Audio Engineering Society

Convention Paper

Presented at the 110th Convention
2001 May 12–15   Amsterdam, The Netherlands

Accurate Sinusoidal Model Analysis and Parameter Reduction by Fusion of Components

Tuomas Virtanen

Tampere University of Technology, Signal Processing Laboratory
TAMPERE, P.O.Box 553, FIN-33101, Finland

ABSTRACT

A method is described, with which two stable sinusoids can be represented with a single sinusoid with time-varying parameters and in some conditions approximated with a stable sinusoid. The method is utilized in an iterative sinusoidal analysis algorithm, which combines the components obtained in different iteration steps using described the method. The proposed algorithm improves the quality of the analysis at the expense of an increased number of components.

INTRODUCTION

Sinusoidal modeling is a powerful parametric representation for audio signals. It represents the periodic components of a signal with sinusoids with time-varying frequencies, amplitudes, and phases. The parameters are updated from frame to frame, and sinusoidal analysis algorithms are usually frame-based, too.

In polyphonic, real-world signals the density of sinusoidal components can be very high. Also the sinusoids are usually not stable, which makes it difficult to estimate their parameters accurately.

There are complex algorithms, which do the analysis in only one pass, and iterative methods that try to get a better estimation of the parameters in each iteration, for example [1]. Because of errors and inaccuracies in the sinusoidal analysis, there might be some harmonic components left in the residual. One approach to correct this phenomenom is to detect sinusoids iteratively from the residual. There are algorithms, which detect only one sinusoid at time, synthesize it, and then remove from the residual, for example [2]. Our system detects several sinusoids at each pass, therefore requiring only two or three iterations.

ITERATIVE ANALYSIS

The sinusoids that are not detected are left in the residual. If the parameters of the detected sinusoids are inaccurate, there remain sinusoids in the residual, the frequencies of which are close to the original ones.

A natural approach to remove the sinusoids from the residual is to analyze the residual iteratively with the same analysis algorithms. If the sinusoids obtained from the residual are combined with the trajectories obtained from the original signal, a sinusoid which parameters were inaccurate becomes presented with two or more sinusoids. Normally, this is an undesirable situation. The proposed method combines the sinusoids obtained in different iterations, therefore reducing the total number of the parameters.

The block diagram of the system is illustrated in Figure 1. In the first iteration, the input signal is analyzed using a conventional sinusoidal analysis system. This block can itself be very complex, but basically any sinusoidal analysis system can be used. In our experiments, sinusoidal likeness measure was used to detect the meaningful sinusoidal peaks [3]. The frequency resolution was improved using quadratic interpolation [4]. The amplitudes and phases are obtained using non-iteratively the least-squares solution proposed in [1]. The peaks are tracked into trajectories by
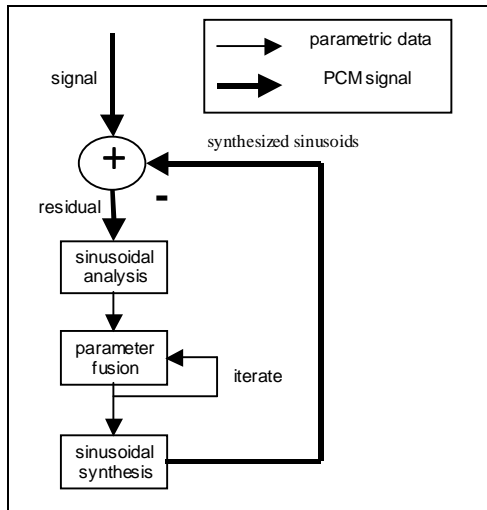
Figure 1: Block diagram of the iterative analysis system.

synthesizing the possible continuations and comparing them to the original signal. The trajectories are filtered using the methods presented in [5]. The obtained trajectories are then synthesized and subtracted from the original signal in time domain to obtain the residual.

In the following iterations, the residuals are analyzed with the same sinusoidal analysis algorithms. The parameters of the analysis, for example the sensitivity in the peak detection, can be varied from iteration to iteration. The sinusoidal trajectories obtained in different iterations are fused together using the methods proposed in the next section. Using the trajectories obtained in the first iteration and the remaining errors obtained in the following iterations, the parameters of the underlying sinusoids can be estimated. Again, the combined sinusoids are synthesized and the iteration continues. The iterative procedure can be repeated as long as desired. For example, the iteration can be stopped if no significant harmonic components are found from the residual. In our analysis system, two iterations was found to be quite enough.

The iterative algorithm is computationally expensive, since each iteration requires one pass of a conventional analysis, and synthesis of the sinusoids, too. Compared to the analysis and syntesis, the fusion of sinusoids is computationally cheap.

FUSION OF TWO SINUSOIDS

Representation of Two Sinusoids with a Single Sinusoid and Time-varying Parameters
Let us have two sinusoids, the amplitudes, frequencies, and phases of which are a_1, a_2, ω_1, $\omega2$, φ_1, and φ_2, respectively. The sum of the sinusoids at time t is denoted by $x(t)$:

$$x(t) = a_1 \sin(\omega_1 t + \varphi_1) + a_2 \sin(\omega_2 t + \varphi_2) \qquad (1)$$

Using the basic trigonometric formulas this can be converted into a form where the two terms have equal frequencies and time-varying amplitudes:

$$x(t) = \sin\left(\frac{(\omega_2 - \omega_1)t + \varphi_2 - \varphi_1}{2}\right)(a_2 - a_1)\cos\left(\frac{(\omega_2 + \omega_1)t + \varphi_2 + \varphi_1}{2}\right) +$$
$$\cos\left(\frac{(\omega_2 - \omega_1)t + \varphi_2 - \varphi_1}{2}\right)(a_2 + a_1)\sin\left(\frac{(\omega_2 + \omega_1)t + \varphi_2 + \varphi_1}{2}\right)$$

The sine and cosine of equal frequency can be combined into a single term, the amplitude and phase of which are time-varying:

$$x(t) = a_3(t)\sin\left(\frac{(\omega_2 + \omega_1)t + \varphi_2 + \varphi_1}{2} + \varphi_3(t)\right),$$

where

$$a_3(t) = \sqrt{a_1^2 + a_2^2 + 2a_1 a_2 \cos((\omega_2 - \omega_1)t + \varphi_2 - \varphi_1)} \qquad (2)$$

and

$$\varphi_3(t) = \arctan\left(\tan\left(\frac{(\omega_2 - \omega_1)t + \varphi_2 - \varphi_1}{2}\right)\right) + \theta, \qquad (3)$$

where correction term θ takes the negative amplitudes into account:

$$\theta = \begin{cases} \pi & \frac{\pi}{2} < \frac{\varphi_2 - \varphi_1}{2} \bmod 2 < \frac{3\pi}{2} \\ 0 & otherwise \end{cases} \qquad (4)$$

By taking a derivative of the phase we can represent the time-varying phase with an initial phase $\varphi_3(0)$ plus a time-varying integral of the frequency $\omega_3(t)$:

$$\varphi_3(0) = \arctan\left(\tan\left(\frac{\varphi_2 - \varphi_1}{2}\right)\right) + \theta, \qquad (5)$$

$$\omega_3(t) = \frac{d}{dt}\varphi_3(t)$$

$$= \frac{1 + \tan^2\left(\frac{(\omega_2 - \omega_1)t + \varphi_2 - \varphi_1}{2}\right)}{1 + \tan^2\left(\frac{(\omega_2 - \omega_1)t + \varphi_2 - \varphi_1}{2}\right)\frac{(a_2 - a_1)^2}{(a_2 + a_1)^2}}\left(\frac{\omega_2 - \omega_1}{2}\right)\frac{(a_2 - a_1)}{(a_2 + a_1)}$$
$$\qquad (6)$$

Now we can represent the original signal $x(t)$ with a single sinusoid with time-varying amplitude and frequency:

$$x(t) = a_3(t)\sin\left(\frac{(\omega_2 + \omega_1)t + \varphi_2 + \varphi_1}{2} + \int_0^t \omega_3(u)du + \varphi_3(0)\right) \quad (7)$$

Approximation with Constant Parameters
In the sinusoidal model, the parameters are assumed constant inside a frame. In certain conditions, the derived time-varying parameters can be approximated with constant values. The conditions in our iterative system are:

1. Time t is near zero. This means that the approximated values are valid only in a small time frame. The parameters of the sinusoidal model are updated from frame to frame, so this condition is fulfilled. The shorter the time frame is, the better.

2. The frequencies are close to each other. When conditions 1 and 2 hold, term $(\omega_2 - \omega_1)t$ in the equations 2 and 3 becomes neglible.

3. The amplitude envelope of the sum of the two sinusoids does not have a local maximum or minimum inside the time frame. This depends on the phases and frequencies of the original sinusoids. The condition is fulfilled if

$$0 \leq (\omega_2 - \omega_1)T + \left(\varphi_2 - \varphi_1 + \frac{\pi}{2}\right) \bmod \pi \leq \pi \,,\; T \text{ being the}$$

length of the frame.

4. The ratio of the amplitudes a_1 and a_2 is large. This happens in situations where the first sinusoid is obtained on the first analysis pass, and the second one is the error remaining from the first one. If this condition is fulfilled, the term $\left(\dfrac{(a_2 - a_1)}{(a_2 + a_1)}\right)^2$ is near unity.

If these conditions are fulfilled, the sinusoid with time-varying parameters can be approximated with a sinusoid with constant parameters:

$$x(t) \approx a_n \sin(\omega_n t + \varphi_n) \tag{8}$$

where constants a_n, ω_n and φ_n are parameters of the new sinusoid which replaces the old ones. The approximations are:

$$a_n = \sqrt{a_1^2 + a_2^2 + 2a_1 a_2 \cos(\varphi_1 - \varphi_2)} \tag{9}$$

$$\omega_n = \frac{\omega_1 a_1 + \omega_2 a_2}{a_1 + a_2} \tag{10}$$

and

$$\varphi_n = \arctan\left(\tan\left(\frac{\varphi_2 - \varphi_1}{2}\right)\frac{a_2 - a_1}{a_2 + a_1}\right) + \frac{\varphi_2 - \varphi_1}{2} + \theta \,. \tag{11}$$

An example of the approximation is illustrated in Figure 2. In synthesis, the parameters of the sinusoids are interpolated from frame to frame. Therefore, it is difficult to measure the validity of the approximation in a single time frame. The amplitudes are interpolated linearly, and if there is no local maxima or minima between the frames, the interpolation should work well. The linear interpolation of the amplitude envelope of a sum of two sinusoids is illustrated in Figure 3. It can be seen clearly that near zero the approximation is better.

In practise, the condition 3 sets the maximum for the difference between the frequencies. The smaller the time frame, the larger the difference can be.

Fusion of Sinusoidal Trajectories
In the sinusoidal model, the harmonic components are represented with trajectories that consist of spectral peaks in successice time frames. Each trajectory has an onset and offset time, which define the range in which the trajectory exists. In the parameter fusion the aim is to combine two closely spaced trajectories. For all trajectory pairs that overlap each other in time, the individual peaks are examined if they fulfil the conditions required for the fusion. In practise, the most important condition is the closeness of the frequencies.

If all the peaks of the two trajectories that overlap with each other fulfil the conditions, new parameters are estimated using the appromations presented above. The old trajectories are replaced with the new one. In practise, not all the peaks have to fulfil all the conditions if the trajectories otherwise match well with each other.

EXPERIMENTAL RESULTS
In complex real-world signals, the density of sinusoidal components can be very high, and there are no obvious numerical ways to measure the performance of a sinusoids+noise analysis system. Therefore the performance of the analysis algorithms was studied by calculating some statistics from analysis and synthesis results obtained for a set of music samples and for a generated test signal.

The same sinusoidal analysis system described in the previous chapter was used for the iterative and non-iterative algorithms. In iterative analysis two iterations were used, so the residual was analysed only once.

Comparison Using a Generated Test Signal
The test signal introduces phenomena usually encountered in musical signals: different kinds of changes in amplitude and frequency, harmonic sounds composed of sinusoids that overlap
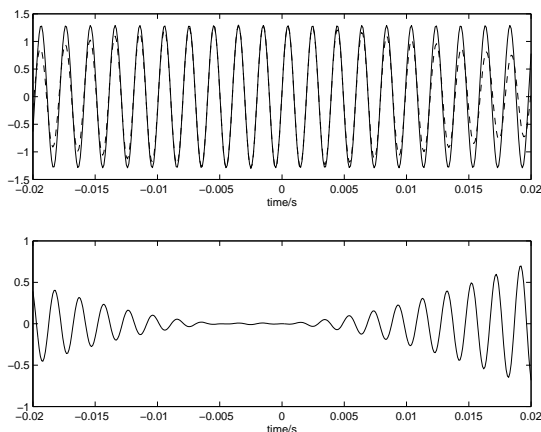

Figure 2: An example of the fusion of two sinusoids. In the upper plot the dashed line is a sum of two sinusoids, the frequencies and of which are 500 and 520 Hz and the amplitudes 1 and 0.3. The solid line is the result of the approximation. In the lower plot is illustrated the error between the two original sinusoids and the one approximated sinoid.
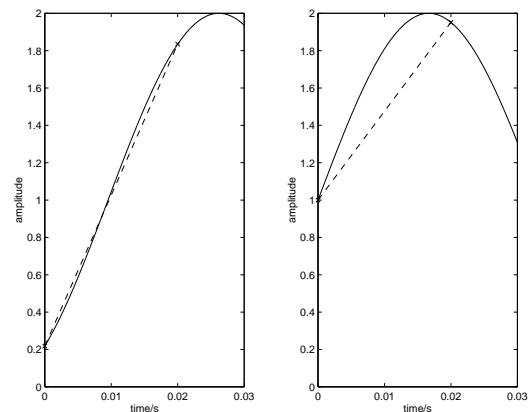

Figure 3: Linear interpolation of the amplitude envelope of a sum of two sinusoids. The solid line is the original amplitude envelope and the dashed line is linear approximation. In the left plot the amplitude envelope has no local extreme values the approximation is valid. In right plot there is a local maximum so the approximation is not valid.

Table 1: Description of the generated test signal.

Section	Signal description. Amplitude is unity (0 dB) unless otherwise stated.
1	Stable sinusoids at different frequencies, one sinusoid at a time.
2	Frequency sweep of a sinusoid from 20 Hz to 10 kHz. The speed of the sweep was exponential on frequency scale.
3	Single sinusoid the amplitude of which fades exponentially from 0 dB to -40 dB
4	Mix of sinusoids with different amplitude and frequency modulations (tremolo and vibrato). The modulation frequencies vary from 0 to 20 Hz, amplitude deviaton from 0 to 1 and frequency deviation from 0 to 1.5 semitones (0 to 9.05% of the center frequency).
5	Frequency crossing of two sinusoids at several different frequencies.
6	Stable harmonic sounds at different fundamental frequencies. All the sounds had 10 first harmonic partials, with unity amplitudes.
7	A frequency sweep of a harmonic sound, ten harmonic partials.
8	Vibrato of a harmonic sound. The modulation frequency and depth of the vibrato were time-varying like in section 4.
9	Different kind of sharp attacks of a Shephard tone. The harmonics were at frequencies 100, 200, 400, ..., 3200, 6400 Hz.
10	Frequency sweep of a harmonic sound, mixed with a constant harmonic sound.

with each other, colliding sinusoids etc. The signal was divided into ten sections, which are described in Table 1.

The generated test signal was analyzed in three different noise conditions: The levels of additive white noise were no noise, low - 14 dB noise and loud +6 dB noise. The reference level 0 dB is a single sinusoid with unity amplitude. The noise energy is for the whole 0-22 kHz frequency range.

Since the test signal is composed of sinusoids only, the remaining error of the residual describes the performance of the analysis system. The signal-to-residual ratios (SRRs) were calculated for all the sections, and averaged over the three noise levels. The results are illustrated in Table 2. The noise removed before calculating the SRRs to get a measure how well the sinusoids have been detected from the noise. It should be noted that for single, stable sinusoids it is easy obtain SRRs of about 50 dB even with quite simple methods in noiseless environment.

Table 2: Signal-to-residual ratios obtained with the iterative and non-iterative analysis system.

Section	SRR without iteration	SRR with iteration	Percentage of additional sinusoids
1	17.4	17.4	0
2	11.0	13.4	50
3	31.0	40.0	0
4	12.4	14.1	15
5	9.8	11.5	10
6	2.3	2.6	6
7	1.7	2.2	50
8	6.7	10.0	15
9	24.6	24.3	2
10	1.4	1.8	39

The generated test signal was made advisedly difficult to bring out the differences between the analysis algorithms. In section 1, the low SRRs are caused mostly by low-frequency sinusoids, which are difficult to detect with a normal analysis window.

The performance of the iterative and non-iterative system was studied by calculating the average error of the parameters and the number of missed peaks, too. These studies show that the improvement in the SRRs is caused mostly by the additional sinusoids detected. In a few cases the parameters become more accurate with the iterative analysis, like the SRRs of the section 3 show: the number of sinusoids is the same but an improvement of about 10 dB is gained. In noiseless conditions the difference was even larger: an improvement of 27 dB (56 to 83) was gained. In noisy environment the improvements are smaller, because the estimation errors can be quite small compared to the noise levels. In most sections the average parameter errors are almost equal with iterative and non-iterative system, and the improvement in the quality comes at the expense of an increased number of components.

Comparison Using Musical Signals

The performance of the iterative analysis was tested with four musical signals, too. In musical signals there are non-periodic components like drums that should not be represented with sinusoids, and signal-to-residual-ratios can be as low as only a couple of dBs even though the sinusoidal analysis was perfect. Therefore, the SRRs should not be the only performance measure for musical signals. To prevent any noise to be presented with sinusoids, a bit higher threshold was used in the peak detection.

The SRRs obtained using only one analysis pass ranged from 2.8 to 9.0 dB. After two iterations, the SRRs ranged from 3.5 to 10.8, and an average improvement of 1.9 dB was gained. The percentage of the additional sinusoids ranged from 75 to 86%. The results were studied by listening to the synthesized sinusoids and residuals, too. The perceptual quality was clearly better with the iterative algorithm. The large number of additional sinusoids shows again that the largest improvement is obtained by finding completely new sinusoids, not by improving the parameters.

Parameter Reduction

Fusion of components has little use in non-iterative systems. It can be used to reduce to number of components, but usually this only makes further analysis more difficult.

The parameter fusion was tested directly with the trajectories obtained from the first iteration. The objective was to reduce the number of the sinusoids without affecting the quality of the synthesized signal. Sinusoidal trajectories analyzed with several different algorithm sets were available, so this test was done also with other analysis methods than the one described earlier.

The average number of the sinusoids was reduced by 1.1%, while the average SRR was reduced by 0.08 dB. As one can expect, that small difference was inaudible. With some signals the number of parameters was reduced by 10%, but the average reduction was still very small. Our system uses quite low frame rate (44 frames/s). With a faster frame rate it might be possible to get more reduction.

CONCLUSIONS

A method is proposed to approximate two sinusoids with a single sinusoid with time-varying parameters. The approximation is utilized in the sinusoidal analysis with an iterative algorithm. The algorithm was compared to a non-iterative analysis system by using a generated test signal and a set of musical signals. In both cases the iterative algorithm can improve the quality of the

analysis, if the remaining energy of the residual is used to judge the performance.

In most cases better quality is obtained at the expense of an increased number of components. In a few cases the accuracy of the parameters is improved without additional components.

References

[1] Depalle, Ph. & Hélie, T. "Extraction of Spectral Peak Parameters Using a Short-Time Fourier Transform And No Sidelobe Windows". IEEE 1997 Workshop on Applications of Signal Processing to Audio and Acoustics". Mohonk, New York, 1997.

[2] George, E. & Smith M. "Speech Analysis/Synthesis and Modification Using an Analysis-by-Synthesis/Overlap-Add Sinusoidal Model". IEEE Transactions on Speech And Audio Processing, Vol. %, No. 5, September 1997.

[3] Rodet, Xavier. "Musical Sound Signal Analysis/Synthesis: Sinusoidal+Residual and Elementary Waveform Models". IEEE Time-Frequency and Time-Scale Workshop 1997, Coventry, Grande Bretagne.

[4] Smith, J.O., Serra, X. "PARSHL: An analysis/synthesis program for non-harmonic sounds based on a sinusoidal representation", Proceedings of the International Computer Music Conference, 1987.

[5] Levine, Scott. "Audio Representation for Data Compression and Compressed Domain Processing". Ph.D.   thesis. Stanford University.