following desirable property: Given a fixed frequency that is high enough to make the Gabor function effectively analytic (i.e. without negative frequencies), the Gabor transform will filter an analytic signal from the original signal using the filtering properties of the Gabor function.
For time-frequency coordinates (*t,f*) where the frequency meets the above requirement we define the instantaneous frequency f_i by

$$f_{i,w}(t,f) = \frac{1}{2\pi}\frac{\partial \varphi_w(t,f)}{\partial t} \qquad (6)$$

Applying the logarithm to (4) and differentiating leads to

$$\frac{\partial \varphi_w(t,f)}{\partial t} = \frac{\partial\,\mathrm{Im}(\ln(G_w(t,f)))}{\partial t} = \mathrm{Im}\left[\frac{\frac{\partial G_w(t,f)}{\partial t}}{G_w(t,f)}\right] \qquad (7)$$

By combining (2), (6), and (7) we obtain

$$f_{i,w}(t,f) = f - \mathrm{Im}\left[\frac{G_{w'}(t,f)}{G_w(t,f)}\right]\Big/ 2\pi \qquad (8)$$

- where *w'* is the derivative of *w*.
From (8) we can calculate the instantaneous frequency f_i without using numerical differentiation. A natural utilization of this formula is when extracting additive components from multicomponent signals according to (1). Such components generate ridges in the spectrogram, as far as the analyzing window is able to separate them. It can be shown that on ridges the instantaneous frequency is equal to the frequency of the ridge [4]. Furthermore, in the neighborhood of a ridge the instantaneous frequency is greatly influenced by the ridge. It is therefore possible to obtain a good estimate of the frequency of the ridge from the instantaneous frequency in its neighborhood. By restricting the Gabor transform to a fixed time we can use the instantaneous frequency to estimate the frequency of spectral peaks.
It should be noted that this method can be applied to all differentiable windowing functions, including frequency modulated analytic windows (chirps).
Figure 1 shows the estimated frequencies of the spectral peaks of a frequency detail from a soprano voice. The horizontal steps of the lower subplot indicate frequency ranges where the instantaneous frequency is governed by the spectral peaks. The horizontal crossings with the diagonal line determine the frequency estimation of the peaks.
Because of the constant instantaneous frequency in the vicinity of the spectral peaks, the frequency estimation can be calculated with small error, from a single frequency line, as long as the analysis frequency is close. Interpolation algorithms can nonetheless easily be adapted, as well as iterative search algorithms [4].
Linear interpolation has been used in figure 1, whereas no interpolation has been used in the analytic derivation method in the experiments in the next section.

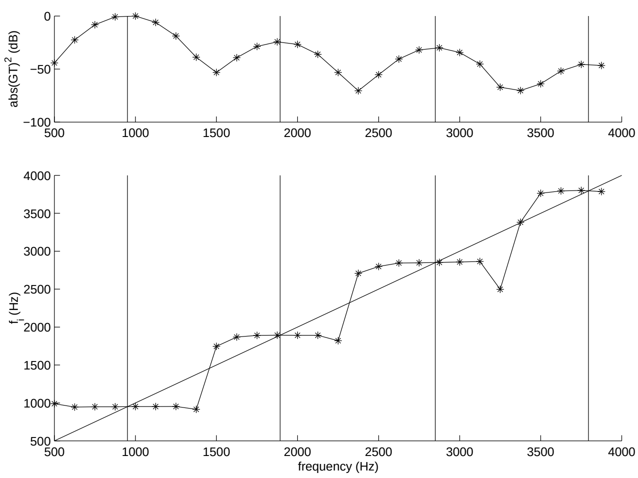

Figure 1. *Top: Section of power spectrum of a soprano voice. Bottom: f_i vs. analysis frequency. The diagonal line is where the instantaneous frequency equals the analysis frequency. In both plots the vertical lines show the estimated frequencies of the spectral peaks.*

4. EXPERIMENTS

In order to test the accuracy of the frequency estimation of the partials of a signal according to (1), three test signals have been constructed. All three signals have eight harmonic partials, each having the same amplitude envelope; zero for 1/8th of a second, then a linear slope from 1800 to 600 for 3/4th of a second, and finally silence for 1/8th of a second. The fundamental frequencies of the three test signals are 30, 100 and 300 Hz. Each test signal has been analyzed using three different methods: peak-interpolation using gauss interpolation, analytic derivation, and time difference of the phase of two $STFT_w$'s one sample apart. All methods use the hanning window, and the window length is 2.8 times the period of the signal. This window length has been found in this work to be the shortest possible with adequate frequency separation of harmonic signals, although [7] uses a window length of 2.5 times the period. The sampling rate is 32000 Hz. All three methods use the same analysis frequencies to determine the frequency and amplitude at each time step. These analysis frequencies are the harmonic components with the fundamental frequency determined algorithmically using an autocorrelation-based pitch tracker [8]. The analysis frequencies are generally worse than the estimated frequencies of all three methods.
The resulting amplitude and frequency tracks can be seen in figure 2 for the fundamental of the 300 Hz test signal. The peak-interpolation results are shown top/left, the analytic derivation results are shown in the middle and the time difference results are shown bottom/right. The time and frequency axes have been shifted in order to improve visibility.
The peak interpolation and time difference methods both determine the amplitude as the maximum of the gauss

interpolation, whereas the analytic derivation method uses the absolute value of the Gabor transform at the estimated frequency.

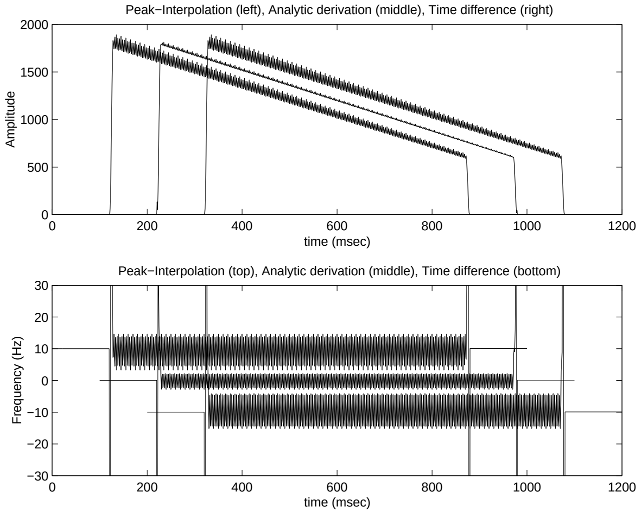

Figure 2. *Amplitude estimation (top) and frequency deviation (bottom) of the fundamental of a 300 Hz test signal. Peak interpolation (left/top), analytic derivation (middle) and time difference (right/bottom).*

The mean errors for the eight partials are shown in figure 3 for the amplitude in dB (top) and the frequency in cents (bottom). The peak interpolation errors are shown left, the analytic derivation errors are shown middle and the time difference errors are shown right.

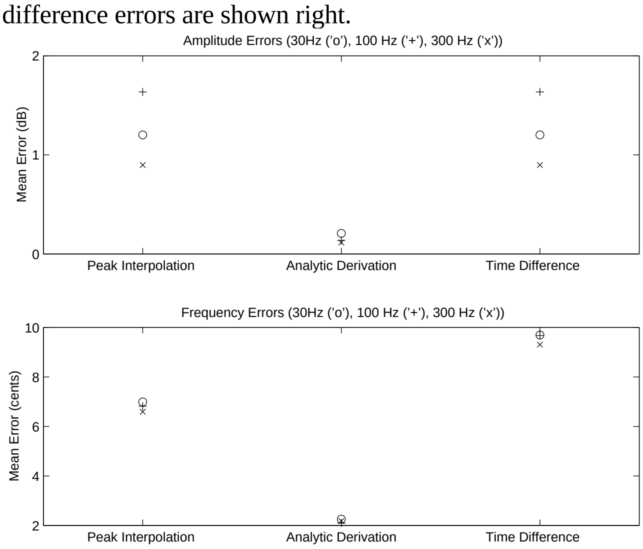

Figure 3. *Amplitude error in dB (top) and frequency error in cents (bottom) for peak interpolation (left), analytic derivation (middle) and time difference (right).*

The analytic derivation results are clearly better than the peak interpolation and the time difference results. The amplitude estimations are identical for the peak interpolation and the time difference analysis. The time resolution is slightly better for the analytic derivation analysis than for the other analysis methods, and it is in the order of two thirds of the period of the fundamental.

Informal listening tests have been performed on a large number of sounds from many instruments, including the piano and the flute. The result of these listening tests are that the sounds analyzed with the peak-interpolation and the time difference methods generally have comparable quality, whereas the sounds analyzed with the analytic derivation method have equal or superior quality. The increase in quality is especially noticeable on high-pitched sounds.

5. CONCLUSIONS

This paper has presented the estimation of frequencies of voiced sounds using analytic derivation. A formula for calculating instantaneous frequencies is derived and it is shown how the instantaneous frequencies can be used to find spectral peaks. Experiments are presented that evaluate the accuracy of the frequency estimation as compared to two commonly used methods: peak interpolation and time difference. The results of these experiments confirm that frequency estimation using analytic derivation is significantly better than the reference methods.

The analytic method is being used by the authors in ongoing research regarding the separation of additive components of musical sounds. The superiority of the method is particularly pronounced when analyzing difficult signals.

6. REFERENCES

[1] Serra, X. and J. Smith III, ÒSpectral Modeling Synthesis: A Sound Analysis/Synthesis System Based on a Deterministic plus Stochastic DecompositionÓ *Computer Music Journal*, 14(4):12-24, 1990.

[2] Borum, S., ÒFrekvensbestemmelse af partialtoner i musikalske lydeÓ Master thesis. Department of Computer Science, University of Copenhagen, 1999.

[3] Jensen, K., Ò*Timbre Models of Musical Sounds*Ó Ph.D. dissertation, Department of Computer Science, University of Copenhagen, DIKU Report N° 99/7, 1999.

[4] Guillemain, P. and R. Kronland-Martinet, ÒCharacterization of Acoustic Signals Through Continuous Linear Time-Frequency RepresentationsÓ *Proceedings of the IEEE*, 84(4):561-585, 1996.

[5] Gabor, D. ,ÓTheory of CommunicationÓ, *The Journal of the Institution of Electrical Engineers*, 93(III):429-457, 1946.

[6] Cohen, L., Ò*Time-Frequency Analysis*Ó Prentice Hall, NJ, 1995.

[7] R. J. McAulay, T. F. Quatieri, ÒSpeech analysis/synthesis based on a sinusoidal representationÓ, *IEEE Trans. on Acoustics, Speech and Signal Proc.*, vol. ASSP-34, No. 4, August 1986.

[8] L. R. Rabiner, ÒOn the use of autocorrelation analysis for pitch detectionÓ, *IEEE Trans. Acoustics, Speech and signal processing*. Vol. ASSP-25, No 1, February 1977.