

PARAMETER ESTIMATION AND TRACKING FOR TIME-VARYING SINUSOIDS

Heiko Purnhagen*

Laboratorium für Informationstechnologie, University of Hannover
Schneiderberg 32, 30167 Hannover, Germany
purnhage@tnt.uni-hannover.de

ABSTRACT

Parametric modeling permits an efficient representation of audio signals and is increasingly utilized for very low bit rate coding applications. Such systems are based on a decomposition of the audio signal into components that are described by appropriate source models and represented by model parameters. Commonly used components types are sinusoidal trajectories, harmonic tones, transients, and noise. Proper estimation and tracking of sinusoids in case of vibrato or portamento is vital, yet difficult due to the uncertainty principle for time-frequency resolution. This paper presents a reliable solution to this problem that was successfully demonstrated in an encoder for MPEG-4 parametric audio coding “Harmonic and Individual Lines plus Noise” (HILN).

1. INTRODUCTION

Parametric modeling of audio signals has long been used for analysis/synthesis of speech and music signals. It allows a compact signal representation as well as flexible and “meaningful” modification of the signal in the parametric domain. While the first property is, for example, exploited by very low bit rate speech coding systems, the second property is frequently utilized in computer music research.

In general, parametric representations are based on a decomposition of the input signal into components that are described by appropriate source models and represented by model parameters. Sinusoidal modeling, i.e. a parametric representation that utilizes only sinusoidal components [1], is very popular because probably most real-world audio signals are dominated by tonal signal components. Equation 1 shows how the input signal $x(t)$ is approximated by a set of I sinusoidal trajectories i with slowly varying parameters for amplitude $a_i(t)$ and frequency $f_i(t)$ and a start phase φ_i .

$$\hat{x}(t) = \sum_{i=1}^I a_i(t) \sin \left(\varphi_i + 2\pi \int_0^t f_i(\tau) d\tau \right) \quad (1)$$

However, noise-like and transient signal components can not be efficiently represented by sinusoidal modeling. Hence, a sinusoidal model is often combined with additional signal models for noise and transients. Overviews of such hybrid models can be found e.g. in [2], [3].

The time-discrete representation of the signal $x(n)$ is denoted

$$x[n] = x((n+1/2)(1/f_s)), \quad (2)$$

* The author is now with Coding Technologies Sweden AB, Stockholm, Sweden.

where f_s is the sampling frequency. Because the parameters for frequency $f_i(t)$ and amplitude $a_i(t)$ of a sinusoidal trajectory i may vary slowly over time, they are sampled at regular intervals $T = N/f_s$ and interpolated during synthesis. This corresponds to frame length (i.e., hop size) of N input signal samples. The parameters for frame l are sampled at the frame center $t_l = (l+1/2)T$.

For the experiments reported in this paper, the MPEG-4 parametric audio coder HILN (“Harmonic and Individual Lines plus Noise”) was used as a framework [4], [5].

In the HILN encoder, the input signal is decomposed into different signal components and then the model parameters for the components are estimated: *Individual sinusoids* are described by their frequencies and amplitudes, a *harmonic tone* is described by its fundamental frequency, amplitude, and the spectral envelope of its partials, and a *noise* signal is described by its amplitude and spectral envelope. The modeling of transient components is improved by optional parameters describing their temporal amplitude envelope. A perceptual model is used to select the perceptually most relevant signal components in order to comply with bit rate constraints. Finally, the component parameters are quantized, coded, and multiplexed to form a bit stream. The target bit rate range of HILN is approximately 6 to 16 kbit/s, and typically an audio bandwidth of 8 kHz (i.e., $f_s = 16$ kHz) and a frame length of $T = 32$ ms (i.e., $N = 512$) are used.

In the HILN decoder, the parameters of the components are decoded and then the component signals are resynthesized according to the transmitted parameters. By combining these signals, the output signal of the HILN decoder is obtained. Sinusoidal components continued from the previous frame (i.e. that are part of a longer trajectory) are synthesized using interpolation of frequency and amplitude parameters to avoid phase discontinuities. For new (“born”) sinusoids, usually the start phase parameters are not transmitted and random values are used instead.

Proper estimation and tracking of sinusoids in case of vibrato or portamento is vital, yet difficult due to the uncertainty principle for time-frequency resolution. This paper describes techniques for signal decomposition and accurate estimation of frequency, sweep rate, amplitude, and phase of sinusoidal signal components (Section 2). Based on these techniques, a new approach for reliable tracking of sinusoids with time-varying parameters is introduced (Section 3). The performance of the old techniques and the new approach is compared for synthetic test signals and also in the context of a complete parametric audio coding system (Section 4). The paper ends with an outlook and conclusions (Section 5).

2. SIGNAL DECOMPOSITION AND PARAMETER ESTIMATION FOR SINUSOIDS

Accurate and robust parameter estimation for signal components is important in order to allow perceptually equivalent resynthesis of components and to facilitate analysis/synthesis-based signal decomposition.

Traditionally, component parameters are estimated once per parameter sampling interval and the estimation is based on a signal segment obtained using a temporal window centered at the parameter sampling point. The windows for consecutive segments typically overlap by 50%.

2.1. Signal Decomposition

Sinusoidal components permit subtractive signal decomposition. This enables an iterative analysis/synthesis approach. Starting from the input signal segment $r_0(t)$, in each step i of the iteration a dominant sinusoidal component in the current residual is extracted. The extraction is realized by estimating the component parameters and then resynthesizing and subtracting the component to calculate a new residual $r_i(t)$. If all sinusoidal components have been extracted properly, only noise-like components are left over in the residual $r_i(t)$. This assumes that transients are modeled sufficiently well by time-varying sinusoids.

2.2. Estimation of Constant Frequency

For the simple case of a single sinusoid with constant frequency in white Gaussian noise, the maximum likelihood estimator is basically given by the location of the maximum in the periodogram, i.e., the squared magnitude of the Fourier transform [6].

$$\hat{f} = \arg \max_f \left| \frac{1}{N} \sum_{n=0}^{N-1} x[n] e^{-j2\pi \frac{f}{f_s} n} \right|^2 \quad (3)$$

Once a frequency estimate \hat{f} is available, amplitude \hat{a} and phase $\hat{\phi}$ can be found by correlation with a complex sinusoid having the estimated frequency.

In order to reduce interference from neighboring sinusoidal components and to meet the assumption of a single sinusoid in noise, a bandpass filter with a passband centered at an initial estimate \hat{f}'_i of the sinusoid's frequency has been introduced. The initial estimate, usually the location of a peak in the discrete Fourier transform of the signal segment, is provided by the decomposition algorithm. In combination with the above mentioned temporal window, estimation thus evaluates only a section of the time-frequency plane (Figure 1(a)). The filter bandwidth has been chosen carefully to achieve a good trade-off between maximum attenuation of neighboring components and minimum attenuation of the sinusoid to be estimated, which may be located off-center due to the error of the initial estimate. For the typical effective segment length of $\Delta_t = T = 32$ ms employed in HILN, a bandwidth of $\Delta_f = 32$ Hz is used, which is the minimum bandwidth according to time-frequency uncertainty $\Delta_t \Delta_f \geq 1$.

To implement bandpass filtering and frequency estimation, the signal is subjected to a frequency shift corresponding to the initial estimate, followed by a lowpass filter [7], typically an FIR filter (Figure 2). Frequency shifting is performed with help of a complex heterodyne signal

$$x_{\text{het},i}[n] = e^{-j2\pi \frac{f_{\text{het},i}}{f_s} n} \quad (4)$$

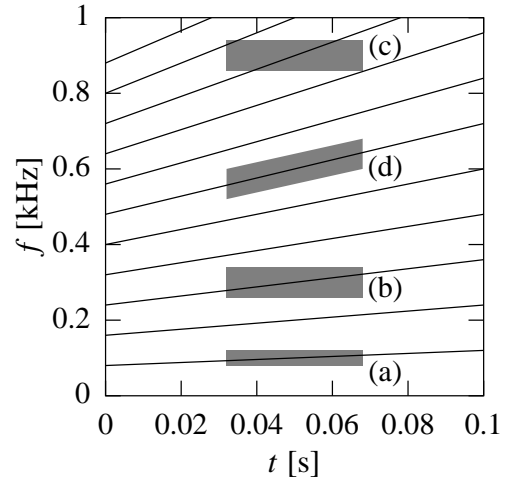


Figure 1: Rectangular (a), (b), (c) and slanted (d) sections of the time-frequency plane for the partials of a harmonic tone with f_0 sweeping from 80 Hz to 120 Hz in 100 ms.

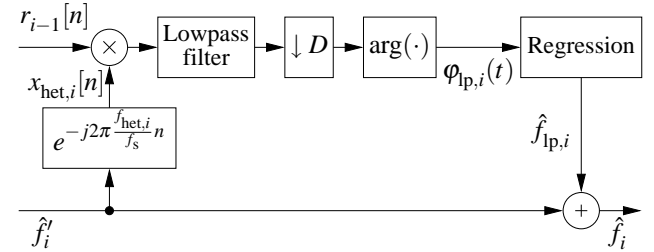


Figure 2: High accuracy estimator for frequency \hat{f}_i based on the initial estimate \hat{f}'_i .

having the initially estimated frequency $f_{\text{het},i} = \hat{f}'_i$. The frequency $\hat{f}_{\text{lp},i}$ of resulting complex lowpass signal is estimated accurately as the slope of a linear approximation of its phase $\phi_{\text{lp},i}(t)$ over time [8] using linear regression. Due to the limited bandwidth of the lowpass signal, it is typically resampled at $f_{\text{s,lp}} = 500$ Hz in order to reduce the computational complexity of lowpass filtering and regression. Finally, the initial estimate \hat{f}'_i is added to obtain an accurate frequency estimate \hat{f}_i for the original signal.

2.3. Estimation of Frequency and Sweep Rate

To accommodate for the time-varying frequency of a sinusoidal trajectory in case of vibrato or portamento, the heterodyne-based frequency estimator has been extended to permit also estimation of the sweep rate of linearly changing frequencies [7]. For this purpose, the bandpass filter bandwidth had to be increased to 64 Hz to cover the frequency range traversed during the duration of the signal segment (Figure 1(b)). Furthermore, linear regression now finds the parameters of a parabolic approximation of $\phi_{\text{lp},i}(t)$, hence estimating the frequency and sweep rate of a sinusoid with instantaneous frequency changing linearly over time.

3. TRACKING OF SINUSOIDS

3.1. Tracking by Parameter Matching

High sweep rates, as e.g. observed for the higher partials of a singing voice, can not be recovered this way because of increasing interference from neighboring partials (Figure 1(c)). To address this problem, algorithms for building sinusoidal trajectories from a time-series of frequency and amplitude parameter estimates have to be considered. A simple trajectory-building approach is based on finding the best matches between frequency and amplitude parameters of the sinusoids estimated independently in consecutive segments. Because of the mentioned problems, the results are not reliable in case of high sweep rates.

3.2. Tracking Sweep Estimation

More reliable results can be obtained if the signal of a sinusoidal component is actually tracked between consecutive segments. Hence, the sweep estimator has been extended to take into account the frequency and phase parameters f_{pre} and ϕ_{pre} in the previous segment $l-1$ in order to provide phase-locked tracking of a supposed sinusoidal trajectory. The difference between the frequency in the previous segment and initial frequency estimate for the current segment provides also an initial estimate of the sweep rate. A correspondingly sweeping heterodyne signal

$$x_{\text{het},i}[n] = e^{-j\phi_{\text{het},i}(\frac{n+1/2}{f_s})} \quad (5)$$

with the frequency

$$\frac{d\phi_{\text{het},i}(t)}{2\pi dt} = f_{\text{het},i}(t) = f_{\text{pre}} + \frac{t-t_{l-1}}{T}(\hat{f}'_i - f_{\text{pre}}) \quad (6)$$

permits to extract a slanted section of the time-frequency plane for the accurate frequency and sweep rate estimation (Figure 1(d)).

Linear regression for the phase $\phi_{\text{lp},i}(t)$ of the lowpass signal now uses a cubic approximation. However, two of the four parameters are predetermined by the frequency f_{pre} and phase ϕ_{pre} of the supposed sinusoidal trajectory in the previous segment. It is convenient to use the center t_{l-1} of the previous segment as origin of a local time axis $t' = t - t_{l-1}$ for the regression.

For a given initial frequency estimate \hat{f}'_i , there can be more than just one potential trajectory to be continued from the previous segment. Hence, those two candidates in the previous segment that match best the initial frequency and amplitude estimate are tested. To account for newly beginning trajectories, also a constant frequency estimation as described in Subsection 2.2 is carried out. Finally, of these three alternatives the one achieving the minimum residual energy after amplitude estimation, resynthesis, and subtraction is selected.

4. RESULTS

In order to assess the performance of the different sinusoidal parameter estimation techniques discussed here, a synthetic test signal sampled at $f_s = 16$ kHz with $L = 5$ sinusoids having a fixed amplitude $a_i(t) = 1$ and constant or varying frequencies $f_i(t)$ was used. The signal has a duration of 1 s and is repeated once, now with additional white noise with the same RMS level as each of the sinusoids.

Figure 3(a) shows the trajectories found by sweep estimation (Subsection 2.3) with subsequent parameter matching to build trajectories (Subsection 3.1). Figure 3(b) shows the trajectories found by tracking sweep estimation (Subsection 3.2)

For the new approach, the only tracking error in the left half of Figure 3(b) can be seen at 0.8 s, 1.2 kHz. In contrast, the old technique (Figure 3(a)) fails for all crossing trajectories. In case of additional noise (right half of Figure 3), the performance of the new approach is reduced, but still significantly better than the old technique.

Furthermore, it should be noted that the more reliable estimation achieved by the tracking sweep estimation presented here also leads to significantly lower residual signal (not shown in Figure 3). This improves the performance of the subtractive signal decomposition and leads to a final residual signal that is much less influenced by estimation errors, hence enabling a better estimation of the noise component parameters in an HILN encoder.

The sinusoidal tracking algorithm introduced here was also employed in the encoder used for the final verification test of the MPEG-4 parametric audio coding tool HILN. In this listening test, HILN has been compared to other state-of-the-art audio coding systems at bit rates of 6 and 16 kbit/s [9]. The test has shown that HILN performs comparable to MPEG-4 TwinVQ when operated at 6 kbit/s and comparable to MPEG-4 AAC when operated at 16 kbit/s. See [9] and [5] for a more detailed analysis of the test results and the additional functionalities provided by HILN.

5. CONCLUSIONS

In this paper, a new approach for reliable tracking and parameter estimation for time-varying sinusoids was introduced. When compared to earlier techniques, significantly improved performance for synthetic test signals was observed. This new approach for tracking sweep estimation was also successfully demonstrated in an encoder for MPEG-4 parametric audio coding HILN.

To further improve the sinusoidal parameter estimation, a joint estimation approach for sinusoids closely spaced in frequency, especially in the case of crossing trajectories, should be investigated.

6. REFERENCES

- [1] R. McAulay and T. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 34, no. 4, pp. 744–754, Aug. 1986.
- [2] X. Rodet, "Musical sound signals analysis/synthesis: Sinusoidal+residual and elementary waveform models," in *Proc. IEEE Time-Frequency and Time-Scale Workshop (TFTS'97)*, Coventry, Aug. 1997.
- [3] H. Purnhagen, "Advances in parametric audio coding," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, Mohonk, New Paltz, Oct. 1999, pp. 31–34.
- [4] ISO/IEC, "Coding of audio-visual objects – Part 3: Audio (MPEG-4 Audio Edition 2001)," ISO/IEC Int. Std. 14496-3:2001, 2001.
- [5] H. Purnhagen and N. Meine, "HILN – the MPEG-4 parametric audio coding tools," in *Proc. IEEE Int. Symposium on Circuits and Systems (ISCAS)*, Geneva, CH, May 2000, pp. III–201 – III–204.

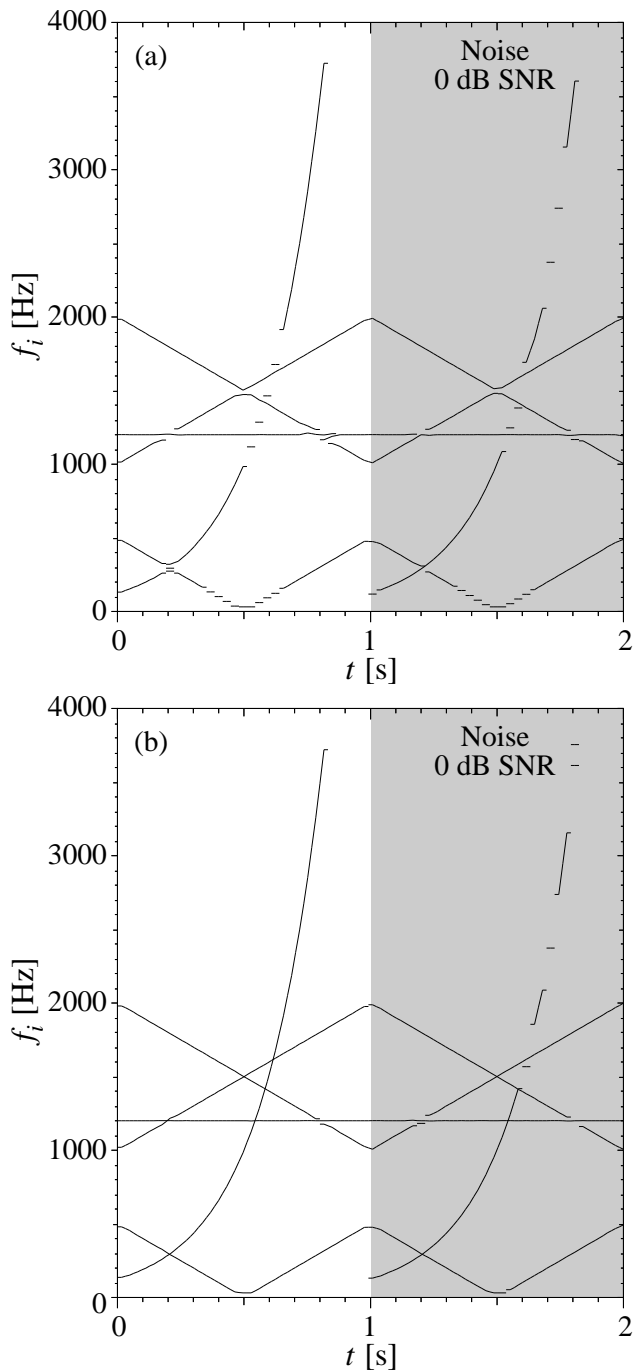


Figure 3: Sinusoidal trajectories found by sweep estimation with parameter matching (a) and tracking sweep estimation (b) for a synthetic signal containing $I = 5$ sinusoids with $a_i(t) = 1$ and additional white noise with $\sigma_n = \sqrt{1/2}$ starting at $t = 1$ s (sampling frequency $f_s = 16$ kHz, frame length $T = 32$ ms).

- [6] S. M. Kay, *Modern Spectral Estimation*, chapter 13: Sinusoidal Parameter Estimation, pp. 407–445, Prentice-Hall, Englewood Cliffs, NJ, US, 1988.
- [7] B. Edler, H. Purnhagen, and C. Ferekidis, “ASAC – analysis/synthesis audio codec for very low bit rates,” in *AES 100th Convention*, Copenhagen, May 1996, Preprint 4179.
- [8] S. Kay, “A fast and accurate single frequency estimator,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, no. 12, pp. 1987–1989, Dec. 1989.
- [9] ISO/IEC JTC1/SC29/WG11, “Report on the MPEG-4 audio version 2 verification test,” ISO/IEC JTC1/SC29/WG11 N3075, Maui, Dec. 1999, available: <http://www.tnt.uni-hannover.de/project/mpeg/audio/public/w3075.pdf>.