An FM-Wavetable-Synthesized Violin with Natural Vibrato and Bow Pressure

Qin Liu[1,*]

[1]Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain
* qin.liu01@estudiant.upf.edu

Abstract. As a matter of fact, the digital representation of musical instruments has long been a challenge for researchers, with the violin being one of the most complex due to its expressive nature and intricate timbre. This study delves into the synthesis of a digital violin, aiming to reproduce its rich sound by incorporating natural vibrato and bow pressure based on performance statistics. Utilizing wavetable and FM synthesis techniques, the research has been able to enhance the fixed spectrum of the violin to yield a more authentic sonic experience. Vibrato and bow pressure elements, once missing in many synthesizers, have been introduced to bring about greater expressiveness to the synthesized sound. The results of our research indicate a marked improvement in the naturalness and expressiveness of the digital violin, although achieving full realism remains a challenge. The advancements made in this study significantly contribute to the field of digital music synthesis, bridging the gap between traditional instruments and their digital counterparts.

Keywords. Digital Violin Synthesis, Wavetable Synthesis, FM Synthesis, Vibrato, Bow Pressure, Expressive Control

1 Introduction

Most of the violin synthesizers on the market sound rigid, mechanical, and unnatural, often failing to capture the violin's nuanced timbre including the expressiveness of vibrato, complex bowing, etc. To address this issue, this project will utilize wavetable synthesis and FM synthesis to synthesize a digital violin and add natural vibrato and bow pressure according to the performance statistics, aiming at synthesizing realistic, natural sounding violin sound. Violin family sound synthesis is an age-old but uncommon research problem due to its more complex musical attributes. Commonly used sound synthesis methods include subtractive synthesis, additive synthesis, FM synthesis, wavetable synthesis, physical modeling synthesis, granular synthesis, sample-based synthesis, formant synthesis. For violin synthesis, the recent studies have mainly used the following three synthesis methods:

Physical modeling synthesis is a method of sound synthesis in which the waveform of the sound to be generated is computed using a mathematical model, a set of equations and algorithms to simulate a physical source of sound. For the violin, physical modeling synthesis would attempt to replicate the physical processes involved in producing sound from the instrument, which may include string vibration,

B. H. Ahmad (ed.), *Proceedings of the 2023 International Conference on Data Science, Advanced Algorithm and Intelligent Computing (DAI 2023)*, Advances in Intelligent Systems Research 180,
https://doi.org/10.2991/978-94-6463-370-2_27

bowing interaction, violin body resonance. Gabriel constructed a simple computer model of violin through physical acoustics, consisting of two arrays carrying the forward and backward waves on the string, appropriate reflection rules at the bridge and finger, a bow of finite width [1], and a combination of parallel two - pole filters to simulate the radioactivity of the violin body. Synthesis Toolkit in C++ (STK) implements digital waveguide modeling to simulate bowed string instruments, capturing the complex interactions between the string, the bow, and the instrument's body [2]. Maestre et al. presents the bowing control modeling framework applied to sound synthesis [3]. Greg presents an audio driven violin controller for real time synthesis, taking advantage of acoustical features extracted from the violin's signal [4].

Sample-based modeling synthesis, in its basic form, is able to capture individual sounds very accurately but does not offer manipulation techniques necessary for an expressive synthesis [5]. For example, extended commercial samplers like Vienna Symphonic Library provide a large number of possibilities in terms of articulations, pitches, etc., whilst explicit bowing control and vivid vibrato expression are still missing components. To realize the manipulation, Garritan Solo Stradivart and Synful Orchestra complete specialized spectral domain sound sample transformations oriented toward real-time expressiveness control using traditional MIDI-based keyboard interfaces. Henrik von Coler [6] combines sample-based modeling with sinusoidal modeling is capable of reacting to expressive gestures for sound manipulation, using the input parameters pitch and intensity. Janer and Maestre present a user-adapted system based on sample concatenation to synthesize sound relying on how subjects sing musical articulations for the violin and other instruments [7].

Recent research has demonstrated machine learning can be employed to learn the nuances and complexities of musical instruments and subsequently produce expressive synthesis. Most of the studies on existing expressive synthesis deal with piano because the piano does not allow a continuous control of the timbre, consequently considering a minimum number of parameters. Zhang et al. present a vector-quantized variational autoencoder (VQ-VAE) with mutual information minimization techniques and self-supervising strategies for content and style disentanglement for expressive piano performance [8], and Flossmann et al. propose YQX, a probabilistic piano performance rendering system based on Bayesian network theory [9].

However, there is rare research on fully automated machine learning-based expressive synthesis on the instruments of violin family due to its more complex musical attributes. The only one study founded is a decision tree-based violin synthesizer adapted from neural parametric singing synthesis technique proposed by Nizami and Lestari [10]. Most of the machine learning research on the violin is for analytical and pedagogical purposes. Li et al. created a dataset containing 11 expressional items, then chose duration, dynamic and vibrato note-level features to classify expressions using Support Vector Machine [11]. Lionello and Ramirez propose a machine learning approach to model vibrato in violin audio performances and a didactic real-time feedback system to assist students in learning vibrato [12]. Visentin et al. investigate the biomechanics of left-hand position changes (shifting) in violin performance [13], Zhao et al. researches the violinist identification based on vibrato [14], and Yang et al. also compare and analyze the vibrato between the Violin and Erhu [15].

The goal of the proposed system is to utilize wavetable synthesis, FM synthesis, and performance statistics to synthesize a digital violin sounding more realistic and add expressive control to the violin. Details on the motivations, analysis, and modeling are presented in Section 2. Section 3 shows the results of subjective qualitive evaluations. Section 5 concludes the contributions and lists perspectives for future developments.

2 Methodology

All code of the project[1] is implemented in Nyquist programming language. Nyquist is a language for sound synthesis and music composition, which can deal with both musical events and signal processing in a single integrated system.

2.1 Wavetable Synthesis Using Additive Synthesis Principles

Wavetable synthesis is a technique used in electronic music and sound design to produce musical tones by playing back and looping through sequences of waveforms. Instead of generating sound waveforms in real-time (like traditional synthesis methods), wavetable synthesis involves the playback of prerecorded waveforms stored in a table of list. Additive synthesis involves or "adding" individual sine waves often referred to as "harmonics" or "partials" together. Given that the spectrum of the violin is relatively consistent, not varying too much according to pitch and amplitude [16], so the project utilizes a fixed spectrum sampled from a substantial violin, empirically 30 harmonics. Furthermore, in the project, instead of generating the sound in real-time using a multitude of oscillators as in traditional additive synthesis, wavetable synthesis in the project employs a table of pre-generated waveforms by `build-harmonic()` function in Nyquist. The example is given in the link D4 note of violin table.wav.

[1] https://github.com/kenzhuti/FM-Wavetable-Synthesized-Violin-/tree/main

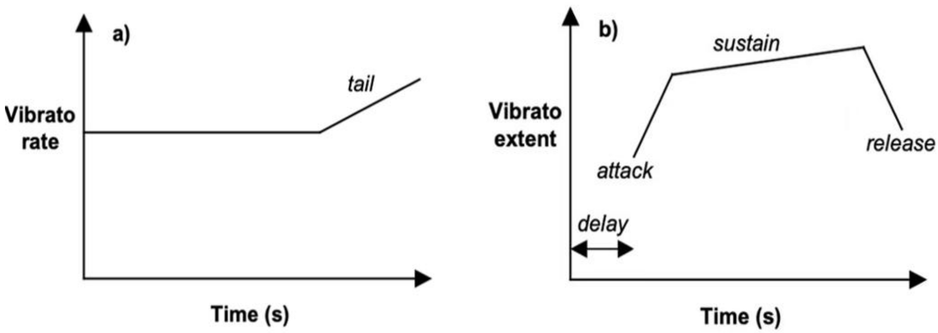

Fig 1. Vibrato Rate and Extend Envelopes (Photo/Picture credit: Original)

2.2 Vibrato

Vibrato plays an important role in the violin, adding warmth, richness, and many expressive subtleties to the performance. Vibrato has two key attributes: vibrato rate (frequency) and vibrato extent (amplitude, depth). Vibrato rate means how quickly the pitch fluctuates, and vibrato extent is the frequency deviation of the pitch. FM synthesis is a method wherein the frequency of a carrier waveform is modulated by the frequency of a modulator waveform, invented by John Chowning [17], which is commonly used for vibrato of instruments. Specifically, modulator frequency determines the vibrato rate, whereas modulator amplitude determines the vibrato extent. It is found in statistical research that vibrato rate ranges between 4 and 10 Hz with 5-6 more often, whilst vibrato extent varies between 0.20 and 0.35 semitones. Moreover, the vibrato in real performance is not constant, both vibrato rate and vibrato extent have regular envelopes with rules: The vibrato rate envelope encompass a constant part, followed by a tail with 15% increase over the 3 final vibrato cycles shown as Fig. 1(a), whereas vibrato extent envelope is made up of a delay (300 ms), attack (350 ms), sustain (duration minus the others), and release(300 ms) part shown in Fig. 1(b) for vibrato more than 1 second, and the note less than 0.35 second will not have vibrato [18]. A 1 Hz "random jitter" for vibrato is added to avoid sounding mechanical and evenly distributed. All the rules are applied in the code, aiming at producing natural and vivid vibrato.

2.3 Bow Pressure

The bow pressure contributes the most to the violin amplitude [19]. Thus, a normal bowing envelope is applied into the synthesized violin [20]. Concretely, bow attack time ranges between 0.08 and 0.3 second, bow release time ranges between 0.2 and 0.7, whilst bow sustain time is the time remained from bow duration, denoted in Fig. 2, leading to more realism.

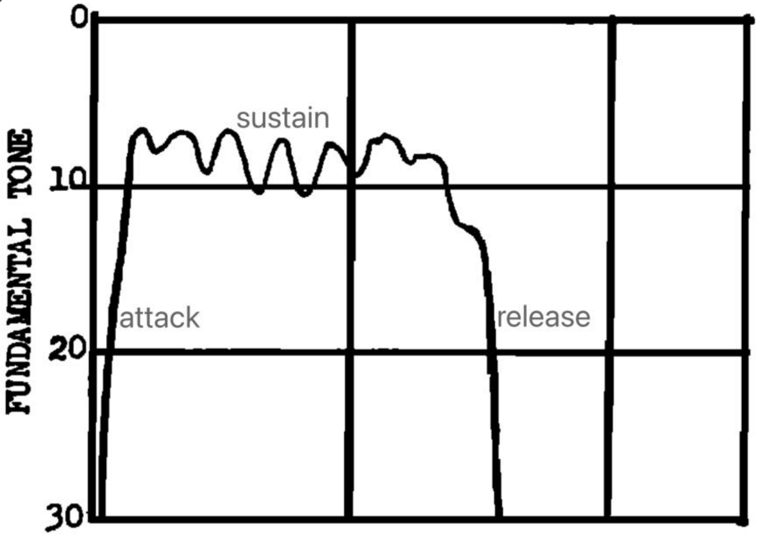

Fig 2. Bow Pressure Envelope (Photo/Picture credit: Original).

2.4 Score Demos

To demonstrate the overall performance, 2 score demos are composed. The first one is a Probabilistic Random Pentatonic Piece within Violin Register and Common Durations (score1.wav) in order to generate various scores for evaluations. The second one is Lightly Row (score2 ~ 0.8.wav, score2 ~ 0.34.wav), an A major piece from Suzuki Etudes, a preliminary textbook for violin beginners, which is handy for comparison with real performance.

3 Evaluations and Results

3.1 Objective

The primary intent of the evaluation was to evaluate the music quality of two distinct models: the baseline model bowed() model from STK [2] and this innovative model introduced by this project. The assessment revolved around three pivotal musical dimensions: Realism, Naturalness, and Expressiveness.

3.2 Method

To ensure an in-depth evaluation, six musicians with professional backgrounds were selected as subjects. These subjects were required to rate the models using a scale from 1 to 10 (with 10 being the highest). The models were evaluated on their ability to perform two identical musical pieces, as described in Section 2.4.

Six professionally educated musicians as subjects rated the two models using a scale from 1 to 10 (with 10 being the highest). Each subject evaluated the models based on their performance in playing two identical scores composed in Section 2.4. This approach offered a balanced representation of each model's capabilities in a standardized and controlled environment.

3.3 Results and Analysis

From the aggregated results in Table 1, the project model demonstrates a commendable edge over the baseline, specifically in the domains of naturalness and expressiveness. Nonetheless, this project slightly lagged behind the baseline in terms of realism. This dichotomy hints at the nuanced improvements brought by the project but also indicates areas that may require further refinement. To summarize, this evaluation underscores the progress made in the domain of violin synthesis. The advances showcased by the project model point to the evolving sophistication of musical instrument synthesis and its ever-increasing potential to capture the essence of real instruments.

Table 1. Subjective Evaluation Scores

	Baseline			**This Project**		
Model / **Subject**	**Realism**	**Naturalness**	**Expressiveness**	**Realism**	**Naturalness**	**Expressiveness**
No. 1	8	5	3	6	9	5
No. 2	7	3	6	7	6	7
No.	9	6	4	5	7	4

3						
No. 4	8	4	5	8	5	8
No. 5	8	5	4	4	8	8
No. 6	6	2	4	6	7	6
Average	7.7	4.2	4.3	6.3	7	6.3
Overall	5.4 = (7.7 + 4.2 + 4.3) / 3			6.53 = (6.3 + 7 + 6.3) / 3		

4 Conclusion

This study advanced the realm of musical synthesis with the introduction of a novel violin synthesizer capable of emulating human-like nuances in sound production. This synthesizer excels in generating a realistic violin sound, capturing the essence of human touch through vibrato and bow pressure dynamics. Furthermore, the provision to adjust vibrato frequently and bow pressure grants users greater control, enhancing the expressiveness of the produced music. However, certain limitations were observed. The current model does not support multiple notes with a single bow very well and lacks intricate performance techniques such as portamento, staccato, spiccato, détaché, pluck, and double stops. A crucial observation was that the consistent spectrum in the synthesizer, which remains invariant with pitch or technique, might be a potential factor in its reduced realism compared to the baseline.

For future improvements, except introducing the more performance techniques, approaches such as spectral interpolation synthesis, physical modeling synthesis, sampled-based synthesis, and spectral modeling synthesis could be considered. Furthermore, the integration of cutting-edge technologies, especially machine learning and DDSP (Differentiable Digital Signal Processing), offers promising avenues for enhancing bowed-string instrument synthesis. The research highlights the significance of capturing the intricate nuances of a musical instrument in synthesis and establishes a benchmark for future endeavors in this domain.

References

1. G. Weinreich, J. Acous. Soc. Ame., **74(S1)**, S52-S52 (1983).
2. P.R. Cook and G.P. Scavone, *The synthesis toolkit (stk)* (ICMC, London, 1999) pp. 1-131.
3. E. Maestre, IEEE Trans. on Audio, Speech, & Lang. Proce., **18(4)**, 855-871 (2010).

4. G. Kellum, *Violin Driven Synthesis from Spectral Models, in Music Technology Group* (UPF, Barcelona, 2007) pp. 1- 77.
5. R. Dannenberg, and I. Derenyi, J. of New Music Res. **27**, 211-238 (1998).
6. H. von Coler, *Statistical Sinusoidal Modeling for Expressive Sound Synthesis*. in Proceedings of the 22nd International Conference on Digital Audio Effects (DAFx), Birmingham, 2019.
7. J. Janer and E. Maestre, *Mapping Phonetic Features for Voice-Driven Sound Synthesis* (Springer, Heidelberg 2009).
8. H. Zhang and H. S. Dixon. *Disentangling the Horowitz Factor: Learning Content and Style From Expressive Piano Performance* n ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2023.
9. S. M. Flossmann, M. Grachten and G. Widmer, Guide to Comp, for Expre. Music Perf., 13, 75-98 (2013).
10. M. Nizami, and D. P. Lestari. *A DT-Neural Parametric Violin Synthesizer*, in 2021 International Conference on Electrical Engineering and Informatics (ICEEI), 2021.
11. P. C. Li, *Analysis of Expressive Musical Terms in Violin Using Score-Informed and Expression-Based Audio Features*, in International Society for Music Information Retrieval Conference, 2015.
12. M. Lionello, *Detection and Regression of Vibrato from Scores, Prediction of violin vibrato features from scores and rendering on mobile devices for didactic purposes*. Project Report, 2018.
13. P. Visentin, Peer J, **3**, e1299 (2015).
14. Y. Zhao, *Violinist identification based on vibrato features*, in 2021 29th European Signal Processing Conference (EUSIPCO), 2021.
15. LYang, E Chew and K Rajab, *Vibrato Performance Style: A Case Study Comparing Erhu and Violin* (Springer, Berlin, 2013).
16. C. E. Gough, *Violin acoustics* (Acoustics Today, London, 2016).
17. J. M. Chowning, J. Audio Eng. Soc., **21(7)**, 526-534 (1973).
18. E. Schoonderwaldt and A. Friberg. *Towards a rule-based model for violin vibrato. in Workshop on Current Research Directions in Computer Music* (Audiovisual Institute, Barcelona, 2001).
19. J. Woodhouse, Acta Acustica united with Acustica, **78** (1973).
20. H. Fletcher, E. D. Blackham and O. N. Geertsen, The J. Acous. Soc. Ame., **37(5)**, 851-863 (1965).