CONTINUOUS STATE MODELING FOR STATISTICAL SPECTRAL SYNTHESIS

Tim-Tarek Grund and *Henrik von Coler*

Audio Communication Group
TU Berlin
voncoler@tu-berlin.de

ABSTRACT

Continuous State Markovian Spectral Modeling is a novel approach for parametric synthesis of spectral modeling parameters, based on the sines plus noise paradigm. The method aims specifically at capturing shimmer and jitter - micro-fluctuations in the partials' frequency and amplitude trajectories, which are essential for the timbre of musical instruments. It allows for parametric control over the timbral qualities, while removing the need for the more computationally expensive and restrictive process of the discrete state space modeling method. A qualitative comparison between an original violin sound and a re-synthesis shows the ability of the algorithm to reproduce the micro-fluctuations, considering their stochastic and spectral properties.

1. INTRODUCTION

1.1. Spectral Modeling

Sounds of musical instruments can be modeled as a combination of sinusoidal and noise-like components. Spectral modeling methods generally perform an analysis of the spectrum of an input signal in order to separate the deterministic, tonal content from the stochastic, while in some cases transients are also considered separately [1]. On the basis of the spectral examination, an output signal can be re-synthesized, with additional means for manipulations. While early methods modeled the tonal as well as the stochastic components using additive synthesis, later models used a different approach for noise-like signal content. The Deterministic plus Stochastic Model [2] expresses any sound as a sum of sinusoids with individual time-varying amplitudes $A_r(t)$ plus a residual noise component $e(t)$, which is modeled by a time-varying filtering of white noise.

$$s(t) = \sum_{r=1}^{R} A_r(t) \cos(\theta_t) + e(t) \tag{1}$$

The Deterministic plus Stochastic Model can be simplified for modeling harmonic sounds, for which each sinusoid is derived from integer multiples of the fundamental frequency. These can be referred to as partials.

$$s_{\text{harm}}(t) = \sum_{r=1}^{R} A_r(t) \cos(2\pi r f_0 t + \phi_r) + e(t) \tag{2}$$

Within the proposed method, the harmonic synthesis of the deterministic content constitutes the basis for modeling the tonal content. The modeling of the stochastic spectral content is outside the scope of this paper.

The original sines plus noise model can re-synthesize musical sounds with high quality and offers extensive means of manipulation. However, spectral models rely on a large set of parameters, making it it challenging to apply them in settings with few control parameters, for example in expressive performance. Ongoing research thus deals with approaches which allow a more direct control or parameter management. An extended source-filter model, presented by Hahn et al. [3], models a database of instrument sounds with different pitches and intensities. The deterministic part is based on a non-white source and a resonator filter. Parameters are modeled by tensor product B-splines (basic-splines), covering the sounds' temporal evolution. The DDSP approach [4] combines classic signal processing with deep learning methods. The end-to-end learning approach enables independent control of loudness and pitch, dereverberation and timbre transfer [5].

The method presented in this work aims at capturing a data set of instrument recordings, based on a statistical analysis of the spectral modeling parameters. Resulting models can be used for expressive real-time synthesis, allowing an interpolation between the data set's samples and different timbres. Statistical spectral modeling grants direct control over the micro-fluctuations.

Irregularities of the amplitude trajectory are generally referred to as *shimmer*, while irregularities within the frequency trajectory are denoted as *jitter*. These fluctuations contribute to the individual timbre of an instrument and are essential for the perceived sound quality of synthesis results [6].

1.2. Stateless Modeling

Statistical spectral modeling aims at capturing the timbre of musical sounds by means of measuring the distribution of spectral modeling parameters. A first implementation [7] captured the distribution functions of amplitude and frequency trajectories for single partials, as shown in Figure 1 for a partial's amplitude. New trajectories could be synthesized with this distribution using the inverse transform sampling method [8], followed by a low-pass filter smoothing.

1.3. Discrete State Modeling

An extended version of the stateless approach models parameter trajectories as Markov processes [9]. It hence captures the distribution properties and spectral properties, without the smoothing needed in the stateless approach. Instead of capturing a single distribution for a parameter, transition probabilities are calculated for a parameter trajectory with length L, quantized with $i = j$ steps:

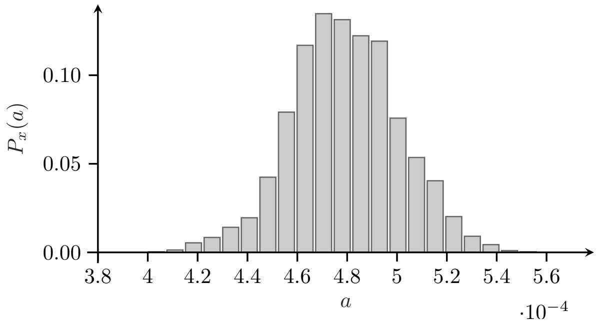


Figure 1: Distribution of a partial's amplitude [9].

$$\text{PMF}(i,j) = \frac{1}{L}\,|\{x[n] \mid x[n+1] = x_j\}|,\; n = 1 \dots L \quad i = 1 \dots 21 \tag{3}$$

This procedure results in a transition probability matrix $PMF(i,j)$, as shown in Figure 2. These matrices can be used for generating a stochastic process with the properties of the original trajectory. Both the stateless and the discrete state modeling allow to interpolate between samples from the analysis data set [7].

2. CONTINUOUS STATE MODELLING

Although the discrete state model presented in the previous chapter is well suited for modeling and synthesizing musical instrument sounds, it has several drawbacks. While it is able to create means of expressive sound synthesis utilizing the intensity dimension of the timbral plane, it lacks the means of altering the micro-structure of frequency and amplitude trajectories. The proposed method however allows for a parametric control of shimmer and jitter.

Parameter trajectories can be modeled as sequences governed by Markov processes. This interpretation could potentially yield more natural sounding synthesis results compared to low-pass filtered white noise disturbed parameter trajectories.

Another potential benefit of this method are the real time morphing capabilities that emerge from the possibility of parametric control over the distribution of events.

Central to this model is the algorithm to for the parametric generation of frequency and amplitude trajectories. Currently, there are two different modes used to mimic the stable trajectory behaviour of real sound sources. These, the Scaled Normal and the Skew Normal model, are both explained in detail later on. Within the Scaled Normal method, the parameter *mean* is parametrized using Markov chains, while for the Skew Normal method the *skew* of the distributions is parametrized this way.

To create a waveform from the trajectories it is necessary to interpolate between the support points. Here a cubic interpolation is used in order to avoid rapid changes in the phase trajectories. In this manner waveforms for each partial can be created.

At this point, it is possible to multiply each partial waveform with a constant partial amplitude in order to preserve an original partial amplitude relationship of a source sound. These individual waveforms can now be added together.

Based on the Markovian approach for spectral modeling synthesis, a parametric algorithm is developed. This evolution has several benefits. The parametric nature allows changes to the sounds properties during run-time and it consumes less memory for storing a model.

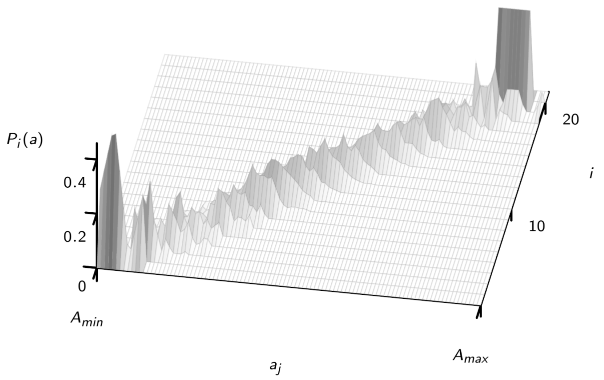


Figure 2: Transition probability matrix for a partial amplitude trajectory [9].

2.1. Parametrization of Mean

In this model, every support point is drawn from a normal distribution with the parameters μ and σ. While σ is freely adjustable, the mean μ of any following support points is dependent on a linear combination of the overall mean x_{mean} and the value of the last support point x_i, with the parameters α and β scaling the influence of each component.

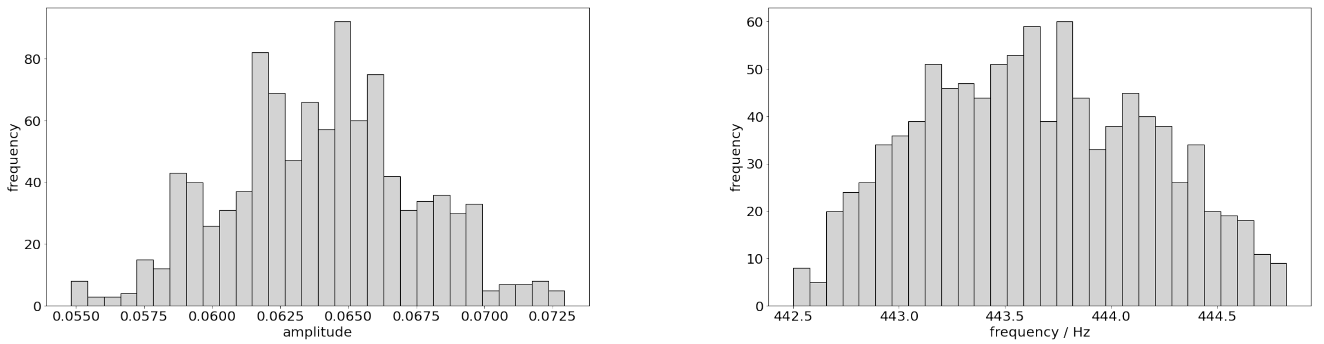
$$\begin{aligned} \mu_{i+1} &= \alpha \cdot x_{\text{mean}} + \beta \cdot x_i \\ \alpha + \beta &= 1 \\ x_{i+1} &\sim \mathcal{N}(\mu_{i+1}, \sigma), \\ \mu_0 &= x_{\text{mean}} \end{aligned} \tag{4}$$

For $\alpha = 1$, the resulting trajectory will be a normally distributed trajectory around the overall mean x_{mean}, for $\beta = 1$ the algorithm will produce an unstable trajectory; a random walk.

2.2. Parametrization of Skewness

For this model, the value of every support point is drawn from a *Skew Normal distribution* with parameters μ, σ and θ. The parameter μ of the following states distribution is solely dependent on the last state of the sequence. The *skew* θ of any following support point is dependent on the difference between the overall mean (target value) and the value of the last support point. In this model, the parameter gamma is used to scale the influence of the deviation of the last state x_i from the target value x_{mean}. The further the last state is away from the target state (and the higher the value of gamma), the more skewed will the one-step transition density be in the direction of the target value, resulting in a likely transition of states towards the center. Both the parameter γ and σ are freely controllable in this algorithm, although σ_{f0} will be multiplied by a factor corresponding to the partial order, so as to maintain a constant partial frequency to standard deviation ratio.

$$\begin{aligned} \mu &= x_i, \\ \theta_{i+1} &= -\gamma * (x_i - x_{\text{mean}})\,,\ \gamma \in [0, \infty] \end{aligned} \tag{5}$$



(a) Histogram of amplitude trajectory of 1st partial of the source sound.

(b) Histogram of frequency trajectory of 1st partial of the source sound.

Figure 3: Original distributions of source material.

For $\gamma = 0$ the distribution of the following value will be a normal distribution without skew, also resulting in an unstable trajectory.

The method to produce Skew Normal distributed random variables is based on a procedure by Henze [10]. Here, two uniformly distributed random numbers U and V suffice to generate a random variable Z_θ, which has the Skew Normal distribution.

$$Z_\theta = \frac{\theta}{\sqrt{1+\theta^2}}|U| + \frac{1}{\sqrt{1+\theta^2}}V \sim \mathcal{SN}(\mu, \sigma, \theta) \qquad (6)$$

3. ANALYSIS

For the analysis phase, the *TU-Note Violin Sample Library* [11] is used as source material. The library contains 336 single sound items and 344 two-note sequences. Within the scope of this project, only the single sound items are used, which consists of 84 pitches in four different dynamics. While the material is provided at a sampling frequency of 96 kHz with a resolution of 24 bit, the sampling frequency has been altered to 44.1 kHz in order to use the Spectral Modeling Synthesis Tools (SMS-Tools) [12]. The SMS-Tools are a set of software tools for sound analysis, transformation and resynthesis written in Python and C.

Before the sound items are analyzed, they need to be preprocessed. TU violin single sound items are provided with manually annotated segmentation documentation, which contain the time stamps for on- and offsets of attack, sustain and release segments via four points A, B, C and D. The sustain part of each sound item is contained with in the space bounded by points C and D. All sound items are prior to the following analysis stages segmented to the sustain part. Modeling attack and release segments is outside the scope of this paper.

The SMS-Tools are employed at this point to extract the frequency and amplitude trajectories of each partial per sound item. To this end, the segmented sustain parts of the single sound items are analysed using the `harmonicModelAnal`-function of the SMS-Tools, utilizing the sinusoidal harmonic model with a fast Fourier Transform (FFT) size of 2048 samples and a hop size of 128 samples. The harmonic analysis yields the frequency, amplitude and phase trajectory for each partial. As the original phases of each partial are not relevant to the synthesis algorithm, they are discarded at this step. Since the amplitude trajectory is returned in decibels, it becomes necessary to convert it.

To further investigate the trajectories, both the amplitude and the frequency trajectories are subjected to an outlier removal eliminating all trajectory values twice the standard deviation in order to account for errors within the peak continuation. From the remaining trajectory values the mean and the standard deviation as well as the trajectory histograms are calculated. Subsequently, the mean value is subtracted from the trajectories to remove the impact of the 0 Hz bin, which eases the calculation of relevant spectral features. Now, spectral centroid, spectral flatness, as well as the lower and upper spectral roll-off frequency (at 15% and 85%) can be calculated for trajectories of each partial of every sound item. For these spectral features the absolute error between each partial of the original material and the synthesized sound can be calculated.

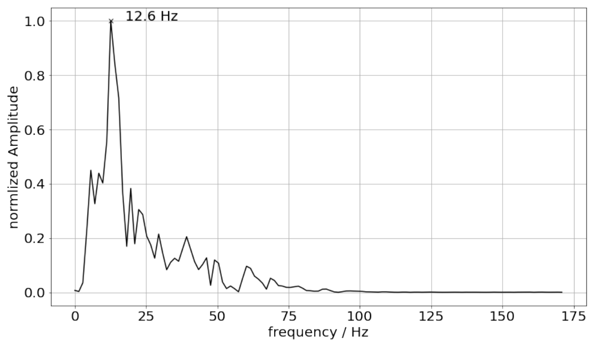
For the spectral analysis, both trajectories are subject to a high-pass FIR filter using the window method with the cutoff frequency at 5 Hz. The employed window is a Blackman window. Since the analysis and the synthesis stage are separated, real-time analysis is not needed. This permits the use of filters of higher order, which is why the filter order used here is 801. As the trajectories were created using a hopsize of 128 samples, the sampling frequency of the parameter trajectories can be calculated as

$$f_{s,t} = \frac{f_{s,x}}{n_{\text{hop}}} = \frac{44\,100\,\text{Hz}}{128} = 344.531\,25\,\text{Hz}. \qquad (7)$$

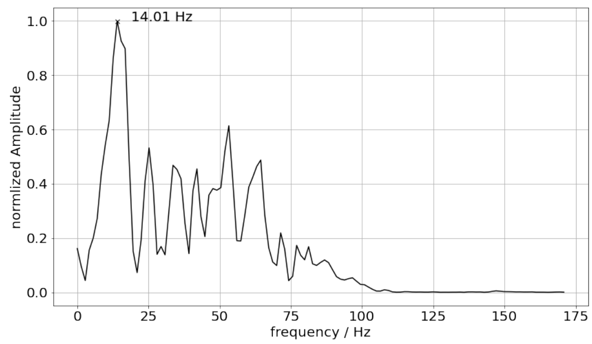
Within the scope of this paper sound item 60 will be used as the single source sound, against which the two generated sound items will be compared. This corresponds to a 443.00 Hz tone with the *fortissimo* dynamic.

The stochastic analysis provides the mean μ and the standard deviation σ for both trajectories for each partial for each sound item. The mean of the frequency trajectory is discarded at this point due to the harmonic nature of the synthesis algorithm. The amplitude mean as well as both standard deviations are stored for later use in the synthesis process, but only the amplitude mean will be used. This is for the reason that at the time of writing no reasonable way of transforming the standard deviations from a statistical measure of the whole sound parameter trajectory into a Markovian model parameter has been identified.

Within Figure 3b an approximately normal distribution of states of the frequency trajectory can be seen. The amplitude trajectory states however seem to follow a more irregular distribu-



(a) Normalized amplitude trajectory FFT of the source material.



(b) Normalized frequency trajectory FFT of the source material.

Figure 4: Original Spectra of source material.

tion with multiple peaks, as can be seen in Figure 3a. The mean frequency for the source sound is 443.64 Hz rounded to two decimal places, which is slightly above the assigned note frequency of 443.00 Hz.

Fourier transformations of both amplitude and frequency trajectories are calculated and subsequently peak-normalised. The FFT of the amplitude trajectory in Figure 4a has its highest peak at around 12.6 Hz and continues to decrease until it approaches 0 at around 100 Hz. The FFT of the frequency trajectory in Figure 4b has its highest peak at around 14 Hz. Afterwards, it falls approaching 0 at around 125 Hz, with a small but prominent peak at around 145 Hz.

4. RESYNTHESIS

4.1. Method

Analyzed sounds can be re-synthesized using the parameter trajectories derived from the continuous space Markovian spectral modeling. For each partial r a unique frequency trajectory $f_{\text{traj, r}}(t)$ and a unique amplitude trajectory $A_{\text{traj, r}}(t)$ are generated.

The unique partial frequency trajectory and amplitude trajectory are produced by the aforementioned methods of parametrization of mean and parametrization of the skewness.

For the parametrization of mean, each new trajectory support point is drawn from a normal distribution, with the mean parameter being calculated by weighting the last state and the target state with the weights α and β, with a second model parameter being the standard deviation. Regarding the frequency trajectory, the starting value for the values drawn from the last state is substituted by the target state, which is the frequency of the current partial. The standard deviation for each partial is an integer multiple of the standard deviation of the fundamental frequency referring to the partial order. For the amplitude trajectory, the starting value for the values drawn from the last state as well as the target state is simply 1 and the standard deviation stays the same for all partial amplitudes. The parameters for the Scaled Normal Markovian model can be found in Table 1.

For the parametrization of skewness, every new support point is drawn from a Skew Normal distribution, where the mean parameter serves as the last state, and the skew parameter is governed by the distance of the last state to the target state, multiplied by a weight γ. Concerning the frequency trajectory, the standard deviation is again an integer multiple of the fundamental frequency standard deviation equal to the partial order, the target state being the partial frequency. For the amplitude trajectory, it is again the same standard deviation for all partials, with the target state being 1. The starting value of the values drawn from the last state is again substituted by the target state for both trajectories. The parameters for the Skew Normal Markovian model can be found in Table 2.

Table 1: Parameters for the Scaled Normal Markovian model.

Parameter	Value
μ_{f0}	$\alpha_{f0} \cdot f_{\text{last state}} + \beta_{f0} \cdot f_{\text{target state}}$
μ_{amp}	$\alpha_{\text{amp}} \cdot A_{\text{last state}} + \beta_{\text{amp}} \cdot A_{\text{target state}}$
σ_{f0}	0.004
σ_{amp}	0.02
α_{f0}	0.0001
α_{amp}	0.001

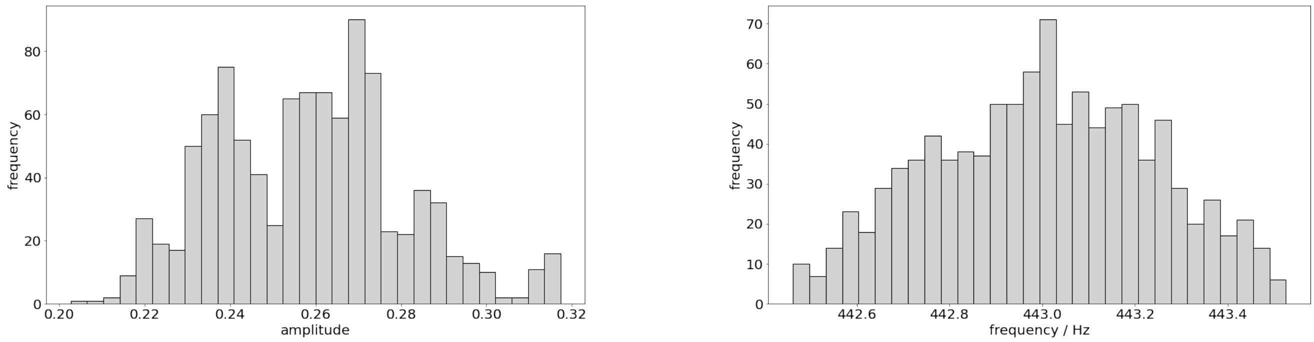
Table 2: Parameters for Skew Normal Markovian model.

Parameter	Value
μ_{f0}	$f_{\text{last state}}$
μ_{amp}	$A_{\text{last state}}$
σ_{f0}	0.004
σ_{amp}	0.03
γ_{f0}	1
γ_{amp}	0.8

Since the amplitude trajectory starting point for each partial is 1, it is imperative to scale the resulting waveform by the mean amplitude of the partial $A_{const,r}$ extracted in the earlier analysis step.

Another important variable in the synthesis process is the distance between support points. A smaller distances will lead to more rapid changes within the trajectories. The distance used in this synthesis context is 512 samples.

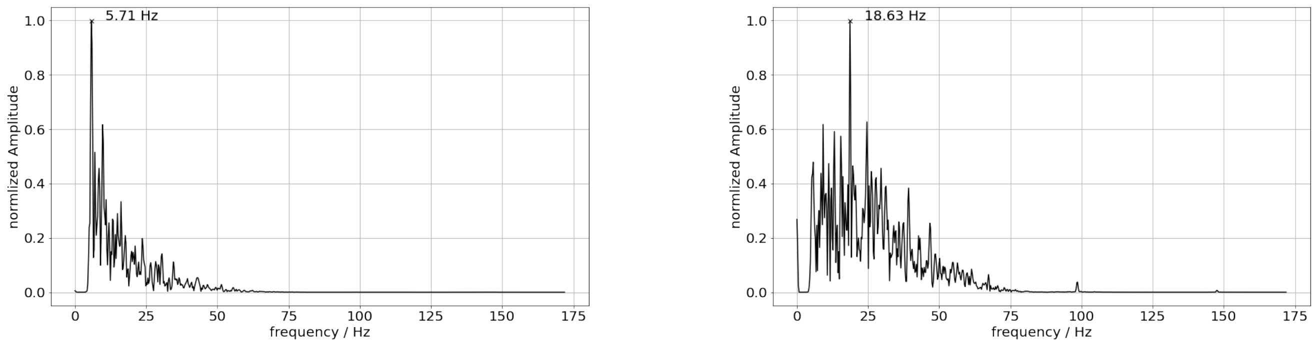
After interpolating between the support points, we can syn-



(a) Histogram of amplitude trajectory of 1st partial of the synthesized sound (Scaled Normal).

(b) Histogram of frequency trajectory of 1st partial of the synthesized sound (Scaled Normal).

Figure 5: Distributions of synthesis result (Scaled Normal).



(a) Normalized amplitude trajectory FFT of the Scaled Normal synthesized material.

(b) Normalized frequency trajectory FFT of the Scaled Normal synthesized material.

Figure 6: Spectra of synthesis result (Scaled Normal).

thesize the sound by creating and summing the waveforms for all partials using the following equation:

$$s_{\text{synth}}(t) = \sum_{r=1}^{R} A_{\text{const},r} A_{\text{traj},r}(t) \cos\left(2\pi f_{\text{traj},r}(t) \cdot t\right) \qquad (8)$$

Synthesis is performed in the time-domain, not frame-by-frame but rather array-wise: The trajectories themselves are created frame-by-frame, resulting in a frequency and an amplitude trajectory.

4.2. Resynthesis Properties

In this section parameters of the synthesized violin sounds are analyzed in the same manner as the original sound item. The preprocessing of the synthesized violin sounds is identical to the preprocessing of the TU-Note violin sound items.

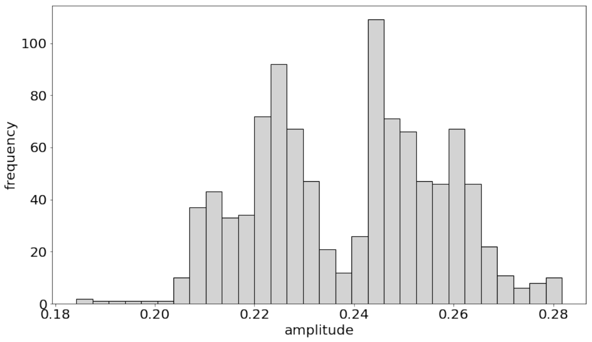
The Figures 5b and 7b both show an approximate normal distribution of the frequency trajectory of the synthesized sounds. The mean frequency across both synthesis methods is 443.00 Hz rounded to two decimal places. Visible in the histogram of the amplitude trajectory of Scaled Normal (Figure 7a) and of the Scaled Normal synthesized sounds (Figure 5a) are different distributions: In the graph for the Skew Normal method a bimodal distribution becomes apparent, in the graph for the Scaled Normal method a more irregular, multimodal distribution can be seen.

Figure 6a shows the fast Fourier transformation of the trajectory of the amplitude of the sound material synthesized using the Scaled Normal Markovian modeling. Here the highest peak is visible at around 5.7 Hz, after which the spectrum falls until it approaches 0 at around 75 Hz.

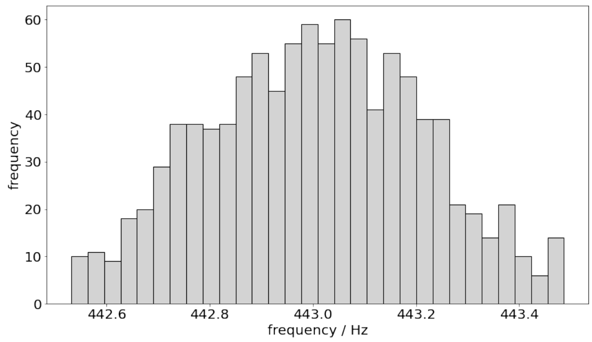
The frequency trajectory FFT in Figure 6b of that method follows a similar pattern, however with several peaks between 5 Hz - 25 Hz, with the highest peak at 18.63 Hz, after which it decays until it approaches 0 at around 75 Hz. However, two small but notable peaks at around 95 Hz and 145 Hz can be identified.

Regarding the sound material of the Skew Normal synthesis, the FFT of the amplitude trajectory in Figure 8a behaves similarly to the one of the Scaled Normal synthesis: its highest peak rests at around 10.6 Hz. It shows a decline thereafter, approaching 0 at around 75 Hz.

The frequency trajectory FFT of the Skew Normal synthesis in Figure 8b also follows the frequency trajectory FFT of the scale Normal synthesis closely: A region of high peaks between 6 Hz - 25 Hz, with the highest peak at around 7.2 Hz. After that, it

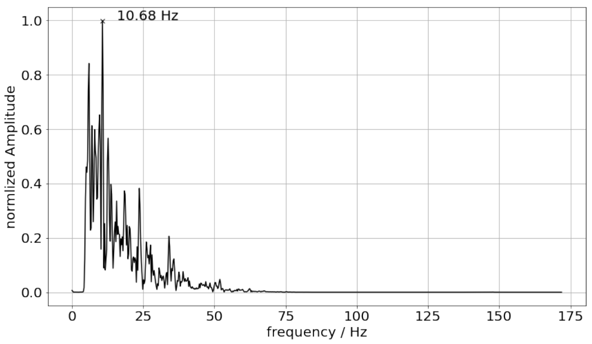


(a) Histogram of amplitude trajectory of 1st partial of the synthesized sound (Skew Normal).

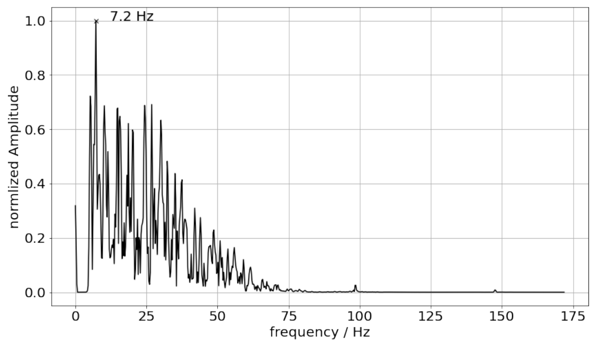


(b) Histogram of frequency trajectory of 1st partial of the synthesized sound (Skew Normal).

Figure 7: Distributions of synthesis result (Skew Normal).



(a) Normalized amplitude trajectory FFT of the Skew Normal synthesized material.



(b) Normalized frequency trajectory FFT of the Skew Normal synthesized material.

Figure 8: Spectra of synthesis result (Skew Normal).

approaches 0 at around 80 Hz with two notable peaks at around 95 Hz and 145 Hz.

5. COMPARISON

5.1. Single Item Comparison

When comparing the frequency distributions from the synthesized sounds (Figures 5b and 7b) to the frequency distribution of the source material (Figure 3b), we can see that although the mean frequency is higher for the original material, all three trajectories seem to follow a similar distribution. However, for the amplitude trajectories the Figures 3a, 7a and 5a show that all three amplitude trajectories follow a different distribution form. While all have in common, that they do not follow a normal distribution, the value ranges leave room for discussion. Since both the Scaled Normal and the Skew Normal sound material were subject to a normalization, the individual partial values differ considerably between the amplitude values of the original and the synthesized material. However, when scaled up to a similar level of amplitude, the standard deviation of the amplitude trajectory of the 1st partial of the source material becomes 0.013, while the standard deviations of the synthesized material are 0.022 for the Scaled Normal and 0.018 for the Skew Normal synthesis method. This means, that the synthesized distributions are wider than the original distribution. The irregular distributions of the amplitude trajectories of the synthesized material are most probably impacted by a Markovian random walk. This is to be expected since the influence of the relevant parameter on containg the effect of a random walk (α for the Scaled Normal model and γ for the Skew Normal model) has been decreased compared to the synthesis of the frequency trajectories.

In the previous section, similarities between the two synthesized sound items have already been highlighted. Furthermore, there are similarities with the FFTs of the source sound item, too: All three share the highest peak within their respective amplitude trajectory FFTs in the region between 5 Hz - 25 Hz, with a generalised decline until they approach 0 at around 75 Hz for the synthesized sounds and 100 Hz for the source sound. The frequency trajectories also follow a similar makeup: A region of highest peaks followed by a decline approaching 0 at around 90 Hz for the synthesized sound items and 125 Hz for the source sound are included in all three spectra. The difference in frequency at which the FFT approaches 0 between the plateaus of source sound and the synthesized sounds can perhaps be explained by a difference in nature: since the source sound trajectory is based on a recording, it might be susceptible to recording noise, in contrast to the digitally synthetic nature of the synthesized sound items.

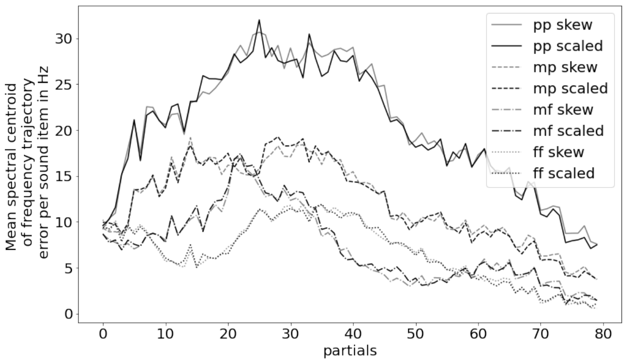


Figure 9: Mean error of spectral centroid between source material and synthesized material across sound items.

5.2. Comparison by Dynamic Level

In order to evaluate the capabilities of the analysis-synthesis approach for the complete sample library, the relationship between dynamic level of the source material, partial order, synthesis mode and deviations between the spectral features of the trajectories can be investigated. The spectral centroid of the frequency trajectory averaged across all sound items is shown in Figure 9. It becomes apparent that for lower dynamics the deviations from the source material are larger than for higher dynamics. It is also becomes evident that within one dynamic group the differences between the Skew Normal and Scaled Normal synthesis are negligible.

6. CONCLUSIONS

Continuous state Markovian spectral modelling has been proposed as a novel approach for data-driven spectral synthesis. The method aims at capturing the micro-fluctuations of sinusoidal parameters, allowing the control of jitter and shimmer. Stochastic and spectral similarities have been identified for a single instrument sound, potentially validating the proposed method and showing the limits of the heuristically tuned synthesis parameters. Notably for the multiplicative increase of the standard deviation across both synthesis algorithms, it can be said that analytic deduction or evolutionary tuning of this parameter could provide more realistic results. Since at the time of writing there is no sensible transformation of the analyzed standard deviations of the frequency and amplitude trajectory into a standard deviation to be used within the Markovian modelling, the next step would be to identify measures that would result in a more truthful representation of the original sound. Possible actions for the future within the context of this project are a more thorough numerical comparison between the source sound and the synthesized sound items. In order to answer the question, whether a a low-pass filtered white noise could potentially yield more convincing results, a listening test can be employed. Artistic and expressive use of the presented algorithms could further be explored in a user study with real time synthesis control.

7. REFERENCES

[1] Julius O. Smith, "Spectral Audio Signal Processing," Available at http://ccrma.stanford.edu/~jos/sasp/, accessed 22.03.2022, online book, 2011 edition.

[2] Xavier Serra, "Musical sound modeling with sinusoids plus noise," in *Musical Signal Processing*, Curtis Roads, Stephen Travis Pope, Aldo Piccialli, and Giovanni De Poli, Eds., pp. 91–122. Lisse, the Netherlands, 1997.

[3] Henrik Hahn and Axel Röbel, "Extended source-filter model for harmonic instruments for expressive control of sound synthesis and transformation," in *Proceedings of the 16th International Conference on Digital Audio Effects (DAFx)*, Maynooth, Ireland, 2013.

[4] Jesse Engel, Lamtharn Hantrakul, Chenjie Gu, and Adam Roberts, "DDSP: Differentiable digital signal processing," *International Conference on Learning Representations (ICLR)*, 2020.

[5] Francesco Ganis, Erik Frej Knudesn, Søren VK Lyster, Robin Otterbein, David Südholt, and Cumhur Erkut, "Real-time timbre transfer and sound synthesis using DDSP," *arXiv preprint arXiv:2103.07220*, 2021.

[6] Akira Nishimura, Mitsumi Kato, and Yoshinori Ando, "The relationship between the fluctuations of harmonics and the subjective quality of flute tone," *Acoustical Science and Technology*, vol. 22, no. 3, pp. 227–238, 2001.

[7] Henrik von Coler, "Statistical sinusoidal modeling for expressive sound synthesis," in *Proceedings of the International Conference of Digital Audio Effects (DAFx)*, Birmingham, UK, 2019.

[8] Luc Devroye, *Non-Uniform Random Variate Generation*, Springer, McGill University, 1986.

[9] Henrik von Coler, *A System for Expressive Spectro-spatial Sound Synthesis*, Ph.D. thesis, TU Berlin, 2021.

[10] Norbert Henze, "A probabilistic representation of the 'skew-normal' distribution," *Scandinavian Journal of Statistics*, vol. 13, no. 4, pp. 271–275, 1986.

[11] Henrik von Coler, Jonas Margraf, and Paul Schuladen, "Tu-note violin sample library," Available at http://dx.doi.org/10.14279/depositonce-6747, 2018.

[12] Xavier Serra, "Spectral modeling synthesis tools," Available at https://www.upf.edu/web/mtg/sms-tools, accessed 29.03.2022, 2013.