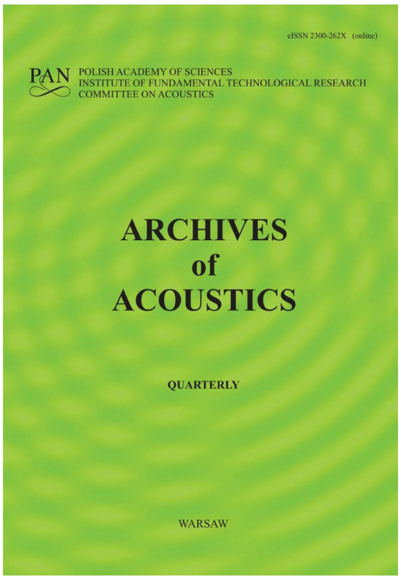
JOURNAL PRE-PROOF



Title: The Impact of Generated and Expressive Modulation of the Synthetic Instrument Sound Parameters on the Impression of Naturalness

Author(s): Marek Pluta

DOI: https://doi.org/10.24423/archacoust.2026.4327

Journal: *Archives of Acoustics*

ISSN: 0137-5075, e-ISSN: 2300-262X

Publication status: In press

Received: 2025-09-08

Revised: 2026-03-04

Accepted: 2026-03-08

Published pre-proof: 2026-03-11

Please cite this article as:

Pluta M. (2026), The Impact of Generated and Expressive Modulation of the Synthetic Instrument Sound Parameters on the Impression of Naturalness, *Archives of Acoustics*, https://doi.org/10.24423/archacoust.2026.4327

The Impact of Generated and Expressive Modulation of the Synthetic Instrument Sound Parameters on the Impression of Naturalness

Marek PLUTA[1,*]

[1] AGH University of Krakow, Department of Mechanics and Vibroacoustics, Krakow, Poland,
https://orcid.org/0000-0002-2519-8135
[*] Corresponding Author e-mail: pluta@agh.edu.pl

Abstract

Despite their different spectral structures, the sound of early instruments from the electrophone group was often considered to be deceptively similar to the sound of wind or bowed string instruments. However, the wavetable synthesizer playing a short, looped sample of a natural instrument is easily distinguishable from the actual instrument. This results from the presence of specific modulatory structures in the sound of some instruments related to expression, which can be a strong clue regarding the identification of the instrument. The control of early electrophones, such as the theremin or Martenot waves, gave the performer expressive capabilities comparable to bowed instruments. Contemporary synthesizers are returning to similar solutions. The aim of this work is to study the impact of various types of modulation on the perceived naturalness of violin sound. Modulation through an automatic low frequency oscillator is compared to expressive modulation by a human using a controller. Two advanced controllers are studied to determine whether simultaneous modulation of more than one parameter brings benefits. A set of sound samples was prepared that included violin recordings and synthesized signals, where different waveforms were combined with various modulation sources and modulated parameters. The effect was assessed by a group of expert listeners. The results indicate that expressive, multi-parameter modulation with advanced controllers brings benefits for waveforms with realistic spectra, close to that of a violin. In less realistic waveforms this kind of modulation may be perceived as less natural than a simple one, obtained through an oscillator.

Keywords: sound synthesis; signal modulation; expressive performance; expressive controller.

Acronyms

MIDI – Musical Instrument Digital Interface,

MPE – MIDI Polyphonic Expression,
LFO – low frequency oscillator,
VCO – voltage controlled oscillator,
VCA – voltage controlled amplifier,
VCF – voltage controlled filter,
f_0 – fundamental frequency,
f – frequency.

1 Introduction

Due to their versatility, sound synthesizers are often considered as a replacement for acoustic instruments. Such scenarios require a great deal of effort to reproduce the sound features of the original instrument. The bare minimum is to produce a sound that can be recognised as a chosen instrument, while the ultimate goal is to convince a listener that the sound is not synthetic, but original. A set of features (McAdams, Bruno, 2012) that need to be reproduced depends on the instrument. For some instruments, it is enough to concentrate on spectral features, while others require a proper reproduction of particular temporal structures, or combination of both (Iverson, Krumhansl, 1993).

The violin, along with the remaining members of the bowed string group, is a particularly difficult case, where synthetic counterparts are still easily distinguished from the original. The difficulty is caused by the continuous control that a performer has over multiple sound parameters. However, this specific way of controlling sound parameters is a means to produce a unique, and highly expressive performance. A key to achieve a convincing performance with synthesizers is therefore not only the ability of a synthesizer to allow control over multiple sound parameters, but also the source of a control data.

The sound of the early instruments belonging to the electrophone group was often considered to be deceptively similar to the sound of wind instruments or bowed strings, even though its spectral features differed from the original. On the other hand, the wavetable synthesizer playing a short, looped sample recorded from a natural instrument, could be easily distinguished from the actual instrument due to the absence of some modulatory structures specific for the instrument and related to expression. These structures may be a stronger clue regarding the identification of the instrument than the sound spectrum. The control of early electrophones, such as gestures used in theremin, or a moving keyboard, a ring, and a touch-sensitive lozenge in Martenot waves, gave the performer expressive capabilities comparable to bowed and wind instruments. Contemporary synthesizers are introducing control solutions based on similar assumptions, allowing to combine multiple parameters in one gesture. This gives an opportunity to study a combination of artificial sound source with natural, expressive modulation.

The synthesis of a violin sound is a known issue, and various attempts have been made to improve it. In a recent study (Liu, 2024), a violin sound has been synthesized using the wavetable synthesis method with the vibrato effect simulated using pitch modulation. The modulation rate and depth varied according to simple few-segment envelopes. The envelope parameters were set according to the performance statistics described in the literature (Schoonderwaldt, Friberg, 2001). Another study (Kim *et al.*, 2025) proposes to model the natural pitch contour using a two-stage diffusion-based synthesis framework. The first stage is responsible for the estimation of the fundamental frequency contour that controls the MIDI pitch bend. The second stage generates mel spectrogram that applies these expressive details. The attempt is similar to a prior study (Wu *et al.*, 2022) that uses Differentiable Digital Signal Processing (DDSP) to generate expressive deviations of the synthesis parameters.

Another approach to the problem, based strongly on instrument physics, has been proposed for the case of spectral synthesis techniques (Pérez Carrillo, 2009). The extensive study presents solutions to issues such as measuring violin performance using recorded sounds, as well as devices recording bow motion and force, establishing the relationship between performance data and sound timbre, and designing a generative model of timbre. The findings have been applied to design an advanced sound synthesizer. However, the applied model of vibrato is relatively simple. It uses a sine modulation of pitch, with depth and rate controlled by a fade-in and fade-out envelope with random deviations based on measurements.

The above presented studies are based either on numerical modelling of physical objects or on some form of machine learning. The comprehensive review of both approaches (Hawley *et al.*, 2020) discusses their capabilities to achieve realistic sound features. The conclusion is that in order to produce convincing sounds of instruments it is not necessary to conduct a detailed physical simulation, which requires a further validation based on measurements with live musicians and real instruments. Instead, very good results can be obtained when either a human expert or a deep neural network selects and tunes a set of salient factors based on knowledge or a very large set of audio recordings.

All of the approaches discussed above present a view of the problem focused mainly on the side of the sound source. They analyse models of instruments and refinements that can be applied to various methods of sound synthesis. A view from a different perspective, centred around human perception of sound, can provide valuable insight into the problem (Fritz *et al.*, 2025). The study compares an actual human performance using a real instrument to a bowing machine playing a real violin, and a hybrid sound synthesis that takes control data from a bowing force recorded with a human violinist. The goal of the comparison is to capture properties perceived in the actual instruments. The main part of the study discusses the methodologies based on listening tests that are applied to evaluate perceived sound qualities. The hybrid method yielded the best results, allowing one to test unlimited variants of instrument properties with the same natural excitation, and showing the best correlations with acoustic measurements. This shows that in the case of studying violin, both controlled playing conditions and natural excitation of the instrument are equally important. It can be seen that player interaction can influence the perception of produced sound even more than the properties of the instrument.

Current research trends appear to focus on one particular use case, where a synthesizer serves as a means to automatically produce an artificial recording of an instrument on the basis of the musical score alone. Such scenario requires an implementation of a simulated expression. There is, however, another use case, where a synthesizer is used in a live performance, using some form of controller, such as a keyboard. Here, all the expression can, and should, come from a performer. There are, however, two limitations: synthesis parameters available to be assigned with the expression data, and limitations of the controller. A simple keyboard does not allow one to control the parameters in a continuous manner. An addition of aftertouch capability, that is, a continuous sensitivity to a force applied to the already depressed key solves the problem. However, if more than one parameter needs to be controlled, which is the case of a violin, the aftertouch is insufficient. Thus, more advanced controllers are required.

Various experiments have been carried out with designs adapted for specific instruments, such as accordion (Gurevich, von Muehlen, 2001) or even for human voice (Donati, Chousidis, 2022). Some attempts have been made with game controllers adapted for controlling sound synthesis parameters, but finally expressive extension (The MIDI Manufacturers Association, 2018) for the original MIDI specification (The MIDI Manufacturers Association, 1996) allowed one to design general purpose multiple degrees of freedom controllers for synthesizers (Robertson, 2011). These controllers attempt to give the musician a possibility to freely manipulate multiple parameters, possibly with a single gesture. Typical solutions include touch-sensitive keys that react not only to a change in pressure, but also to a change of finger placement on the key surface, which allows a single finger to control three parameters. Other solutions include keys with additional rotational axes.

When combined with a sufficiently complex synthesizer, advanced, multiple degree of freedom controllers have a potential to significantly improve features of a synthesized sound by adding a natural expressive performance. So far, this possibility has not undergone a thorough study, and remains open.

The aim of this work is to study the impact of various types of modulation on the perceived naturalness of violin sound, and to compare modulation through an automatic low-frequency oscillator to expressive modulation by a human using a controller. Two advanced controllers are studied to determine whether simultaneous modulation of more than one parameter brings benefits. A synthesizer with architecture open for controller-related modifications has been designed as a base for the current study and its future continuation. Using the synthesizer, a set of sound samples was prepared that included violin recordings and synthesized signals, where different waveforms were combined with various modulation sources and modulated parameters. The effect was assessed by a group of expert listeners. The results indicate that expressive, multi-parameter modulation with advanced controllers brings benefits for waveforms with realistic spectra, close to that of a violin. However, in less realistic waveforms, this kind of modulation may be perceived as less natural than a simple one, such as that obtained through an oscillator.

2 The experiment

The aim of the experiment was to present a group of expert listeners with a set of sound samples. The samples contained a middle section of a single violin note, a^1 (f_0 = 440 Hz). They represented combinations of various synthesized sounds with several variants of modulation. Some modulations were generated using LFO, some were recorded by a violinist using an expressive controller. The listeners were asked to assess how natural each sample was. The analysis of listeners' assessments should give information regarding the impact of particular sound features on the impression of naturalness of a violin sound.

Fig. 1 presents the procedure and elements of the experiment. The main element is a program that serves three purposes. It is a sound synthesizer, and plays a sound in response to data from the controller. The controller, operated by a violinist, produces a stream of modulation data, which is recorded by the program. With data from the controller recorded, the program can use it to modulate selected synthesis parameters and record generated audio signal. The supervisor chooses a desired base signal and adds LFO or recorded modulation to the selected parameters. The resulting audio signals are recorded to be used in the listening test.

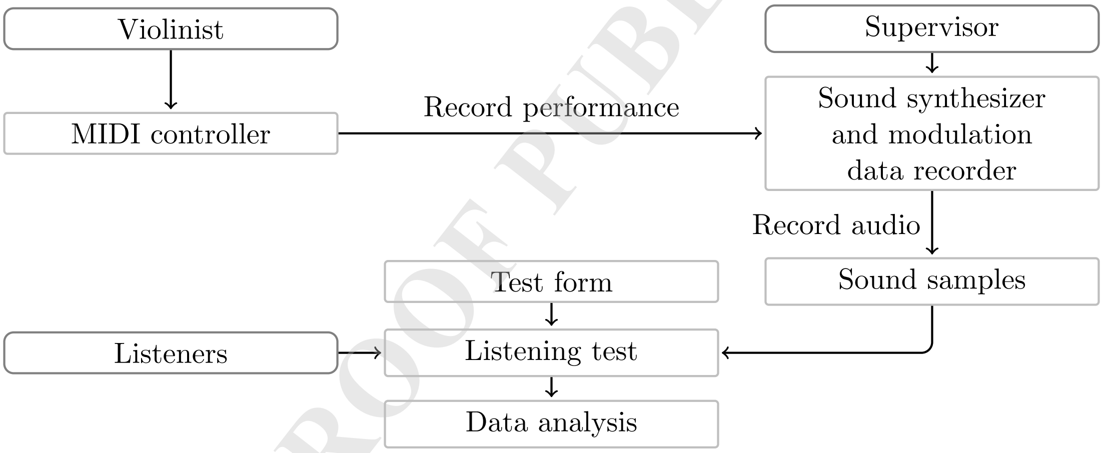


Fig. 1. The overall diagram of the experiment.

2.1 Synthesis and modulation

The program is implemented in the Max/MSP visual programming language, which is capable of handling data from MPE-compatible controllers. It consists of three modules. The first is responsible for real-time performance and reacts to data from the controller by synthesizing a sound and adjusting its parameters accordingly, providing a performer with auditory feedback.

The second module catches and stores a stream of MPE data sent by the controller. This data is the recording of expressive performance of the violinist. Three separate streams can be recorded simultaneously: pitch bend data, amplitude data, and aftertouch data. The pitch bend data controls the instantaneous deviation of the fundamental frequency, which is a main component of the violin vibrato. The amplitude stream is independent from a single-number

MIDI velocity parameter, sent when a key is pressed. It can control the amplitude envelope or add amplitude modulation which, to some extent, usually occurs in parallel to frequency modulation. The aftertouch data are interpreted as a timbre parameter, which controls the cut-off frequency of the low-pass filter applied to the signal. Again, it may be used to add an envelope or oscillatory modulation.

The last program module uses the same synthesizer as the first real-time module to reproduce a sound with selected modulation. The source of modulation can be an internal LFO, or a stored performance stream. The selected signal is combined with the selected modulation source and applied to selected parameters: pitch, amplitude, or cut-off frequency. The synthesized sound is played and stored in an audio file.

The synthesizer can produce three types of constant signal: sawtooth, filtered sawtooth, and a single looped period of a real violin sound, recorded in an anechoic chamber using a close microphone (Fig. 2). The first two may be considered a case of subtractive synthesis. The last one is based on the wavetable principle and reproduces real violin spectrum but is devoid of any internal evolution. Filtered sawtooth uses three resonant low-pass filters to roughly shape lower regions of the spectrum, similarly to the violin spectrum. Thus, the three signal types are graded from purely synthetic (sawtooth), to a violin-like (wavetable). The selected waveform passes through a controllable low-pass resonant filter and a controllable amplifier. The frequency deviation of the waveform, the filter cut-off frequency, and the signal amplitude can be modulated either by a common internal LFO, or separate data streams recorded from the controller and implemented as envelope generators, as shown in Fig. 3.

Violin recordings, carried out in an anechoic chamber, were analysed and used as a reference for setting ranges for modulation parameters. They were compared and verified against values found in the literature (Schoonderwaldt, Friberg, 2001). The modulation parameters are presented in Table 1.

Table 1: Ranges of modulation parameters.

Modulation source	Modulation rate	Pitch-bend	Amplitude	Filter cut-off
LFO	6 Hz	± 25 cent	± 1 dB	3000–6000 Hz
Controller	Variable	± 25 cent	100%	3000–6000 Hz

2.2 Controllers

Performance data streams were recorded using two controllers: Expressive E Osmose, and Joue Play. The former is a novel keyboard controller. The latter is an universal controller with interchangeable templates. Both allow controlling three parameters with a single finger, although each has its own limitations.

Expressive E Osmose (Fig. 4) is a keyboard with an additional axis: the keys can be pressed down and moved sideways. Pressing is divided into two regions: in the first region, the key moves lightly and in the second region there is a perceptible key resistance. The boundary

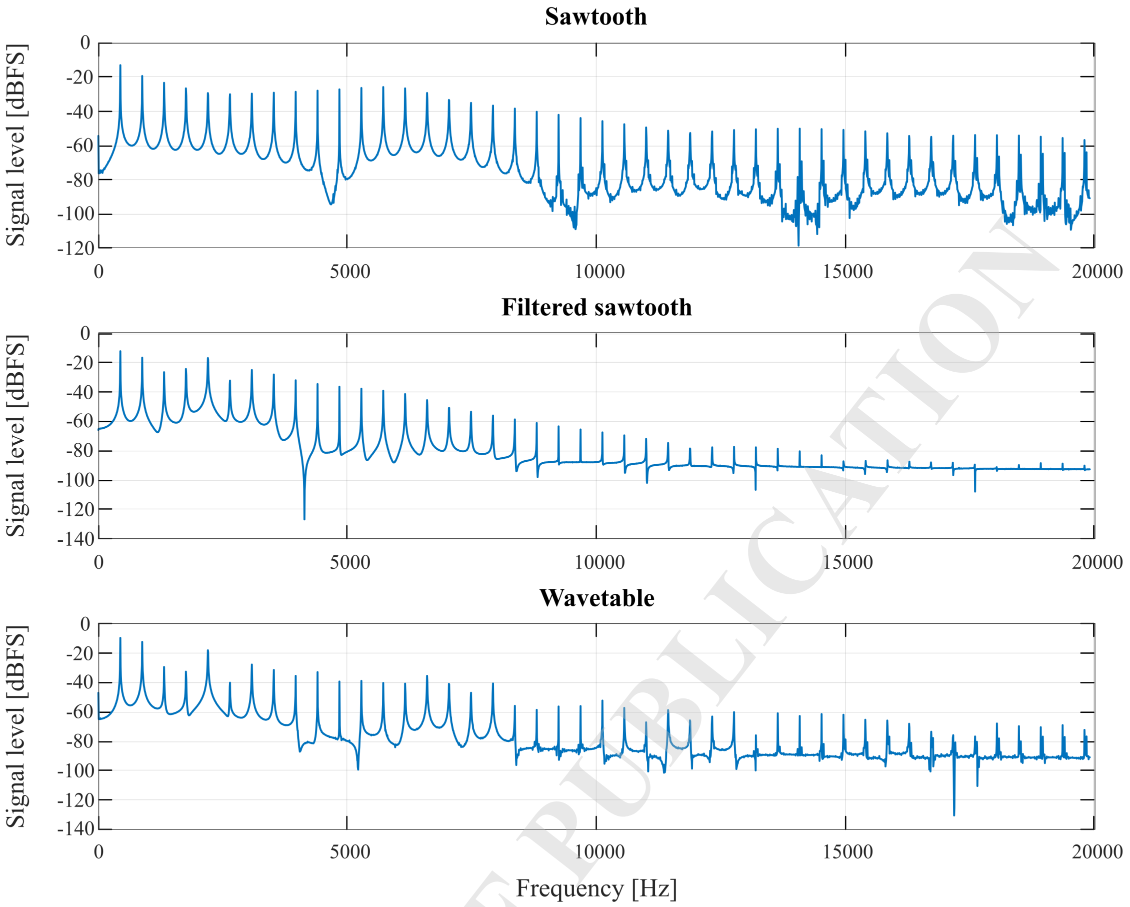


Fig. 2. Spectra of the synthesized signals. Base waveforms are filtered with a default setting of aftertouch-controlled low-pass resonant filter.

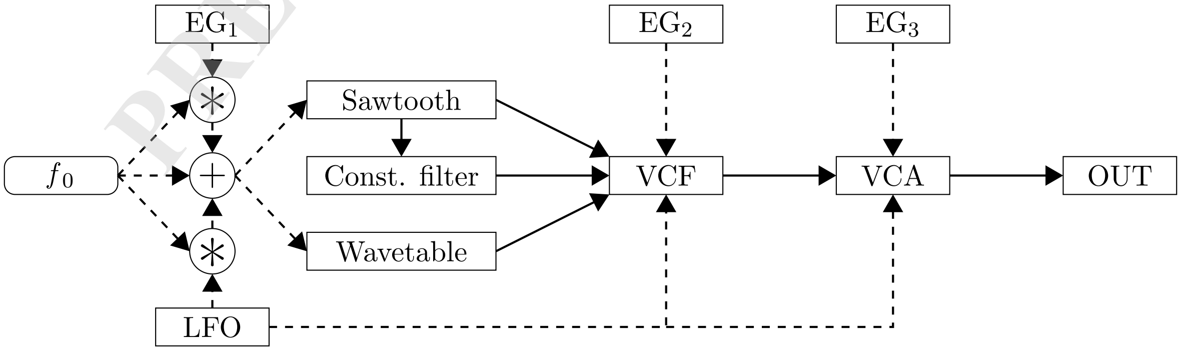


Fig. 3. The diagram of the synthesizer (VCF – adjustable low-pass filter, VCA – adjustable amplifier). Pitch, filter cut-off and amplitude are modulated either by internal LFO, or separate envelope generators (EG$_1$–EG$_3$). One of three available waveforms is used at a time.

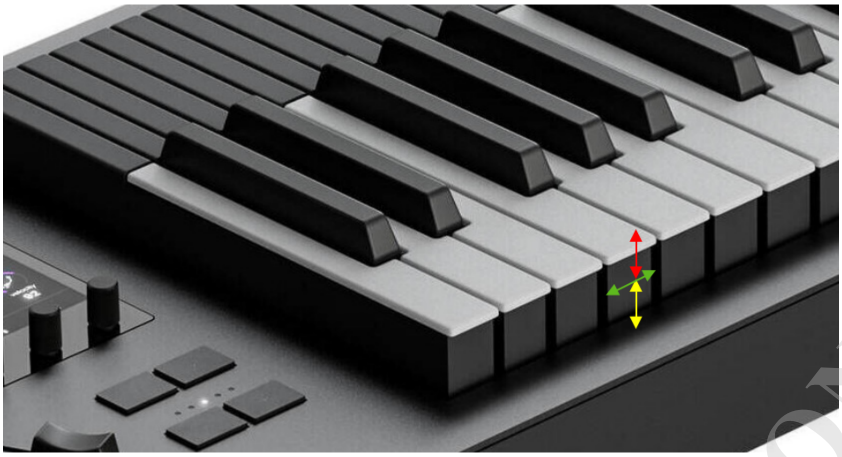


Fig. 4. Controller Expressive E Osmose (green arrow shows pitch bend control, red – amplitude control, yellow – aftertouch).

between both regions is easily perceptible and each region controls its own parameter. In a default mode, which has been used in research, the light region controls the amplitude, the resistant region controls the aftertouch, and the sideways movement controls the pitch bend. A clear limitation is the inability to simultaneously change the amplitude and aftertouch, which belong to different regions of the same key travel.

Joue Play (Fig. 5) was used with a template consisting of 17 identical key-straps. Each strap works as an XY touch-sensitive pad. Moving a finger sideways controls the pitch bend, moving it upwards or downwards controls the aftertouch, and the pressure of the finger controls the amplitude. All three parameters can be controlled simultaneously. However, the actual resolution and sensitivity varies on the surface, which limits the combinations of parameter values achievable.

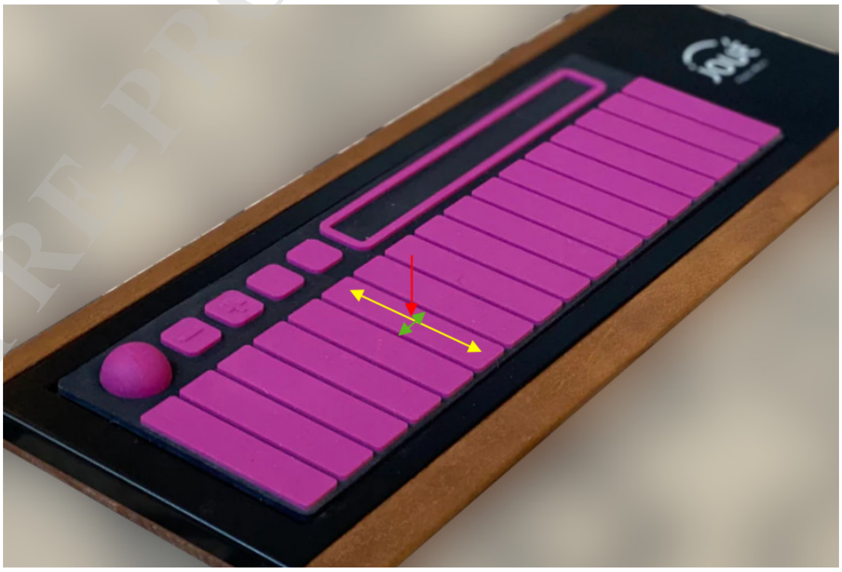


Fig. 5. Controller Joue Play (green arrow shows pitch bend control, red – amplitude control, yellow – aftertouch).

2.3 Recorded performance

The person using both controllers was a violinist experienced in bowed strings and electronic keyboard instruments. The synthesizer was set to reproduce the violin wavetable. The task of the musician was to perform a single, long note that would be as close as possible to the sound of the violin, with regard to expression. The violinist used an auditory feedback to refine the sound, until a satisfactory performance had been recorded. The recorded data streams are shown in Fig. 6. Amplitude and aftertouch are represented by unsigned 7-bit integer values. Pitch bend has a theoretical 14-bit resolution and is represented with floating point values.

The limitations of both controllers are clearly visible. Osmose is used in the aftertouch range; therefore, simultaneous amplitude control is impossible, and its value stays constant, apart from the moments of pressing and releasing a key. Joue seriously limits the practical aftertouch range due to large finger movement (along the entire length of a strap), and its amplitude stream shows some discontinuities.

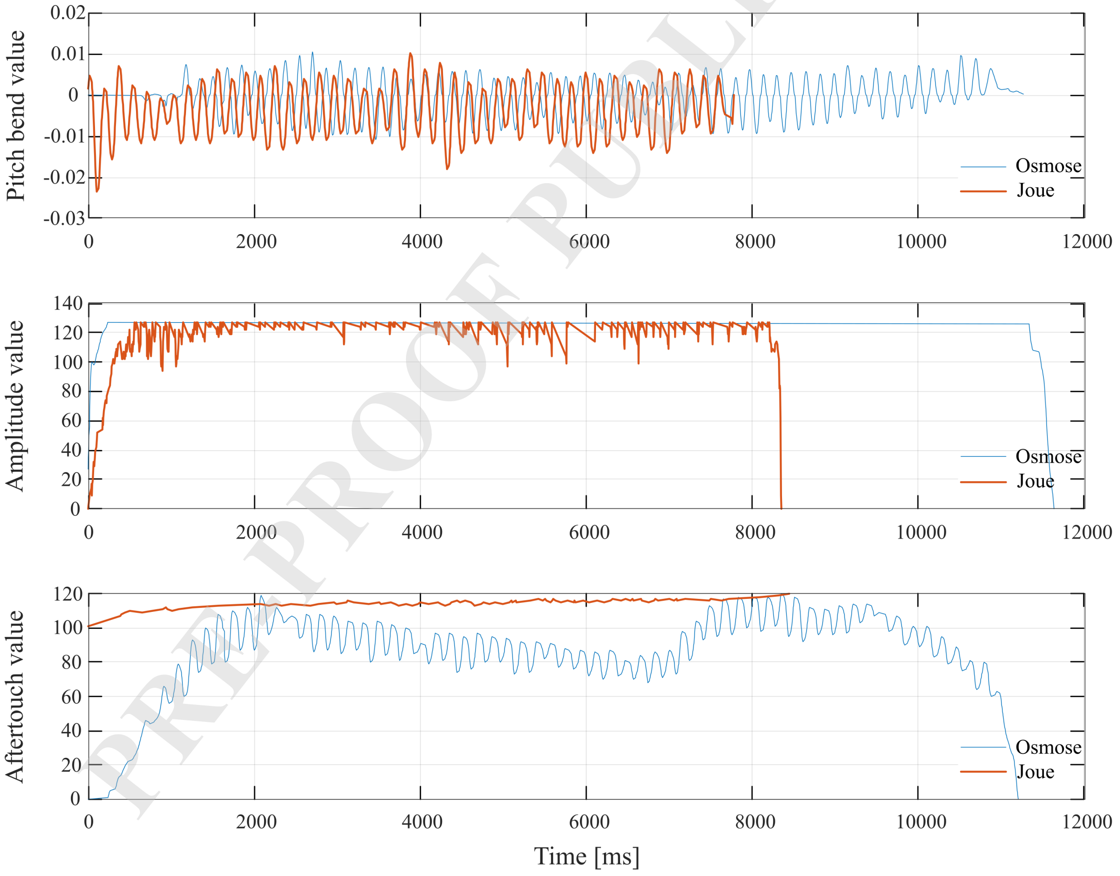


Fig. 6. Raw data recorded from controllers. Pitch bend controls signal fundamental frequency, amplitude controls signal level, and aftertouch controls filter cut-off frequency.

2.4 Sound samples

Using both internal LFO and recorded performance data for modulation of selected parameters, a total of 67 different sound samples has been created. The samples differed in the choice of the base waveform, the modulation source, and the set of modulated parameters. Selection included:

1) recording of the original instrument from an anechoic chamber;

2) 3 synthesized samples (different base waveforms) without modulation;

3) 63 synthesized samples with modulation – all combinations of the following variants:
   - 3 waveforms: violin wavetable (Vn), filtered sawtooth (Sf), and non-filtered sawtooth (Sn),
   - 3 modulation sources: Joue (J), Osmose (O), internal LFO (L),
   - 7 combinations of modulated parameters: amplitude (A), filter cut-off frequency (T – timbre), fundamental frequency (P – pitch).

Despite the fact that modulation of the amplitude and filter cut-off frequency affected the signal level, sound samples were kept at their recorded levels and were not level-matched in loudness to reflect the original impact of expression on the signal. Excluding the original violin recording, the maximum peak level difference among sound samples reached 1.24 dB and the maximum RMS level difference reached 4.54 dB. Including the violin recording, the maximum peak level difference reached 6 dB and the maximum RMS level difference reached 10.74 dB.

Only 3 seconds fragments from the middle section of the recordings were selected for the test to exclude the influence of the attack phase, which was not relevant to the study. 50 ms fade in and fade out have been applied to all samples. Examples of three sound samples are presented in Figs. 7, 8, and 9. All sound samples are available for download from http://home.agh.edu.pl/~pluta/varia/AudioExamples260303.zip.

2.5 Listening test

The effect of modulation was evaluated by a group of expert listeners. They rated the naturalness of each sound sample using a 5 grade scale: excellent (5), good (4), fair (3), poor (2) and bad (1). The listeners were not informed about the details of sound processing. They were only asked to evaluate the samples as the sustain phase of a violin sound. It was possible to listen to the samples multiple times.

The responses were collected using an online form. The form included four questions regarding the nature of the listeners' experience: 1) experience playing the violin, 2) experience playing other instruments or singing, 3) experience in audio signal processing, and 4) experience in sound recording or music production. Additionally, five sound samples occurred twice in test in order to evaluate coherence of the listener's responses. All samples were arranged in random order.

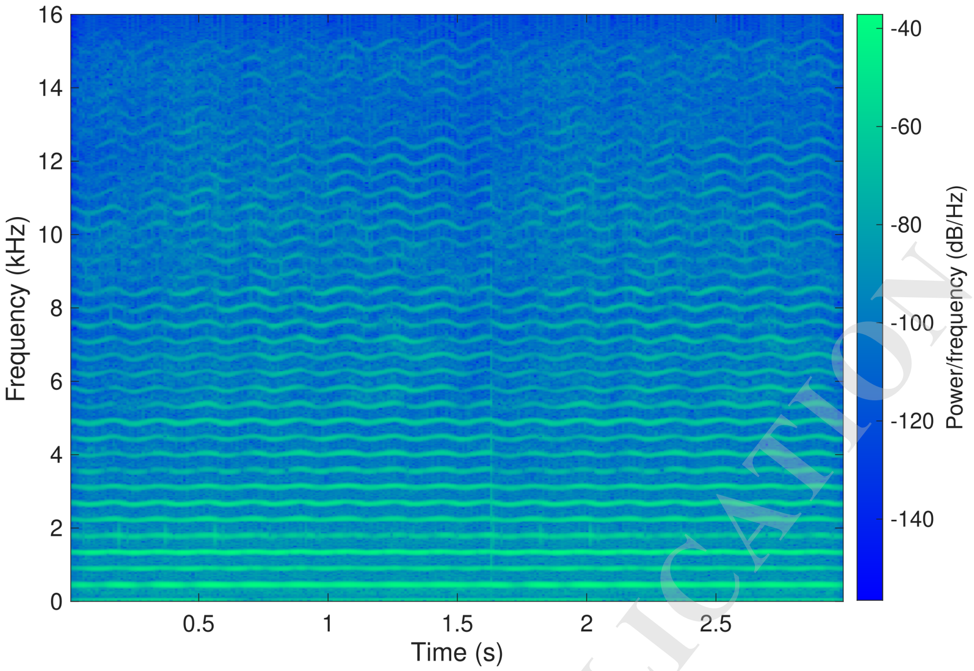


Fig. 7. Spectrogram of the violin recording.

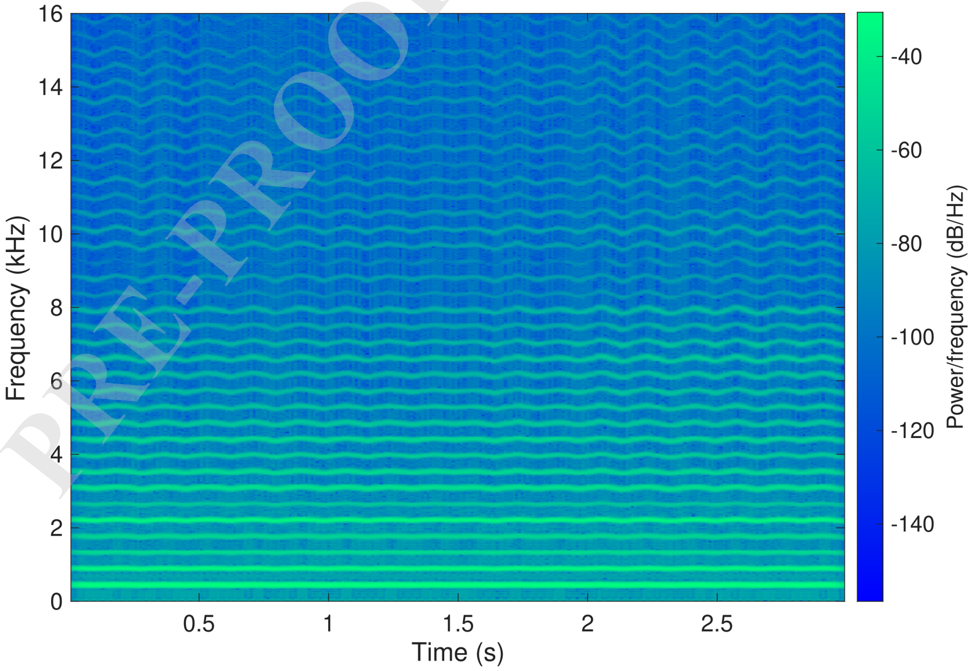


Fig. 8. Spectrogram of the violin wavetable; amplitude, cut-off and pitch modulated by Osmose.

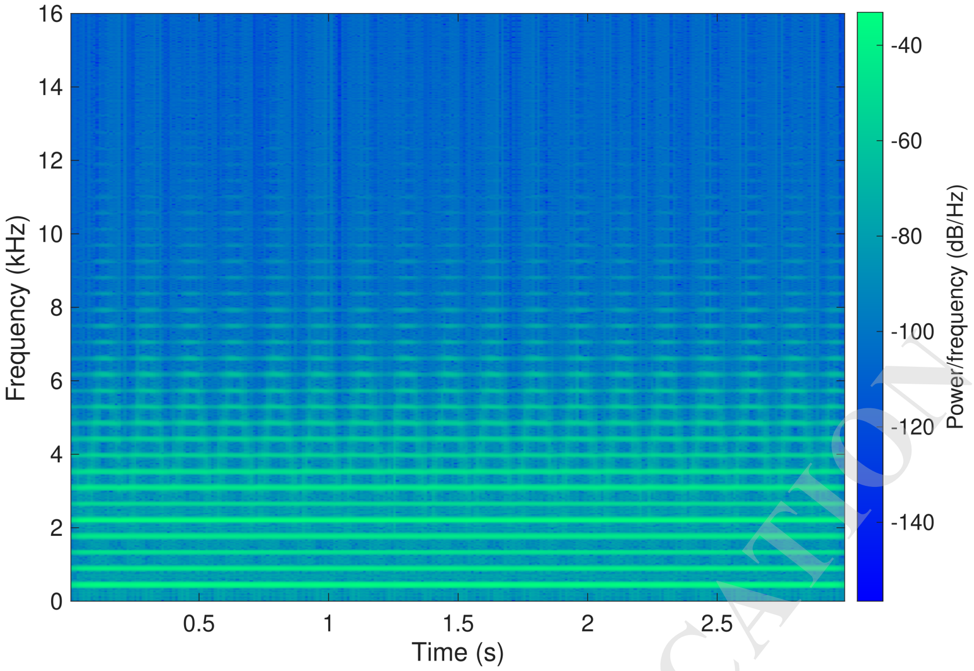


Fig. 9. Spectrogram of the filtered sawtooth; amplitude and cut-off modulated by LFO.

A total of 16 listeners participated in the test. Every listener had at least some experience (up to 4 years) in at least 2 categories or high experience in at least one category. 13 listeners had high experience in at least 2 categories, and 6 listeners had high experience in 3 categories. This translated well into coherence of responses. If a response to a repeated question differed by not more than 1, such a response was considered coherent. All listeners were coherent in at least 4 of 5 repeated questions, and 14 were coherent in all repeated questions. Therefore, all responses in the test were considered valid.

3 The results

All results have been presented in a form of histograms of the responses of the listeners. Histograms are divided into cases that represent combinations of various waveforms, modulation sources, and modulated parameters. Symbols used in charts are explained in Section 2.4. All the scores presented in histograms are summarised in tables as means and medians.

For the baseline, the results for samples without any modulation for three types of waveform are presented in Fig. 10 and in Table 2. As expected, more resemblance to the spectrum of the violin yields better results, although even the results for the violin wavetable are not better than fair.

The second baseline is the violin recording, which in Fig. 11 and in Table 3 is compared to the most complete modulation of all three parameters (ATP), with cases representing three different modulation sources: two controllers and LFO. For the violin, half of the responses give the maximal score (5), which cannot be matched by any synthetic source. At the same time, it can be seen that with full modulation controllers yield better results than LFO, which is the

Table 2: A summary of ratings for all types of waveform without modulation.

Sample	Mean score	Median
Vn	2.00 ± 0.79	2.00
Sf	1.56 ± 0.70	1.00
Sn	1.00 ± 0.00	1.00

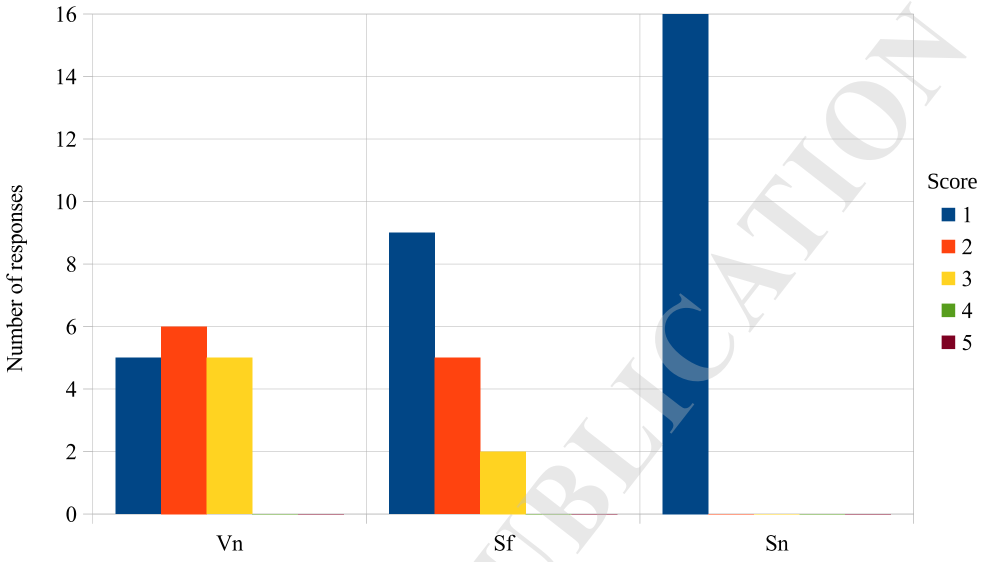


Fig. 10. Responses for all types of waveform without modulation.

least natural. The results of both controllers are similar.

Table 3: A summary of ratings for the violin recording and for cases of complete, three parameter modulation (ATP).

Sample	Mean score	Median
Violin	4.00 ± 1.17	4.50
J Vn ATP	2.56 ± 1.00	2.00
O Vn ATP	2.50 ± 1.06	2.00
L Vn ATP	1.75 ± 0.75	2.00

Figs. 12–14 and Table 4 present a complete set of the results obtained. Unsurprisingly, almost all responses for non-filtered sawtooth are bad. This kind of waveform cannot convince listeners even with all three parameters (ATP) modulated with expressive controllers. Both controllers with modulation of three parameters are the best cases here, as well as Osmose with pitch vibrato supplemented with modulation of the second parameter. However, most of the responses are bad. When comparing filtered sawtooth (Sf) to the wavetable (Vn), a larger spread of results is seen in case of the wavetable.

In Figs. 15–16 and in Table 5 a selection of cases is presented to illustrate the impact of multi-parameter modulation, separately for the filtered sawtooth and for the wavetable. For

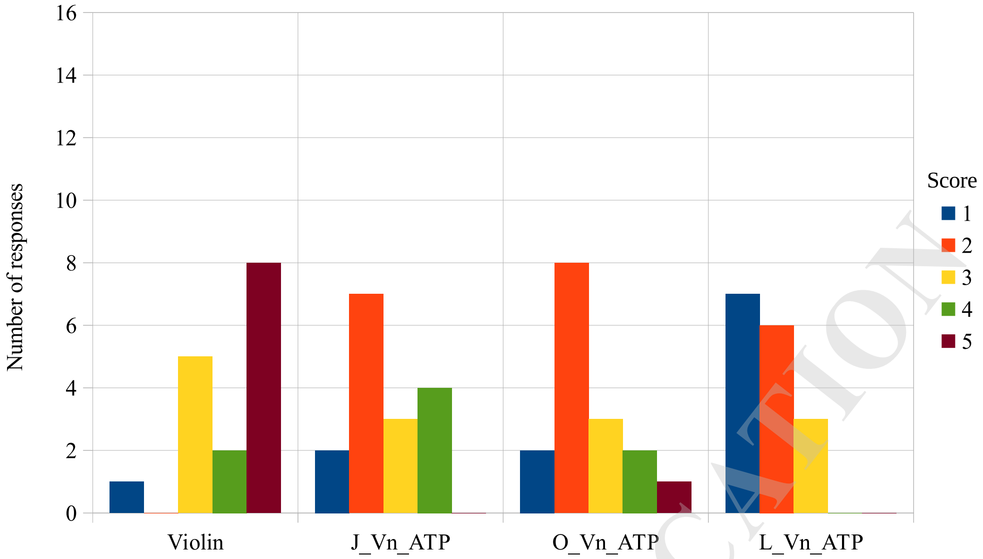


Fig. 11. Responses for the violin recording and for cases of complete, three parameter modulation (ATP).

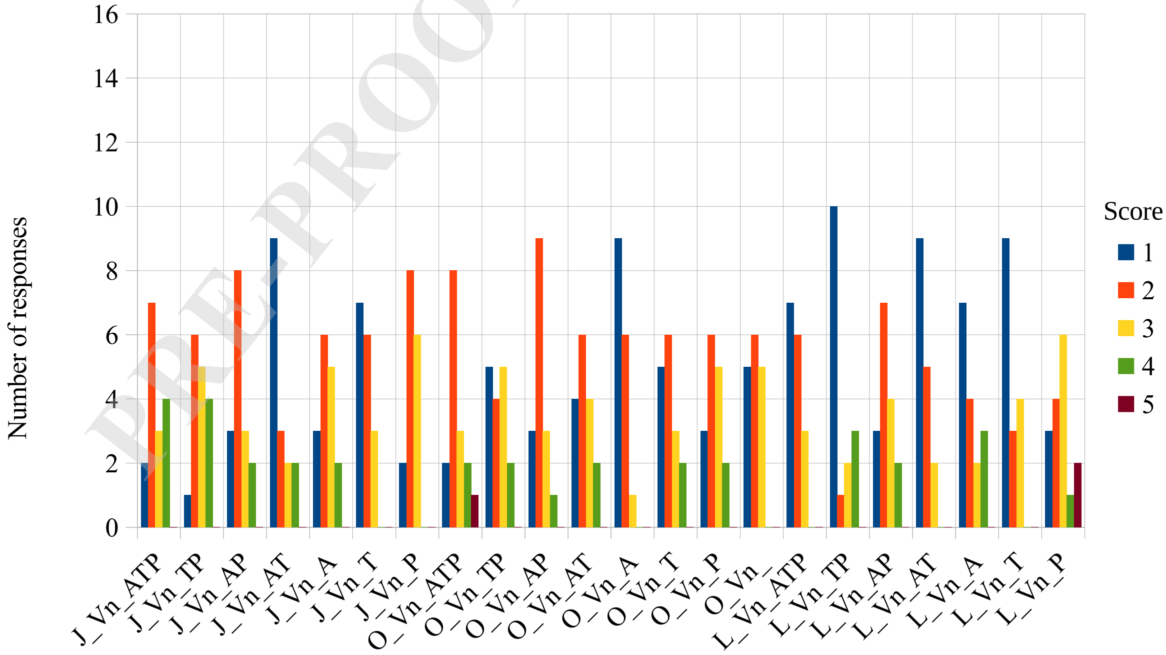


Fig. 12. Responses for all cases of wavetable waveform (Vn).

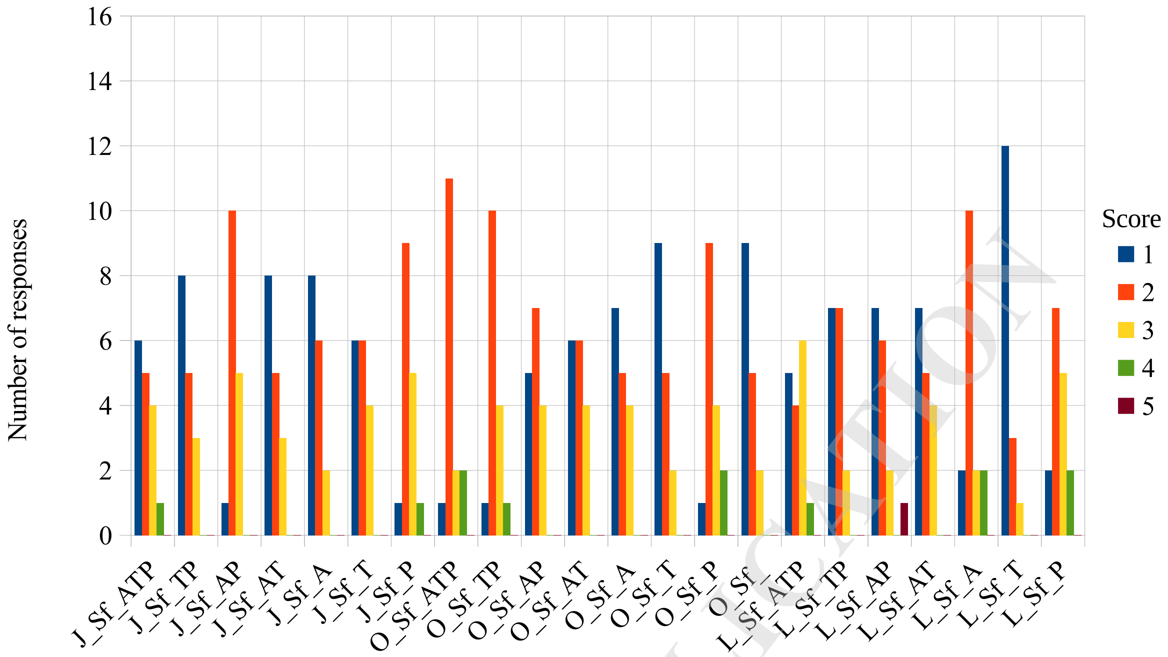


Fig. 13. Responses for all cases of filtered sawtooth waveform (Sf).

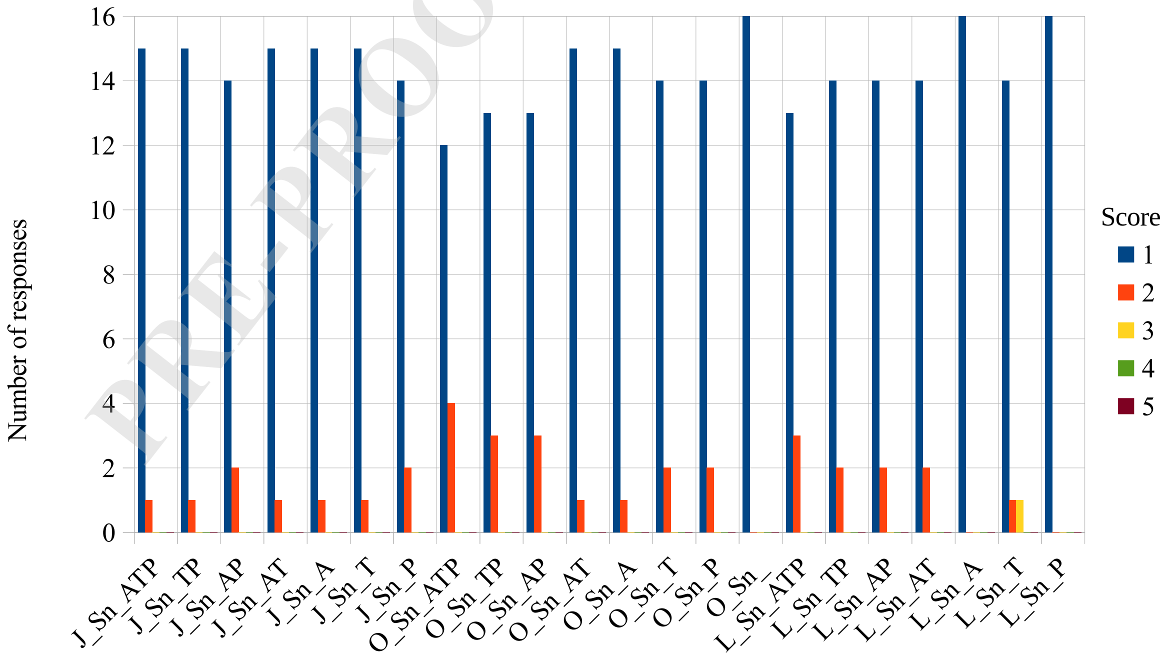


Fig. 14. Responses for all cases of non-filtered sawtooth waveform (Sn).

Table 4: A summary of ratings for all samples.

Sample	Mean score	Med.	Sample	Mean score	Med.	Sample	Mean score	Med.
J Vn ATP	2.56 ± 1.00	2.00	J Sf ATP	2.00 ± 0.94	2.00	J Sn ATP	1.06 ± 0.24	1.00
J Vn TP	2.75 ± 0.90	3.00	J Sf TP	1.69 ± 0.77	1.50	J Sn TP	1.06 ± 0.24	1.00
J Vn AP	2.25 ± 0.90	2.00	J Sf AP	2.25 ± 0.56	2.00	J Sn AP	1.13 ± 0.33	1.00
J Vn AT	1.81 ± 1.07	1.00	J Sf AT	1.69 ± 0.77	1.50	J Sn AT	1.06 ± 0.24	1.00
J Vn A	2.38 ± 0.93	2.00	J Sf A	1.63 ± 0.70	1.50	J Sn A	1.06 ± 0.24	1.00
J Vn T	1.75 ± 0.75	2.00	J Sf T	1.88 ± 0.78	2.00	J Sn T	1.06 ± 0.24	1.00
J Vn P	2.25 ± 0.66	2.00	J Sf P	2.38 ± 0.70	2.00	J Sn P	1.13 ± 0.33	1.00
O Vn ATP	2.50 ± 1.06	2.00	O Sf ATP	2.31 ± 0.77	2.00	O Sn ATP	1.25 ± 0.43	1.00
O Vn TP	2.25 ± 1.03	2.00	O Sf TP	2.31 ± 0.68	2.00	O Sn TP	1.19 ± 0.39	1.00
O Vn AP	2.13 ± 0.78	2.00	O Sf AP	1.94 ± 0.75	2.00	O Sn AP	1.19 ± 0.39	1.00
O Vn AT	2.25 ± 0.97	2.00	O Sf AT	1.88 ± 0.78	2.00	O Sn AT	1.06 ± 0.24	1.00
O Vn A	1.50 ± 0.61	1.00	O Sf A	1.81 ± 0.81	2.00	O Sn A	1.06 ± 0.24	1.00
O Vn T	2.13 ± 0.99	2.00	O Sf T	1.56 ± 0.70	1.00	O Sn T	1.13 ± 0.33	1.00
O Vn P	2.38 ± 0.93	2.00	O Sf P	2.44 ± 0.79	2.00	O Sn P	1.13 ± 0.33	1.00
O Vn	2.00 ± 0.79	2.00	O Sf	1.56 ± 0.70	1.00	O Sn	1.00 ± 0.00	1.00
L Vn ATP	1.75 ± 0.75	2.00	L Sf ATP	2.19 ± 0.95	2.00	L Sn ATP	1.19 ± 0.39	1.00
L Vn TP	1.88 ± 1.22	1.00	L Sf TP	1.69 ± 0.68	2.00	L Sn TP	1.13 ± 0.33	1.00
L Vn AP	2.31 ± 0.92	2.00	L Sf AP	1.88 ± 1.05	2.00	L Sn AP	1.13 ± 0.33	1.00
L Vn AT	1.56 ± 0.70	1.00	L Sf AT	1.81 ± 0.81	2.00	L Sn AT	1.13 ± 0.33	1.00
L Vn A	2.06 ± 1.14	2.00	L Sf A	2.25 ± 0.83	2.00	L Sn A	1.00 ± 0.00	1.00
L Vn T	1.69 ± 0.85	1.00	L Sf T	1.31 ± 0.58	1.00	L Sn T	1.19 ± 0.53	1.00
L Vn P	2.69 ± 1.21	3.00	L Sf P	2.44 ± 0.86	2.00	L Sn P	1.00 ± 0.00	1.00

both waveforms, the case of no modulation is the least natural. When modulation in the filtered sawtooth is applied only to pitch, which is the simplest form of vibrato, LFO and both controllers score very similar results. Strangely, for three-parameter modulation results for LFO and one controller (Joue) turn worse, mostly due to increased number of bad (1) scores. The results of the second controller (Osmose) with modulated three parameters do not change much in comparison to the modulation of one parameter. Clearly, multi-parameter modulation does not help with a waveform that is not very similar to the real instrument. The situation changes with the wavetable waveform. Here, transition from one parameter to three parameter modulation improves the results for both controllers, and worsens for the LFO. However, LFO yielded better results than controllers for single-parameter (pitch) modulation. With three parameters modulated Osmose is considered the most natural. An obvious conclusion is that any modulation makes the violin sound more natural. However, in order for expressive controllers to show their benefits, the spectrum of the signal needs to be similar to that of the original instrument. Only such cases are able to convince some respondents to reward the synthetic signal with the best score (5).

Fig. 17 shows how the improvement of the waveform impacts the score when only the pitch is modulated. Here, gains for the case of LFO modulation are the largest. One of the

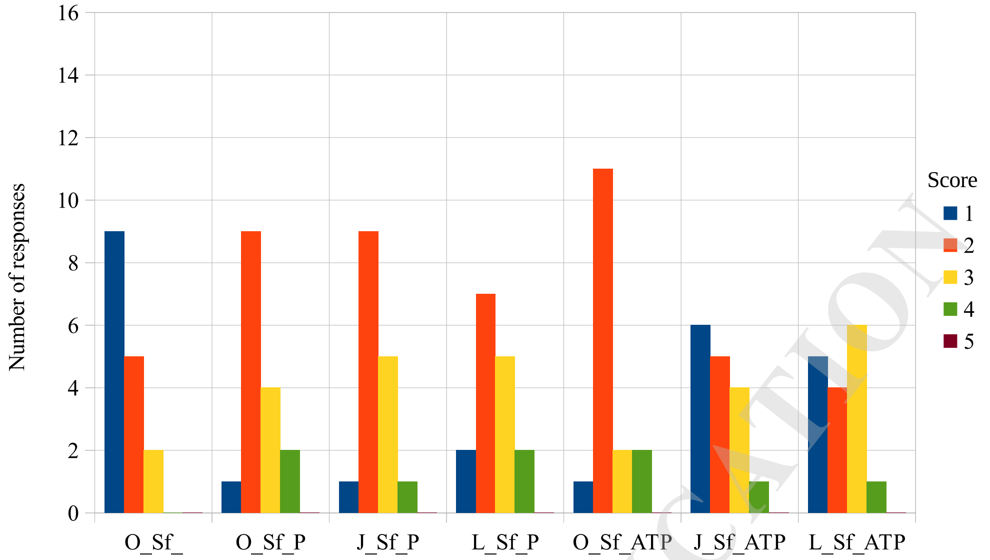


Fig. 15. Responses for cases of zero, one and three parameters modulated in the filtered sawtooth waveform.

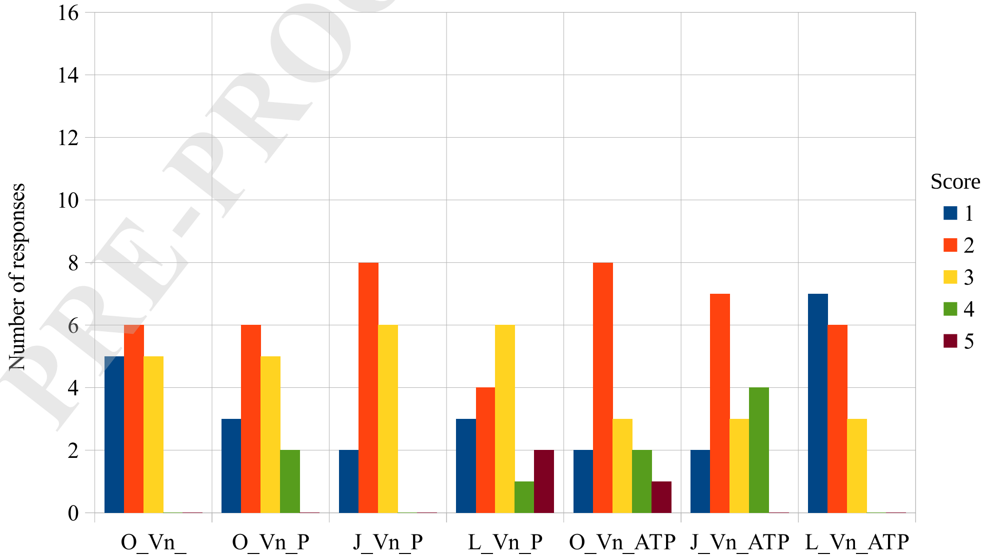


Fig. 16. Responses for cases of zero, one and three parameters modulated in the wavetable waveform.

Table 5: A summary of ratings for cases of zero, one and three parameters modulated in the filtered sawtooth and in the wavetable waveform.

Sample	Mean score	Median	Sample	Mean score	Median
O Sf	1.56 ± 0.70	1.00	O Vn	2.00 ± 0.79	2.00
O Sf P	2.44 ± 0.79	2.00	O Vn P	2.38 ± 0.93	2.00
J Sf P	2.38 ± 0.70	2.00	J Vn P	2.25 ± 0.66	2.00
L Sf P	2.44 ± 0.86	2.00	L Vn P	2.69 ± 1.21	3.00
O Sf ATP	2.31 ± 0.77	2.00	O Vn ATP	2.50 ± 1.06	2.00
J Sf ATP	2.00 ± 0.94	2.00	J Vn ATP	2.56 ± 1.00	2.00
L Sf ATP	2.19 ± 0.95	2.00	L Vn ATP	1.75 ± 0.75	2.00

controllers (Joue) sees a decrease in the score when the filtered sawtooth is changed to wavetable. Everything changes when three parameters are modulated (Fig. 18 and Table 6). In this case, the controllers produce the best scores for the wavetable, but the LFO score decreases.

Table 6: A summary of ratings for various waveforms with pitch modulation only and with full (three-parameter) modulation.

Sample	Mean score	Median	Sample	Mean score	Median
O Sn P	1.13 ± 0.33	1.00	O Sn ATP	1.25 ± 0.43	1.00
O Sf P	2.44 ± 0.79	2.00	O Sf ATP	2.31 ± 0.77	2.00
O Vn P	2.38 ± 0.93	2.00	O Vn ATP	2.50 ± 1.06	2.00
J Sn P	1.13 ± 0.33	1.00	J Sn ATP	1.06 ± 0.24	1.00
J Sf P	2.38 ± 0.70	2.00	J Sf ATP	2.00 ± 0.94	2.00
J Vn P	2.25 ± 0.66	2.00	J Vn ATP	2.56 ± 1.00	2.00
L Sn P	1.00 ± 0.00	1.00	L Sn ATP	1.19 ± 0.39	1.00
L Sf P	2.44 ± 0.86	2.00	L Sf ATP	2.19 ± 0.95	2.00
L Vn P	2.69 ± 1.21	3.00	L Vn ATP	1.75 ± 0.75	2.00

The fact that sound samples were not level-matched in loudness might have affected the results, with some samples more pronounced than others. However, post-recording level-matching might have affected the results as well, because the level differences were caused by various combinations of temporal and spectral signal modifications, which are perceived differently. A more in-depth study is needed to properly address this issue.

4 Conclusions

A common way to improve the sound produced by a synthesizer leads through enhancing or refining its synthesis algorithm. However, current synthesis algorithms are products of a long way of improvements, and further refinements are difficult. This paper proposed a different approach that can be used in real-time performances. A study was carried out to determine the impact of various types of modulation on the perceived naturalness of the violin sound. Modula-

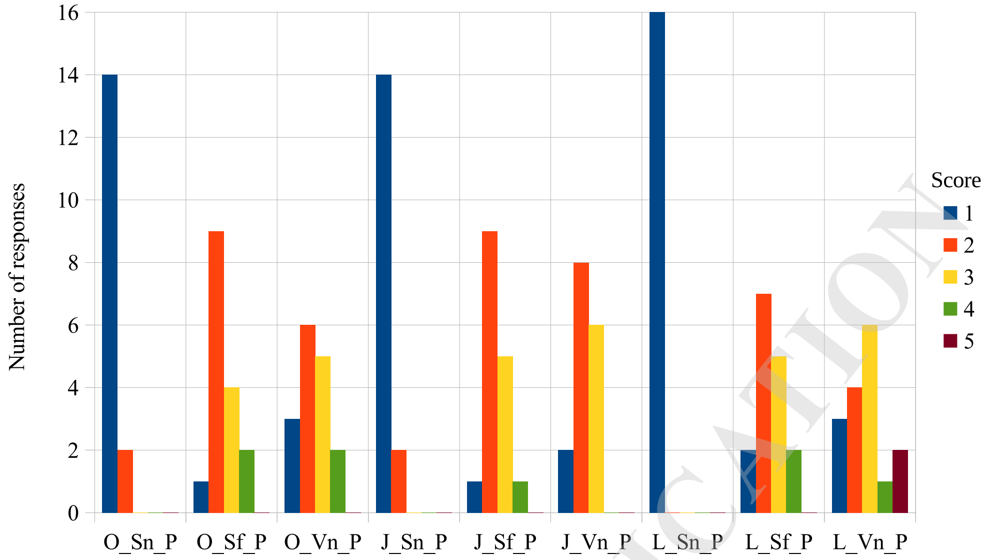


Fig. 17. Responses for various waveforms with pitch modulation only.

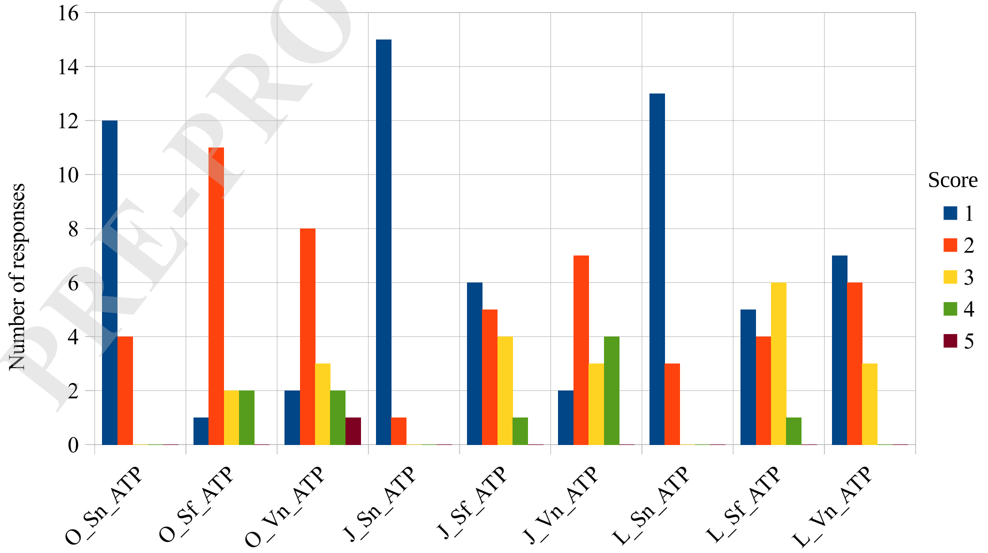


Fig. 18. Responses for various waveforms with full (three-parameter) modulation.

tion through an automatic low-frequency oscillator was compared with expressive modulation caused by a human using a controller. Two advanced controllers were studied to determine whether simultaneous modulation of more than one parameter can cause a synthetic sound to be perceived as more natural. This has a potential to improve the sound quality of existing synthesizers by changing or refining their controllers.

A synthesizer with an architecture open for controller-related modifications has been designed and used to prepare a set of sound samples, where different waveforms were combined with various modulation sources and modulated parameters. The set was supplemented with real violin recordings. The effect was assessed by a group of expert listeners. The results show a relatively complex situation that requires further study. It has been observed that expressive multi-parameter modulation with advanced controllers brings benefits for waveforms with realistic spectra, close to that of a violin. However, in less realistic waveforms, this kind of modulation may be perceived as less natural than a simple one, such as that obtained through an oscillator.

The conclusion from the study is that unlike the case of simple MIDI controllers that can be successfully used with their default settings, advanced controllers bring both improvements and problems that require solutions. Firstly, one finger cannot effectively control three parameters independently. Both studied controllers handled well control of two parameters, but the third could not have been precisely adjusted, or its range was severely limited. It is partially a design limitation, but further study with other than default mappings between parameters and controller degrees of freedom may bring a solution. Secondly, an expression must be coherent with a signal. If the signal features differ too much from the original instrument, the overly expressive control characteristic of this instrument is perceived as unnatural. Finally, continuous multi-parameter control closes the gap between real instruments, such as the violin, and synthesizers. This implies that performance with such synthesizers will be much more demanding than with simple MIDI devices. They would require a much more thorough learning and training process on the side of the performer. However, as the results show, the perception of expressive sounds clearly improves the synthesis effect if the process is implemented properly, which gives an impulse for future studies.

Fundings

This research was funded by the Department of Mechanics and Vibroacoustics of AGH University of Krakow, Poland, grant number 16.16.130.942.

Conflict of interest

The author declares that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Authors' contribution

Marek Pluta conceptualized the study, wrote the original draft, wrote computer programs used within the study, prepared sound samples, performed data interpretation and analysis, and reviewed the final manuscript.

References

[1] Donati E., Chousidis C. (2022), Electroglottography based real-time voice-to-MIDI controller, *Neuroscience Informatics*, **2**(2), https://doi.org/10.1016/j.neuri.2022.100041.

[2] Fritz C., Stoppani G., Igartua U., Woodhouse J. (2025), Developing methodologies to study perceived sound qualities of violins, *Acta Acustica*, **9**(32), https://doi.org/10.1051/aacus/2025014

[3] Gurevich M., von Muehlen S. (2001), The Accordiatron: A MIDI Controller For Interactive Music, [in:] *Proceedings of the International Conference on New Interfaces for Musical Expression*, Seattle, pp. 27–29, https://doi.org/10.5281/zenodo.1176364.

[4] Hawley S.H., Chatziioannou V., Morrison A. (2020), Synthesis of Musical Instrument Sounds: Physics-Based Modeling or Machine Learning? *Acoustics Today*, **16**(1): 20–28, https://doi.org/10.1121/AT.2020.16.1.20.

[5] Iverson P., Krumhansl C.L. (1993), Isolating the dynamic attributes of musical timbre, *Journal of Acoustical Society of America*, **94**(5): 2593–2603, https://doi.org/10.1121/1.407371.

[6] Kim D., Dong H.-W., Jeong D. (2025), ViolinDiff: Enhancing Expressive Violin Synthesis with Pitch Bend Conditioning, [in:] *Proceedings of ICASSP 2025 – 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Hyderabad, India, pp. 1–5, https://doi.org/10.1109/ICASSP49660.2025.10890613.

[7] Liu Q. (2024), An FM-Wavetable-Synthesized Violin with Natural Vibrato and Bow Pressure, [in:] *Proceedings of the 2023 International Conference on Data Science, Advanced Algorithm and Intelligent Computing (DAI 2023)*, pp. 243–250, Shanghai, https://doi.org/10.2991/978-94-6463-370-2_27.

[8] McAdams S., Bruno G.L. (2012), The perception of musical timbre, [in:] *Oxford Handbook of Music Psychology*, Hallam S., Cross I., Thaut M.H. [Eds.], pp. 72–80, Oxford Library of Psychology (2008; online edn, Oxford Academic, Oxford, https://doi.org/10.1093/oxfordhb/9780199298457.013.0007.

[9] The MIDI Manufacturers Association (1996), *MIDI 1.0 Detailed Specification*, https://midi.org/midi-1-0-detailed-specification (access: 6.08.2025).

[10] The MIDI Manufacturers Association (2018), *MIDI Polyphonic Expression*, https://midimpe.neocities.org/rp53spec.pdf (access: 6.08.2025).

[11] Pérez Carrillo A.A. (2009), *Enhancing Spectral Synthesis Techniques with Performance Gestures using the Violin as a Case Study*, Ph.D. Thesis, Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona.

[12] Robertson A. (2011), Seaboard: a New Piano Keyboard-related Interface Combining Discrete and Continuous Control. [in:] *Proceedings of the International Conference on New Interfaces for Musical Expression*, Oslo, https://doi.org/10.5281/zenodo.1178081.

[13] Schoonderwaldt E., Friberg A. (2001), Towards a rule-based model for violin vibrato, [in:] *Proceedings of the Workshop on Current Research Directions in Computer Music*, pp. 61–64.

[14] Wu Y. *et al.* (2022), MIDI-DDSP: Detailed control of musical performance via hierarchical modeling", [in:] *International Conference on Learning Representations (ICLR)*, pp. 1–27