ADAPTIVE MODELING OF SYNTHETIC NONSTATIONARY SINUSOIDS

Marcelo Caetano[*]
INESC TEC
Sound and Music Computing Group
Porto, Portugal
mcaetano@inesctec.pt

George Kafentzis
University of Crete
Department of Computer Science
Heraklion, Greece
kafentz@csd.uoc.gr

Athanasios Mouchtaris[†]
FORTH
SPL - Institute of Computer Science
Heraklion, Greece
mouchtar@ics.forth.gr

ABSTRACT

Nonstationary oscillations are ubiquitous in music and speech, ranging from the fast transients in the attack of musical instruments and consonants to amplitude and frequency modulations in expressive variations present in vibrato and prosodic contours. Modeling nonstationary oscillations with sinusoids remains one of the most challenging problems in signal processing because the fit also depends on the nature of the underlying sinusoidal model. For example, frequency modulated sinusoids are more appropriate to model vibrato than fast transitions. In this paper, we propose to model nonstationary oscillations with adaptive sinusoids from the extended adaptive quasi-harmonic model (eaQHM). We generated synthetic nonstationary sinusoids with different amplitude and frequency modulations and compared the modeling performance of adaptive sinusoids estimated with eaQHM, exponentially damped sinusoids estimated with ESPRIT, and log-linear-amplitude quadratic-phase sinusoids estimated with frequency reassignment. The adaptive sinusoids from eaQHM outperformed frequency reassignment for all nonstationary sinusoids tested and presented performance comparable to exponentially damped sinusoids.

1. INTRODUCTION

Music and speech contain different types of nonstationary oscillations. The attack of many musical instruments presents transients due to nonstationarities [1]. Percussive sounds feature very sharp onsets with highly nonstationary oscillations [2]. Expressiveness in performance such as *tremolo*, *vibrato*, *glissando*, and *portamento* generally results in amplitude and frequency modulations [3, 4]. Similarly, speech sounds such as consonants contain transients [5]. Consonants known as *plosives* feature a sharp onset [6]. Expressivity in speech used to convey emotions, for example, results in prosodic contours [5] or modulations in frequency and amplitude, while vibrato can be said to characterize singing [7].

Sinusoidal modeling is a popular parametric representation for speech and music. Sinusoidal models are widely used in speech and music processing for coding [8, 9, 10], analysis and synthesis [11, 12, 13, 14, 15, 16, 17], enhancement [18, 19, 20, 21], modifications and transformations [12, 15, 22, 23, 24, 25, 26].

The general problem of fitting a sum of sinusoids to a signal is of great interest in many scientific areas. Thus, many algorithms have been developed for accurate estimation of the sinusoidal parameters. For speech and musical sounds, the algorithms can be separated into four categories, namely spectral peak-picking [11, 12], analysis-by-synthesis [15, 27, 28, 29], least squares [30, 31, 32], and subspace methods [33, 34, 35].

Polynomial phase signals [36] have been used to model nonstationary oscillations. McAulay and Quatieri [11] were possibly the first to propose to interpolate the phase values estimated at the center of the analysis window with cubic polynomials. Quadratic polynomials [37] were proposed as an alternative. Girin *et al.* [38] investigated the impact of the order of the polynomial used to represent the phase. They concluded that a polynomial of order 5 does not improve the modeling performance considerably to justify the increased complexity.

The time-frequency reassigned spectrogram [39] was developed to better represent nonstationary oscillations with the short-time Fourier transform. Reassignment is widely used [40, 41, 42] to estimate the parameters of the sinusoidal model. The derivative analysis method [43, 44] was later shown [45] to be theoretically equivalent to the reassignment method.

More recently, adaptive sinusoidal models [31, 32, 46] have gained attention due to their ability to adapt to the local characteristics of the signal via an iterative parameter re-estimation process. Previous works have modeled speech [16, 32] and monophonic musical instrument sounds [17, 47] as a sum of adaptive sinusoids. These studies focused on modeling speech and musical instrument sounds recorded under controlled conditions instead of expressive conversations or music performances. Consequently, the sounds do not feature the prosodic contours or embelishements that result in challenging nonstationary modulations.

[*] This work is financed by the FCT - Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) within project "UID/EEA/50014/2013."

[†] This work is supported in part by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant 644283.

In this work, we investigate the ability of adaptive sinusoids from eaQHM [46] to model nonstationary oscillations. We generated synthetic nonstationary sinusoids with different amplitude and frequency modulations and compared the modeling performance of adaptive sinusoids estimated with eaQHM, exponentially damped sinusoids estimated with ESPRIT, and log-linear amplitude quadratic-phase sinusoids estimated with time-frequency reassignment. We designed nonstationary sinusoids with controlled amplitude and frequency modulations that mimic specific features of nonstationary oscillations found in expressive music and speech, such as *tremolo* and *vibrato*. In this article, we focus on modeling monocomponent signals composed of one of these nonstationary sinusoids to compare the ability of each model to capture that specific feature. We measure the modeling accuracy in the time domain with the signal-to-reconstruction-error ratio (SRER). The SRER is the ratio in dB between the energy in the original signal and in the modeling residual.

The next section briefly describes the underlying sinusoidal model for the exponentially damped sinusoidal model (EDSM), reassigned sinusoidal model (RSM), and extended adaptive quasi-harmonic model (eaQHM). Then we describe the synthetic nonstationary sinusoids used in this work. Next, we present the modeling performance of EDSM, RSM, and eaQHM for the nonstationary signals designed, followed by a discussion of the results. Finally, the conclusions and perspectives are presented.

2. SINUSOIDAL MODELS

Sinusoidal models implicitly assume that each partial (e.g., oscillatory mode) can be described by a time-varying sinusoid $s\left(t\right)$ as

$$s\left(t\right) = A\left(t\right)\cos\left[\Phi\left(t\right)\right], \tag{1}$$

where $A\left(t\right)$ is the time-varying amplitude and $\Phi\left(t\right)$ is the time-varying phase, jointly called the instantaneous components of the signal. $A\left(t\right)$ and $\Phi\left(t\right)$ describe respectively the long-term amplitude and frequency modulations of each partial along the total duration of the sound. Usually, these long-term variations are approximated by piece-wise functions inside short-term signal frames $x\left(t\right)$ typically lasting miliseconds obtained as

$$x_k\left(t\right) = s\left(t\right)w\left(t - k\tau\right), \;\; 0 \leq k \leq N-1, \tag{2}$$

where k is the frame number, τ is the time shift (hop size), and $w\left(t\right)$ is a window function with length L that is zero outside the support L. Typically, $\tau < L$ so that the frames overlap and $s\left(t\right)$ is modeled as N frames $x_k\left(t\right)$ viewed through a sliding window $w\left(t - k\tau\right)$ centered at τ as follows

$$s\left(t\right) = \sum_{k=0}^{N-1} x_k\left(t\right) = \sum_{k=0}^{N-1} s\left(t\right)w\left(t - k\tau\right). \tag{3}$$

Eq. (3) holds for windows $w\left(t\right)$ that satisfy the constant overlap-add (COLA) [48] constraint $\sum_{k=0}^{N} w\left(t - k\tau\right) = 1$, valid only for specific values of τ.

When $s\left(t\right)$ is assumed to be locally stationary, $x\left(t\right)$ becomes

$$x\left(t\right) = A\cos\left(\omega t + \theta\right) \tag{4}$$

modeled as a sinusoid with constant amplitude A, constant frequency $\omega = 2\pi f_0$ and constant phase shift θ. In this case, the long-term model for the partial $s\left(t\right)$ is composed of piece-wise stationary oscillations only capable of capturing relatively stable amplitude and frequency modulations. However, nonstationary oscillations commonly vary enough inside the frame to require a dedicated short-term model. In what follows, we describe the underlying short-term signal model $x\left(t\right)$ for the nonstationary sinusoidal models used in this work, namely exponentially damped sinusoidal model (EDSM), reassigned sinusoidal model (RSM), and extended adaptive quasi-harmonic model (eaQHM).

2.1. Exponentially Damped Sinusoidal Model (EDSM)

EDSM assumes that $x\left(t\right)$ can be approximated by the underlying signal model

$$x\left(t\right) = \exp\left(\lambda + \mu t\right)\cos\left(\omega t + \theta\right), \tag{5}$$

where $A\left(t\right) = \exp\left(\lambda + \mu t\right)$ is the temporal envelope and $\Phi\left(t\right) = \omega t + \theta$ is the time-varying phase. The short-term frame $x\left(t\right)$ in EDSM is simply modeled as a stationary sinusoid with constant frequency ω modulated in amplitude by an exponential envelope controlled by λ and μ. $A\left(t\right)$ grows exponentially when $\mu > 0$, decays when $\mu < 0$, and is constant if $\mu = 0$.

The literature has shown [33, 34, 35, 49] that subspace methods render accurate parameter estimation for EDSM. This work uses ESPRIT to fit the parameters of EDSM [35].

2.2. Reassigned Sinusoidal Model (RSM)

RSM can be shown to render good modeling performance [45] when $x\left(t\right)$ can be approximated by the underlying signal model

$$x\left(t\right) = \exp\left(\lambda + \mu t\right)\cos\left(\psi t^2 + \omega t + \theta\right), \tag{6}$$

where $A\left(t\right) = \exp\left(\lambda + \mu t\right)$ is the temporal envelope and $\Phi\left(t\right) = \psi t^2 + \omega t + \theta$ is the time-varying phase. The short-term frame $x\left(t\right)$ in RSM is approximated as a sinusoid with quadratic phase (quadratic frequency ψ, linear frequency ω, and phase shift θ) modulated in amplitude by an exponential envelope controlled by λ and μ similarly to EDSM.

The parameters of the model are estimated using the time-frequency reassignment method [40, 41, 42, 45]. This work uses the DESAM [50] toolbox to fit the parameters of RSM.

2.3. The extended adaptive Quasi-Harmonic Model (eaQHM)

The assumption behind eaQHM is that speech and musical sounds can be approximated by a sum of M quasi-harmonic, highly nonstationary, AM-FM modulated partials $s_m\left(t\right)$. Each partial is further modeled *inside* the analysis frame as a short-term $x\left(t\right)$ which can be approximated by the underlying signal model

$$x\left(t\right) = \left(\lambda + \mu t\right)\cos\left(\psi t^2 + \omega t + \theta\right), \tag{7}$$

where $A\left(t\right) = \left(\lambda + \mu t\right)$ is the temporal envelope and $\Phi\left(t\right) = \psi t^2 + \omega t + \theta$ is the time-varying phase. The short-term frame $x\left(t\right)$ in eaQHM is implicitly modeled as a sinusoid with quadratic phase (quadratic frequency ψ, linear frequency ω, and phase shift θ) modulated in amplitude by a *linear* envelope controlled by λ and μ. $A\left(t\right)$ grows linearly when $\mu > 0$, decays when $\mu < 0$, and is constant if $\mu = 0$.

The parameters λ, μ, ψ, ω, and θ are iteratively adapted from successive steps of parameter estimation using least squares [46]. Adaptation arises from a sequence of parameter re-estimation steps based on successive refinements of the model basis functions, which

come directly from (7). The complete parameter estimation algorithm is described elsewhere [46].

3. SYNTHETIC NONSTATIONARY SINUSOIDS

The algorithms will be tested on synthetic nonstationary sinusoids to show the properties of each model, along with their corresponding advantages and disadvantages. The following parameters were used to generate the synthetic nonstationary sinusoids, sampling frequency $F_s = 16$ kHz and total length $N = 1600$ samples, corresponding to 100 ms.

All the synthetic nonstationary sinusoids generated are a combination of an amplitude envelope (A) and phase (P). The amplitude envelopes are constant (C), exponential (E), linear (L), cubic (C3), sinusoidal (S), and exponential-sinusoidal (ES). The phases are linear (L), quadratic (Q), cubic (C3), or sinusoidal (S). The synthetic nonstationary sinusoids are described next.

3.1. Constant Amplitude Linear Phase (CA-LP)

This is simply a stationary sinusoid used as reference. All the methods are expected to perform very well for stationary sinusoids.

$$s\left(t\right) = A\cos\left(\omega t + \theta\right), \tag{8}$$

where A is the constant amplitude, $\omega = 2\pi f_0$ is the constant frequency, and θ is the phase shift. The parameter values were $A = 1$, $f_0 = 100$, and $\theta = \frac{-\pi}{2}$.

3.2. Exponential Amplitude Linear Phase (EA-LP)

This corresponds to the underlying model from EDSM, thus we expect EDSM to perform very well for this particular case.

$$s\left(t\right) = \exp\left(\lambda + \mu t\right)\cos\left(\omega t + \theta\right), \tag{9}$$

where λ and μ are respectively the constant and damping factors. Thus $A\left(t\right)$ grows linearly when $\mu > 0$, decays when $\mu < 0$, and is constant if $\mu = 0$. For the phase, $\omega = 2\pi f_0$ is the constant frequency, and θ is the phase shift. The parameter values were $\lambda = 0$, $\mu = -50$, $f_0 = 100$, and $\theta = \frac{-\pi}{2}$.

3.3. Exponential Amplitude Quadratic Phase (EA-QP)

This is the underlying model from RSM, thus we expect RSM to fit this signal very well.

$$s\left(t\right) = \exp\left(\lambda + \mu t\right)\cos\left(\psi t^2 + \omega t + \theta\right), \tag{10}$$

where λ and μ are respectively the constant and damping factors, ψ, is the quadratic frequency $\omega = 2\pi f_0$ is the linear frequency, and θ is the phase shift. The parameter values were $\lambda = -0.5$, $\mu = -5$, $\psi = \left(2\pi\right)^2 f_1$ with $f_1 = 100$, $\omega = 2\pi f_0$ with $f_0 = 440$, and $\theta = \frac{-\pi}{2}$.

3.4. Constant Amplitude Cubic Phase (CA-C3P)

This is a particularly challenging signal because the cubic phase has a large range of variation. Depending on the location of the roots, the phase can vary slowly at first and suddenly grow very fast. We expect the C3P to be challenging for all models mainly because it does not match the underlying signal used by any.

$$s\left(t\right) = A\cos\left(\phi t^3 + \psi t^2 + \omega t + \theta\right), \tag{11}$$

where A is the constant amplitude, ϕ is the cubic phase, ψ, is the quadratic frequency $\omega = 2\pi f_0$ is the linear frequency, and θ is the phase shift. The parameter values were $A = 1$, $\phi = \left(2\pi\right)^3 f_2$ with $f_2 = 4,597.7$, $\psi = \left(2\pi\right)^2 f_1$ with $f_1 = 1,661.1$, $\omega = 2\pi f_0$ with $f_0 = 156.8$, and $\theta = \left(\frac{-\pi}{2}\right)^3$. The phase parameters were chosen to place the roots of C3P at respectively 100, 240, and 580 samples from a total of $N = 1600$ samples.

3.5. Exponential Amplitude Cubic Phase (EA-C3P)

This signal is more challenging than before because the C3 phase is modulated in amplitude by an exponential envelope. We expect this signal to be challenging for all models.

$$s\left(t\right) = \exp\left(\lambda + \mu t\right)\cos\left(\phi t^3 + \psi t^2 + \omega t + \theta\right). \tag{12}$$

The parameter values were $\lambda = 0$ and $\mu = -50$. The C3P parameters are the same as used previously in 3.5.

3.6. Linear Amplitude Cubic Phase (LA-C3P)

$$s\left(t\right) = \left(\lambda + \mu t\right)\cos\left(\phi t^3 + \psi t^2 + \omega t + \theta\right), \tag{13}$$

where λ is the vertical shift and μ is the slope of the amplitude envelope. The parameter values were $\lambda = 1$, $\mu = -10$. The C3P parameters are the same as in 3.5.

3.7. Cubic Amplitude Cubic Phase (C3A-C3P)

$$s\left(t\right) = \left(\lambda + \mu t + \gamma t^2 + \beta t^3\right)\cos\left(\phi t^3 + \psi t^2 + \omega t + \theta\right), \tag{14}$$

where the parameters λ, μ, γ, and β control the time-varying behavior of $A\left(t\right)$. The C3 amplitude envelope can be designed to vary considerably in short time frames by placing all the roots of the polynomial inside the frame. The amplitude parameter values were $\lambda = 0.059$, $\mu = -16.68$, $\gamma = -1110$, $\beta = 19305$. The C3A is simply the C3P signal normalized between 0 and 1. The C3P parameter values are the same as in 3.5.

3.8. Sinusoidal Amplitude Sinusoidal Phase (SA-SP)

This example contains both classic AM and FM modulations. We expect this signal to pose a challenge for all models.

$$s\left(t\right) = \left[A + B\cos\left(\omega_A t\right)\right]\cos\left(\omega_0 t + \theta_0 + \alpha\cos\left(\omega t\right)\right), \tag{15}$$

where A is the constant gain, B is the amplitude of the sinusoid, and ω_A is the constant frequency of the amplitude envelope. The sinusoidal amplitude envelope mimics the classic amplitude modulation signal and it arises in cases when there is *tremolo* or *beating* frequencies. The frequency parameter ω controls the rate of temporal variation inside the frame. The parameter values are $A = 0.7143$, $B = 0.2857$, $\omega_A = 2\pi f_A$ with $f_A = 50$, $\omega_0 = 2\pi f_0$ with $f_0 = 1000$, $\theta_0 = \frac{-\pi}{2}$, $\alpha = 1$, $\omega = 2\pi f_P$ with $f_P = 130$.

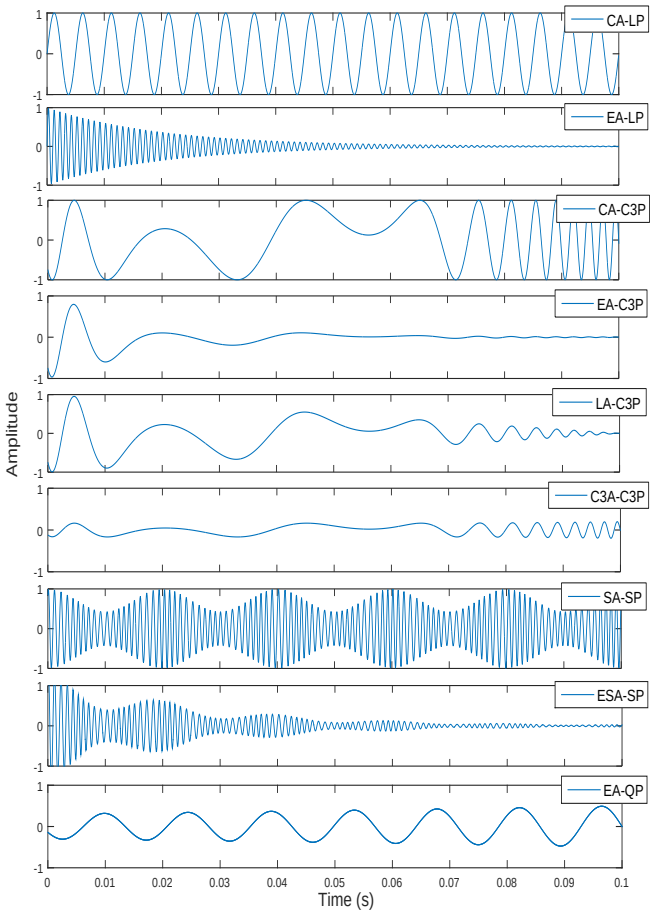

Figure 1: Illustration of the waveform for each synthetic nonstationary sinusoid from section 3.

3.9. Exponentially Damped Sinusoidal Amplitude Sinusoidal Phase (ESA-SP)

The ESA-SP sinusoid is $s(t) = A(t)\cos[\Phi(t)]$ with $A(t)$ and $\Phi(t)$ given below.

$$A(t) = \exp(\lambda + \mu t)\left[A + B\cos(\omega_A t)\right], \tag{16}$$

$$\Phi(t) = \omega_0 t + \theta_0 + \alpha\cos(\omega t), \tag{17}$$

where λ and μ are respectively the constant and damping factors and A is the constant gain, B is the amplitude of the sinusoid, and ω_A is the constant frequency. This amplitude envelope is simply the multiplication of the exponential (E) and the sinusoidal (S) envelopes, thus the result is a sinusoid modulated by the exponential. The parameter values are the same as in 3.8.

Figure 1 shows the synthetic nonstationary sinusoids described above to illustrate the resulting waveforms. Note that each signal presents very different characteristics, imposing different challenges for the models.

4. SYNTHETIC NONSTATIONARY SINUSOID MODELING ACCURACY

This section presents the experiment performed to compare the modeling accuracy of eaQHM, RSM, and EDSM for the nonstationary sinusoids described in section 3. We modeled each synthetic nonstationary sinusoid described earlier with eaQHM, RSM, and EDSM and measured the resulting modeling accuracy with the SRER as described next.

4.1. Measuring Modeling Accuracy

In what follows, we assume that the following relation holds

$$s(t) = y(t) + \hat{y}(t), \tag{18}$$

where $s(t)$ is the original synthetic signal, $y(t)$ is the model reconstruction after resynthesis, and $\hat{y}(t)$ is the modeling residual obtained by subtraction of $y(t)$ from $s(t)$ in the time domain. Then, the signal-to-reconstruction-error ratio (SRER) is defined as

$$\text{SRER} = 20\log_{10}\frac{\text{RMS}[s(t)]}{\text{RMS}[\hat{y}(t)]}\ \text{dB}. \tag{19}$$

Thus the SRER is the ratio in dB of the energy in the original synthetic signal $s(t)$ and the modeling residual $\hat{y}(t)$. Positive values indicate that $s(t)$ has more energy than $\hat{y}(t)$, while negative values indicate the opposite. Note that $\hat{y}(t)$ will only have low energy when the model $y(t)$ follows $s(t)$ very closely, which indicates good modeling performance. Consequently, higher SRER values indicate a better fit.

4.2. Analysis Parameters

All the synthetic nonstationary sinusoids were split into overlapping frames prior to analysis. The hop size was $H = 0.001F_s$ or 1 ms and the window size L varied between 10 ms and 70 ms as shown in Table 1.

EDSM uses a square window $w(t)$ for analysis and a Hamming window for overlap-add (OLA) resynthesis. RSM uses Hamming windows for both analysis and OLA resynthesis, while eaQHM uses a Hamming window for analysis and analytic resynthesis directly from (7).

4.3. Results

Table 1 shows the SRER value in dB for each synthetic signal from section 3 for each L indicated. RSM resulted in negative values for some values of L.

5. DISCUSSION

Table 1 shows that not all models performed as expected. While EDSM and eaQHM presented consistent performance for all the synthetic nonstationary sinusoids tested, RSM did not present robust performance. The SRER measure is very strict because it compares the waveforms directly. So small errors in only one parameter, such as the phase shift θ, for example, will lead to poor performance when measured with the SRER because the resulting waveform will be different. However, the instability in performance is likely due to the implementation used (the DESAM [50] toolbox) rather than the RSM method itself.

Table 1: SRER values in decibels (dB) for each synthetic signal when the window size L varies between 10 ms and 70 ms. C denotes *Constant*, E denotes *Exponential*, L denotes *Linear*, Q denotes *Quadratic*, C3 denotes *Cubic*, and S denotes *Sinusoidal* for either the amplitude envelope (A) or the phase (P). See section 3 for details.

		SRER (dB) for each window Size L(ms)						
sinusoid	algorithm	$L = 10$ ms	$L = 20$ ms	$L = 30$ ms	$L = 40$ ms	$L = 50$ ms	$L = 60$ ms	$L = 70$ ms
CA-LP	eaQHM	286.7	286.6	286.5	286.5	286.4	286.4	286.4
	RSM	−6.0	28.2	−6.0	25.1	−6.0	23.4	−6.0
	EDSM	282.1	278.1	274.0	272.8	264.6	263.4	268.7
EA-LP	eaQHM	66.6	53.4	46.6	42.7	40.0	38.7	34.9
	RSM	−6.0	−6.0	−6.0	−6.0	−6.0	−6.0	−6.0
	EDSM	280.3	269.8	267.7	265.7	262.2	271.0	270.2
EA-QP	eaQHM	51.5	10.0	3.4	2.1	1.5	0.6	0.7
	RSM	−2.3	−1.6	−0.7	−0.8	−0.5	−7.7	−0.1
	EDSM	41.1	6.7	3.5	2.3	1.9	1.4	1.2
CA-C3P	eaQHM	65.4	49.3	18.1	11.7	6.6	4.1	2.7
	RSM	−6.1	−5.2	−3.1	−5.4	−4.3	−21.5	−15.9
	EDSM	46.1	24.4	12.8	7.7	5.7	4.8	4.2
EA-C3P	eaQHM	57.5	45.7	5.2	2.7	2.4	2.0	1.5
	RSM	−3.6	−2.6	−1.1	−2.3	−1.4	−14.7	−1.8
	EDSM	49.3	24.9	11.1	5.0	3.6	3.4	3.1
LA-C3P	eaQHM	69.8	52.2	2.9	1.1	0.8	0.6	0.4
	RSM	−0.5	−0.3	−0.3	−0.3	−3.2	−3.9	−0.2
	EDSM	52.5	15.5	5.4	3.1	2.6	1.8	1.6
C3A-C3P	eaQHM	29.8	7.6	4.1	3.7	3.4	3.4	3.4
	RSM	2.6	3.7	3.1	3.2	3.2	3.2	3.2
	EDSM	11.2	4.4	3.8	2.9	3.6	3.5	2.5
SA-SP	eaQHM	24.2	7.2	4.1	4.0	3.6	3.3	3.6
	RSM	3.0	3.4	3.2	3.7	3.6	3.6	3.7
	EDSM	12.0	5.7	5.5	4.7	4.9	4.8	4.8
ESA-SP	eaQHM	107.4	33.4	33.6	13.1	15.4	17.5	17.5
	RSM	−6.0	−6.0	24.6	−6.0	−0.1	20.6	−6.0
	EDSM	165.9	139.8	123.6	112.0	102.9	95.5	89.1

Nonstationary sinusoids with time-varying frequency are more challenging to model with longer windows. The modeling performance of eaQHM and EDSM decreased when L increased for all the synthetic nonstationary sinusoids except when the phase was linear, namely CA-LP and EA-LP.

Both eaQHM and EDSM present very high SRER for stationary sinusoids (CA-LP). As expected, EDSM outperformed eaQHM for its underlying sinusoid (EA-LP). EDSM also outperformed eaQHM for exponentially damped sinusoidal amplitude modulation and sinusoidal phase (ESA-SP). RSM performed poorly for its underlying sinusoid (EA-QP), presenting negative values throughout.

In general, eaQHM presented the best performance of all models for most signals tested, indicating that adaptation is able to represent well even signals that are different from its underlying model. EDSM also presents better performance than RSM possibly due to the use of ESPRIT to estimate the parameter values.

6. CONCLUSIONS AND PERSPECTIVES

In this paper, we propose to model non-stationary oscillations with adaptive sinusoids from the extended adaptive quasi-harmonic model (eaQHM). We generated synthetic non-stationary sinusoids with different amplitude and frequency modulations and compared the modeling performance of adaptive sinusoids estimated with eaQHM, exponentially damped sinusoids (EDS) estimated with ESPRIT, and log-linear-amplitude quadratic-phase sinusoids estimated with time-frequency reassignment (RSM). Modeling performance is measured with the signal-to-reconstruction-error ratio (SRER), which uses the waveforms directly. The adaptive sinusoids from eaQHM outperformed RSM for all the signals tested and presented performance comparable to EDSM.

Future work should focus on applying eaQHM to modeling recordings of expressive speech and music performance. In previous works, eaQHM has been shown to perform well when modeling relatively stable speech utterances and musical instrument sounds. However, modeling the modulations from expressive speech and music performance would be challenging. Presently, eaQHM only handles monophonic sounds. Therefore, it would be also very interesting to investigate parameter estimation strategies for polyphonic music.

7. REFERENCES

[1] N. H. Fletcher, "The nonlinear physics of musical instruments," *Reports on Progress in Physics*, vol. 62, pp. 723–764, 1999.

[2] Rolf Bader and Uwe Hansen, "Modeling of musical instruments," in *Handbook of Signal Processing in Acoustics*, David Havelock, Sonoko Kuwano, and Michael Vorländer, Eds., pp. 419–446. Springer New York, 2008.

[3] P. Herrera and J Bonada, "Vibrato extraction and parameterization in the spectral modeling synthesis framework," in *Vibrato Extraction and Parameterization in the Spectral Modeling Synthesis framework*, 1998, pp. 107–110.

[4] Henrik Von Coler and Axel Roebel, "Vibrato Detection Using Cross Correlation Between Temporal Energy and Fundamental Frequency," in *131st AES Convention*, 2011.

[5] Nicolas Obin, Christophe Veaux, and Pierre Lanchantin, "Exploiting alternatives for text-to-speech synthesis: From machine to human," in *Speech Prosody in Speech Synthesis: Modeling and generation of prosody for high quality and flexible speech synthesis*, Keikichi Hirose and Jianhua Tao, Eds., Prosody, Phonology and Phonetics, pp. 189–202. Springer Berlin Heidelberg, 2015.

[6] G. P. Kafentzis, O. Rosec, and Y. Stylianou, "On the modeling of voiceless stop sounds of speech using adaptive quasi-harmonic models," in *Interspeech*, 2012.

[7] L. Regnier and G. Peeters, "Singing voice detection in music tracks using direct voice vibrato detection," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2009, pp. 1685–1688.

[8] E. B. George and M. Smith, "A new Speech Coding Model based on a Least-Squares Sinusoidal Representation," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Apr 1987, pp. 1641–1644.

[9] R. J. Mcaulay and T. F. Quatieri, "Low-rate speech coding based on the sinusoidal model," in *Advances in Speech Signal Processing*, S. Furui and M. M. Sondhi, Eds. Marcel Dekker Inc., New York, 1992.

[10] S. Ahmadi and A. S. Spanias, "Low bit-rate speech coding based on an improved sinusoidal model," *Speech Communication*, vol. 34, no. 4, pp. 369 – 390, 2001.

[11] R. J. McAulay and T. F. Quatieri, "Speech Analysis/Synthesis based on a Sinusoidal Representation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, pp. 744–754, 1986.

[12] X. Serra, *A System for Sound Analysis, Transformation, and Synthesis based on a Deterministic plus Stochastic Decomposition*, Ph.D. thesis, Stanford University, 1989.

[13] T.F. Quatieri and R.J. McAuley, "Audio signal processing based on sinusoidal analysis/synthesis," in *Applications of Digital Signal Processing to Audio and Acoustics*, Mark Kahrs and Karlheinz Brandenburg, Eds., chapter 9, pp. 343–416. Kluwer Academic Publishers, 2002.

[14] J. Laroche, Y. Stylianou, and E. Moulines, "HNM: A Simple, Effecient Harmonic plus Noise Model for Speech," in *Workshop on Appl. of Signal Proc. to Audio and Acoustics (WASPAA)*, New Paltz, NY, USA, Oct 1993, pp. 169–172.

[15] E. B. George and M. J. T. Smith, "Speech analysis/synthesis and modification using an analysis-by-synthesis/overlap-add sinusoidal model," *IEEE Transactions on Speech and Audio Processing*, vol. 5, no. 5, pp. 389–406, Sep 1997.

[16] Y. Pantazis, G. Tzedakis, O. Rosec, and Y. Stylianou, "Analysis/Synthesis of Speech based on an Adaptive Quasi-Harmonic plus Noise Model," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2010, pp. 4246–4249.

[17] M. Caetano, G. P. Kafentzis, A. Mouchtaris, and Y. Stylianou, "Adaptive sinusoidal modeling of percussive musical instrument sounds," in *Proc. European Signal Processing Conference (EUSIPCO)*, 2013, pp. 1–5.

[18] Michael E. Deisher and Andreas S. Spanias, "Speech enhancement using state-based estimation and sinusoidal modeling," *The Journal of the Acoustical Society of America*, vol. 102, no. 2, pp. 1141–1148, 1997.

[19] J. Jensen and J.H.L. Hansen, "Speech enhancement using a constrained iterative sinusoidal model," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 7, pp. 731–740, 2001.

[20] E. Zavarehei, S. Vaseghi, and Qin Yan, "Noisy speech enhancement using harmonic-noise model and codebook-based post-processing," *IEEE Trans. on Audio, Speech, and Lang. Processing*, vol. 15, no. 4, pp. 1194–1203, 2007.

[21] Y. Stark and J. Tabrikian, "MMSE-based speech enhancement using the harmonic model," in *Electrical and Electronics Engineers in Israel, 2008. IEEEI 2008. IEEE 25th Convention of*, 2008, pp. 626–630.

[22] T.F. Quatieri and R.J. McAulay, "Shape-Invariant Time-Scale and Pitch Modifications of Speech," *IEEE Trans. on Acoust., Speech and Signal Processing*, vol. 40, pp. 497–510, 1992.

[23] Y. Stylianou, J. Laroche, and E. Moulines, "High-Quality Speech Modification based on a Harmonic + Noise Model," *Proc. European Conference on Speech Communication and Technology (EUROSPEECH)*, 1995.

[24] Naotoshi Osaka, "Timbre interpolation of sounds using a sinusoidal model," in *Proceedings of the International Computer Music Conference*, 1995.

[25] Riccardo Di Federico, "Waveform preserving time stretching and pitch shifting for sinusoidal models of sound," in *Proceedings of the COST-G6 Digital Audio Effects Workshop*, 1998, pp. 44–48.

[26] G. P. Kafentzis, G. Degottex, O. Rosec, and Y. Stylianou, "Time-scale Modifications based on an Adaptive Harmonic Model," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2013, pp. 8193–8197.

[27] M. Goodwin, "Matching pursuit with damped sinusoids," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1997, pp. 2037–2040.

[28] T. S. Verma and T. H. Y. Meng, "Sinusoidal modeling using frame-based perceptually weighted matching pursuits," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1999, pp. 981–984.

[29] R. Heusdeuns, R. Vafin, and W. B. Kleijn, "Sinusoidal modeling using psychoacoustic-adaptive matching pursuits," *IEEE Signal Processing Letters*, vol. 9, no. 8, 2002.

[30] Y. Stylianou, *Harmonic plus Noise Models for Speech, combined with Statistical Methods, for Speech and Speaker Modification*, Ph.D. thesis, E.N.S.T - Paris, 1996.

[31] Y. Pantazis, O. Rosec, and Y. Stylianou, "Adaptive AM-FM signal decomposition with application to speech analysis," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, pp. 290–300, 2011.

[32] G. Degottex and Y. Stylianou, "Analysis and synthesis of speech using an adaptive full-band harmonic model," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 10, pp. 2085–2095, 2013.

[33] J. Nieuwenhuijse, R. Heusdens, and E. Deprettere, "Robust exponential modeling of audio signals," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1998, pp. 3581–3584.

[34] J. Jensen, R. Heusdens, , and S. Jensen, "A perceptual subspace approach for modeling of speech and audio signals with damped sinusoids," *IEEE Transaction on Speech and Audio Processing*, vol. 12, no. 2, pp. 121–132, March 2004.

[35] R. Badeau, B. David, and G. Richard, "A new perturbation analysis for signal enumeration in rotational invariance techniques," *IEEE Transaction on Signal Processing*, vol. 54, no. 2, pp. 492–504, February 2006.

[36] G. Zhou, G.B. Giannakis, and A. Swami, "On polynomial phase signals with time-varying amplitudes," *Signal Processing, IEEE Transactions on*, vol. 44, no. 4, pp. 848–861, 1996.

[37] Yinong Ding and Xiaoshu Qian, "Processing of musical tones using a combined quadratic polynomial-phase sinusoid and residual (quasar) signal model," *Journal of the Audio Engineering Society*, vol. 45, no. 7/8, pp. 571–584, 1997.

[38] L. Girin, S. Marchand, Joseph Di Martino, A. Robel, and G. Peeters, "Comparing the order of a polynomial phase model for the synthesis of quasi-harmonic audio signals," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2003, pp. 193–196.

[39] Sean A. Fulop and Kelly R. Fitz, "Algorithms for computing the time-corrected instantaneous frequency (reassigned) spectrogram, with applications," *Journal of the Acoustical Society of America*, vol. 119, no. 1, pp. 360–371, 2006.

[40] Jeremy J. Wells and Damian T. Murphy, "High-accuracy frame-by-frame non-stationary sinusoidal modelling," in *Proc. International Conference on Digital Audio Effects (DAFx)*, 2006, pp. 253–258.

[41] Saso Musevic and Jordi Bonada, "Generalized reassignment with an adaptive polynomial phase fourier kernel for the estimation of nonstationary sinusoidal parameters," in *Proc. International Conference on Digital Audio Effects (DAFx)*, 2011.

[42] Sylvain Marchand, "The Simplest Analysis Method for Non-stationary Sinusoidal Modeling," in *Proc. International Conference on Digital Audio Effects (DAFx)*, 2012, pp. 23–26.

[43] Sylvain Marchand and Philippe Depalle, "Generalization of the derivative analysis method to non-stationary sinusoidal modeling," in *Proc. International Conference on Digital Audio Effects (DAFx)*, 2008.

[44] B. Hamilton and P. Depalle, "A unified view of non-stationary sinusoidal parameter estimation methods using signal derivatives," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2012, pp. 369–372.

[45] S. Musevic and J Bonada, "Comparison of non-stationary sinusoid estimation methods using reassignment and derivatives," in *Sound and Music Computing Conference*, 2010.

[46] G. P. Kafentzis, Y. Pantazis, O. Rosec, and Y. Stylianou, "An extension of the adaptive quasi-harmonic model," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2012, pp. 4605–4608.

[47] M. Caetano, G. Kafentzis, G. Degottex, A. Mouchtaris, and Y. Stylianou, "Evaluating how well filtered white noise models the residual from sinusoidal modeling of musical instrument sounds," in *Proc. Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2013.

[48] C. Borss and R. Martin, "On the construction of window functions with constant-overlap-add constraint for arbitrary window shifts," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2012, pp. 337–340.

[49] K. Hermus, W. Verhelst, P. Lemmerling, P. Wambacq, and S. van Huffel, "Perceptual audio modeling with exponentially damped sinusoids," *Signal Processing*, vol. 85, no. 1, pp. 163–176, January 2005.

[50] Mathieu Lagrange, Roland Badeau, Bertrand David, Nancy Bertin, Jose Echeveste, Olivier Derrien, Sylvain Marchand, and Laurent Daudet, "The DESAM toolbox: spectral analysis of musical audio," in *Proc. International Conference on Digital Audio Effects (DAFx)*, 2010, pp. 254–261.