

ANALYSIS AND EVALUATION OF IRREGULARITY IN PITCH VIBRATO FOR STRING-INSTRUMENT TONES

William L. Martens

Faculty of Architecture, Design and Planning
University of Sydney, Sydney NSW 2006, Australia
Email: william.martens@sydney.edu.au

Abstract

In describing musical performances, the use of the term vibrato can imply any periodic (or quasi-periodic) fluctuation in pitch, amplitude, or timbre of a sustained musical tone. The current study focused only upon analysis and evaluation of ‘pitch vibrato’ observed in recorded performances on a number of string instruments (three violins, a viola, a cello, and a bass violin). In order to gain a better understanding of the identifiable auditory attributes associated with the perception of pitch vibrato as performed, a multi-parameter vibrato synthesis algorithm was employed in the creation of a set of stimuli for evaluation by human listeners. Besides rate and depth of pitch modulation, a third parameter was included in the synthesis that allowed for a manipulation of the quasi-periodic nature of simulated vibrato intended to mimic performed vibrato. Control for this third parameter, effectively capturing the amount of irregularity in pitch modulation, was enabled via adjustment of the Q value of a resonant low-pass filter that was used to either spread or concentrate a modulation signal’s energy around the nominal pitch modulation frequency (vibrato rate). A high Q value was associated with pitch modulation that sounded very regular, practically sinusoidal at the sub-audio vibrato rate (when the Q value exceeded around 30). Lower Q values were associated with irregular sounding pitch modulation that was heard as more rough, and could become very rough at the lowest Q values (below $Q=3$). Performances recorded without substantial pitch vibrato were processed via a delay-modulation algorithm that employed a collection of the synthesized modulation signals in an attempt to match the character and quality of vibrato performances recorded on the same instruments. A group of fifteen listeners was employed to determine how detectably different the synthetic vibrato was as the Q value was varied, and to what extent changes in Q value influenced the perceived ‘*fluctuation strength*’ of the synthetic vibrato.

1. Introduction

The perception of vibrato in musical performance on string instruments is a complex matter. For example, when a musician performs a sustained musical tone on a violin, the periodic (or quasi-periodic) fluctuation in pitch of the tone is typically associated with a coordinated fluctuation in amplitude as well as the timbre of the tone. Furthermore, the time course of the fluctuation itself is not a simple phenomenon to describe, even if a musician succeeds in the attempt to perform a nearly periodic fluctuation in pitch (recognising that perfectly periodic pitch variation is a machine-realizable goal that is out of a human performer’s reach). In contrast, the perfectly periodic pitch variation that can be produced in electronic or algorithmic synthesis of musical tones typically is perceived as artificial, and can be associated with an undesirable feature of synthetic sound. Considerable effort has been made in the design of synthesizers and vibrato effects processors to enable them to generate more

natural-sounding musical tones. It was in the interest of improving the natural character of such synthetic vibrato that the current study was undertaken, particularly through the experimental study of the perception of more complex pitch modulation than is typically applied to synthetic musical tones.

The approach taken here in order to investigate synthetic pitch modulation for musical applications was to focus on the capacity of human listeners to respond differentially to subtle variations in pitch modulation, both in terms of the detection of differences between similar musical tones undergoing slightly different pitch modulation, and in terms of the discrimination of the magnitude of those differences particularly with respect to an auditory attribute Terhardt [1] termed ‘*Schwankungsstärke*’ or ‘*fluctuation strength*.’ Subsequently, Zwicker and Fastl [2] developed a standardized scale for this attribute using a perceptual unit of measure termed the ‘*vacil*’ that was established relative to a standard reference stimulus (a 1kHz tone, reproduced at 60dB, with 100% amplitude modulation (AM) at 4Hz). Fastl [3] showed that the perceived magnitude of this auditory attribute could be matched successfully between this AM reference stimulus and tones exhibiting frequency modulation (FM), but such a matching procedure was not used in the current study, which required of listeners only that they report which of two stimuli undergoing subtle pitch modulation had the greater *fluctuation strength*.

How these subtle variations in pitch modulation were produced for the current investigation was via a delay-modulation-based process applied to recorded string instrument sounds that were performed without noticeable vibrato. For comparison sake, performances of the same notes by the same musicians on the same instruments were recorded with vibrato, which was accomplished by having the musicians perform each note in alternation between open strings and the same pitch fingered on the immediately adjacent lower string (e.g., for the violin, performances could be compared between the D on the G string versus the open D string, the A on the D string versus the open A string, the E on the A string versus the open E string). Thus a number of examples of naturally produced vibrato were available for blind comparison with very similar notes undergoing synthetically introduced vibrato. It should be noted, however, that the synthetically introduced vibrato did not exhibit the coordinated fluctuation in amplitude and timbre typical of natural tones, as the delay-modulation-based process creating pitch modulation held constant both overall amplitude and relative amplitude of spectral components. That being said, such non-pitch-related features must be difficult to detect given the small excursions in pitch modulation studied here, as will be revealed subsequently in the results of the current study.

The current study was thus focused primarily upon analysis and evaluation of ‘pitch vibrato’ heard in reproduced string instrument performances. In order to gain a better understanding of the identifiable auditory attributes besides that describing as the *fluctuation strength* of perceived pitch vibrato, an additional attribute associated with the irregularity of the vibrato was manipulated via a multi-parameter vibrato synthesis algorithm. Besides rate and depth of pitch modulation, a third parameter was included in the synthesis that allowed for control of the amount of chaotic behaviour of the quasi-periodic simulated vibrato that was intended to mimic the irregularity in performed vibrato.

2. Methods

2.1 Stimuli

Although the stimulus preparation for the current study was based upon an analysis of pitch vibrato observed in recorded performances on a number of string instruments (three violins, a viola, a cello, and a bass violin), the sound recordings selected for presentation in the two experiments reported here were taken from performances on a single violin, recorded on an open string (i.e., with no pitch vibrato). The rate and depth of pitch modulation, which was implemented via a delay-modulation-based process (as described in Dattorro [5]), were set respectively to a constant value of 5.5 Hz and a maximum delay of 1.2 ms. Since the fractal process that was used in the generation of the modulation signals had the typical spectral energy distribution termed ‘ $1/f$ ’ (see [4] for background on this detail), a resonant low pass filter was used to ensure that the modulation signal had its peak energy at the desired frequency, which in this case was the 5.5 Hz average observed in the analysis of vibrato performances on the six string instruments investigated here (an analysis not examined in this paper).

For a number of fractal sequences, the Q value of the filter was adjusted to one of three values in order to vary the amount of regularity in the modulation signals, yielding the three filter responses shown in the left panel of Figure 1. Three examples of the modulation signals that resulted from applying such filtering to the fractal sequences are shown in the right panel of Figure 1. Two of the plotted modulation signals resulted from applying the resonant low pass filter with $Q=30$, while one of the plotted modulation signals resulted from a low pass filter setting of $Q=3$. Although it may be clear from visual inspection that it is the middle plot that differs from the topmost and bottommost plots (a visual distinction that is aided by red versus blue colour in this graph), the auditory detection of this difference was fairly difficult when the plotted signals were used to control pitch modulation.

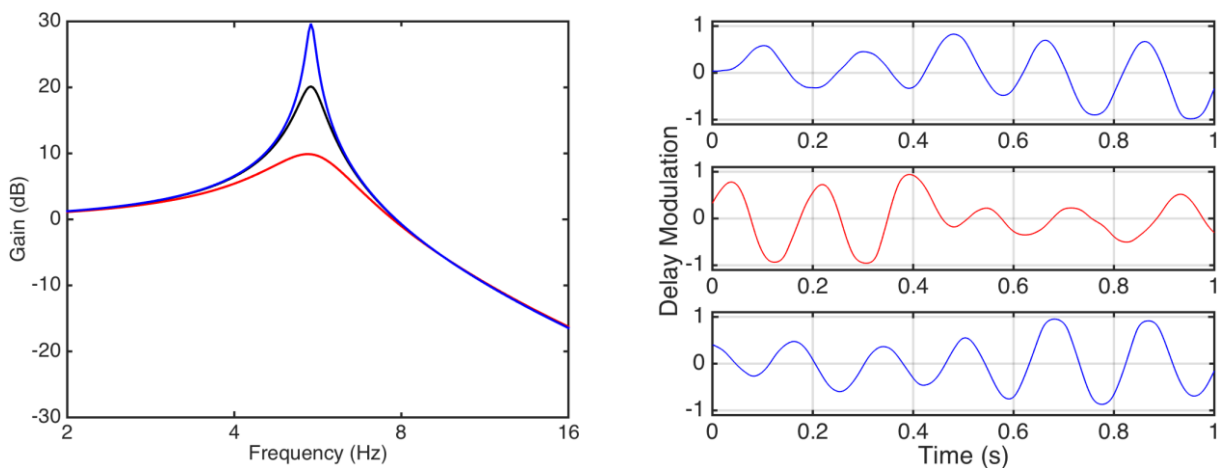


Figure 1. The left panel shows a set of three responses for the resonant low pass filter that was used to manipulate the frequency content of the modulation signals used in stimulus generation for the current study, each of which show peak energy at 5.5 Hz, but having bandwidth values set to $Q=3$ (red curve, with lowest gain at 5.5 Hz), $Q=30$ (blue curve, with highest gain at 5.5 Hz), and $Q=10$ (black curve, in between the other curves). The right panel shows three examples of the modulation signals that resulted from applying such filtering to the fractal sequences. The y-axis values are normalised delay values that are multiplied by the desired maximum delay value before being applied to the signal by the delay-modulation signal processing routine. Note that the topmost and bottommost plots in the right panel resulted from applying the resonant low pass filter with $Q=30$ (curves drawn in blue), while the other plotted modulation signal resulted from a low pass filter setting of $Q=3$ (the middle plot, with curve drawn in red).

2.2 Procedure

Two experiments were designed to test human listeners on their perception of synthetic vibrato, as the filter Q applied to the delay-modulation signal was varied between three values for a number of fractal sequences. The first experiment required listeners to make three-alternative forced choice (3AFC) judgments regarding how detectably different the synthetic vibrato sounded as the Q value was varied. In this 3AFC detection task, listeners were instructed to “pick the odd one out” when presented with a sequence of three similar musical tones. The second experiment employed a paired-comparison paradigm in which listeners made two-alternative forced choice (2AFC) judgments regarding which of two synthetic vibrato tones exhibited the greater perceived ‘*fluctuation strength*,’ when those two synthetic vibrato tones differed predominantly in terms of the filter Q applied to the modulation signal for each tone. Listeners were acquainted with the use of the term ‘*fluctuation strength*’ through an introductory lecture featuring sound examples in order to avoid any possible confusion about the auditory attribute to which they were to attend. It should be clear from context that the rates of modulation for the vibrato tones in the current study correspond to only one of the two different kinds of auditory sensation that listeners might experience when listening to modulated signals, this difference depending on the speed of modulation. In the case of low modulation frequencies (typically less than about 20 Hz) the resultant sensation has been termed ‘*fluctuation strength*’ rather than

‘roughness’ (as originally distinguished by Terhardt [1], and subsequently reinforced by Fastl and Zwicker [2]). No listeners reported any difficulty in understanding either of the two tasks set for them. A total of 15 listeners participated in the first experiment, with 10 participating in the second experiment. None reported any hearing loss.

2.2.1 Discrimination Task

The 2AFC discrimination task was designed to determine whether one synthetic vibrato tone, produced with higher-Q modulation, would be heard as having greater perceived ‘*fluctuation strength*,’ when compared to another synthetic vibrato tone produced with lower-Q modulation. As was the case for the 3AFC task, the 2AFC task again was double blind, but not just in terms of which combinations of stimuli were presented on each trial. Indeed, the listeners were unaware of what combinations of modulation parameters were to be presented in each pair of stimuli, and the parameters under investigation also were not revealed until afterward completion of all experimental trials). Also, to be clear about what differed for each pair of stimuli, it should be stated that between the two string sounds that were compared on each trial, the only difference was the Q value (i.e., one recorded string sound was given vibrato of two different types). Assuming that the two tones were matched in apparent modulation rate and depth, the task from the listener’s standpoint was simply to judge which stimulus had the greater fluctuation strength, the first or the second (which randomly varied between that with higher-Q modulation and that with lower-Q modulation).

2.2.2 Detection Task

The 3AFC detection task was designed to determine whether one synthetic vibrato tone, produced with lower-Q modulation, could be detected as different from two other synthetic vibrato tones produced with higher-Q modulation. This double-blind task was made difficult by having all three modulation signals generated using independent fractal sequences, so it was not simply a matter of finding which two signals were identical in order to determine which one was processed using the lower-Q modulation signal. From the listener’s standpoint, the task was to find which of three tones seemed to have a greater apparent irregularity in its pitch modulation than the other two, even though all three potentially had differing apparent irregularity (though two had a matching filter Q value).

3. Results

3.1 Discrimination Task

For the 2AFC discrimination task, the hypothesis underlying the preferred method of paired-comparison data analysis is that the stimuli can be arranged along a linear perceptual scale, which is associated with the verbal descriptor “*fluctuation strength*” in the current case. The reasoning is as follows: When listeners are presented with two sounds, they may not make the same dominance judgments, and so the proportion of times that one stimulus is chosen over another is taken as a measure of the extent to which one stimulus dominates the other in terms of the attribute of interest. Indeed, the data collected for the group of ten listeners revealed that none of the compared stimuli dominated any of the other stimuli unanimously (i.e., there was always some disagreement). Nonetheless, some stimuli were most often dominating all other stimuli for most listeners, as can be seen by the proportions presented in Table 1.

According to the hypothesis underlying the paired-comparison data analysis, the choice proportions reported in Table 1 may be analysed to yield a coordinate for each of the six stimuli along a linear perceptual scale following Thurstone’s [8] indirect scaling method. The first step was to convert the choice proportion data into the Z-Score values shown in Table 2, which are taken to indicate the magnitude of the underlying perceptual differences between pairs of stimuli. The final row of Table 2 shows the sum of the values in each column, which constitute the scale values determined for each stimulus in manner consistent with Thurstone’s [8] Case IV. The values on this derived scale are effectively normalized so that the sum of the six values is equal to zero, with the

negative values balancing the positive values assigned to stimuli along the scale. The left panel of Figure 2 plots these derived scale values as a function of the Q value of the stimuli, with results for each of the two fractal sequences distinguished by the plotting symbols, cyan-coloured circles for the first sequence, and yellow-coloured squares for the second sequence. Thus, the results of the 2AFC paired-comparison discrimination experiment reveal that different fractal sequences can produce greater differences in perceived ‘*fluctuation strength*’ than the differences introduced by varying the Q value of the filter applied to the modulation signal, although the increase in ‘*fluctuation strength*’ with increasing Q value is also quite clear in the plot.

Table 1. The paired-comparison data collected from ten listeners who indicated on each of 15 trials which of two stimuli had the greater ‘*fluctuation strength*.’ The upper triangular matrix shows the proportion of trials on which the column stimulus dominated the row stimulus (i.e., ‘C>R’), and the lower triangular matrix is derived from the upper triangular matrix by subtracted the observed proportion from 1. The values on the diagonal (in cells coded red) were set to 0.5 based upon the assumption that this proportion best estimates that which characterises the expected result when comparing two identical stimuli. The first three columns correspond to stimuli generated using the first set of three fractal modulation signals (coded cyan), and the last three columns correspond to stimuli generated using the second set of three fractal modulation signals (coded yellow).

Prop C>R Fluct. Str.	Q=3_(1)	Q=30_(1)	Q=10_(1)	Q=3_(2)	Q=30_(2)	Q=10_(2)
Q=3_(1)	0.5	0.6	0.6	0.9	0.9	0.8
Q=30_(1)	0.4	0.5	0.3	0.1	0.7	0.7
Q=10_(1)	0.4	0.7	0.5	0.8	0.8	0.9
Q=3_(2)	0.1	0.9	0.2	0.5	0.2	0.4
Q=30_(2)	0.1	0.3	0.2	0.8	0.5	0.7
Q=10_(2)	0.2	0.3	0.1	0.6	0.3	0.5

Table 2. Z-Score values that were computed for the proportions shown in Table 1 comprise the first six rows of the matrix, which are followed by a final row showing the sum of the values in each column, which constitute the scale values determined for each stimulus in manner consistent with Thurstone’s [8] Case IV indirect scaling method. As in Table 1, the values on the diagonal (in cells coded red) were derived from the assumed proportions associated with the expected result when comparing two identical stimuli. Again, the first three columns correspond to the first set of three fractal modulation signals (coded cyan), and the last three columns correspond to the second set of three fractal modulation signals (coded yellow).

Z-Score Fluct. Str.	Q=3_(1)	Q=30_(1)	Q=10_(1)	Q=3_(2)	Q=30_(2)	Q=10_(2)
Q=3_(1)	0	0.25	0.25	1.28	1.28	0.84
Q=30_(1)	-0.25	0	-0.52	-1.28	0.52	0.52
Q=10_(1)	-0.25	0.52	0	0.84	0.84	1.28
Q=3_(2)	-1.28	1.28	-0.84	0	-0.84	-0.25
Q=30_(2)	-1.28	-0.52	-0.84	0.84	0	0.52
Q=10_(2)	-0.84	-0.52	-1.28	0.25	-0.52	0
SUM	-3.91	1.01	-3.24	1.94	1.28	2.92

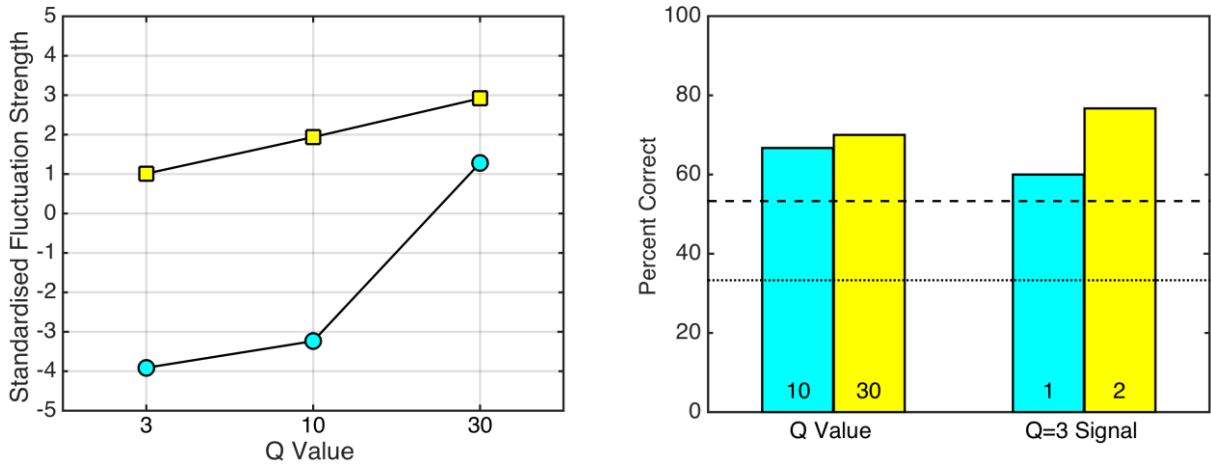


Figure 2. The left panel shows the results of the 2AFC paired-comparison discrimination experiment, in which listeners made two-alternative forced choice judgments regarding which of two stimuli had the greater apparent ‘*fluctuation strength*.’ The right panel shows the results of the 3AFC detection experiment, in which listeners chose which of three tones seemed to have a greater apparent irregularity in its pitch modulation than the other two. The first two bars (labelled ‘10’ and ‘30’) show the percent correct detection of the ‘odd one out’ that corresponded to the stimulus with differing Q-value (i.e., a Q value of 3). The second two bars (labelled ‘1’ and ‘2’) show the percent correct detection of the odd stimulus for the two different fractal sequences generated for the Q=3 modulation signal, irrespective of whether the filter Q value of the other two stimuli was equal to 10 or 30. Note that the cyan and yellow colour coding in the left panel corresponds to the rightmost pair of bars in the right panel labelled ‘1’ and ‘2’ (and not the bars labelled ‘10’ and ‘30’).

3.2 Detection Task

For the 3AFC detection task, the analysis was quite a bit simpler than it was for the 2AFC discrimination task. The reasoning here is that listeners hearing three stimuli in sequence may choose the ‘odd one out’ (i.e., the stimulus with the differing Q value of 3) by chance alone on 33% of all trials. Therefore, the plot of the percent correct detection rates in the right panel of Figure 2 includes a horizontal dotted line at the 33% level. For a statistically significant percent correct detection of the odd stimulus (having lower-Q modulation) at an error criterion of $p < .05$, the observed percent correct rate must exceed the 53% level, which is indicated in the figure with a dashed line. As can be seen in the right panel of Figure 2, all four cases examined exceed this criterion 53% level of performance. Therefore, the experimental results for the 3AFC detection task support the conclusion that vibrato resulting from higher-Q modulation (whether at Q=10 or Q=30) can be distinguished from that resulting from lower-Q modulation (at Q=3). The plotted results also indicate that when the odd stimulus was generated using the first of two different fractal sequences, a lower percent correct detection rate was observed than when using the second of two different fractal sequences. This result is consistent with the results from the 2AFC paired-comparison discrimination experiment, in that the stimuli with greater apparent ‘*fluctuation strength*’ (plotted using yellow-filled square symbols) were also the stimuli associated with higher percent correct detection rates (plotted using yellow-filled bars).

4. Conclusions

The results of the two experiments reported in this paper show that subtle variations in synthetic vibrato may be detected and discriminated by human listeners under controlled conditions. In particular, for musical tones performed by a given musician on a given instrument, with and without vibrato, informal evaluation indicated that the employed synthetic vibrato unit can be used to process string-instrument performances recorded without vibrato to produce an output with vibrato sounding

similar to that performed on the same instrument. Although the current study made no direct experimental comparison between these synthetic vibrato stimuli and stimuli exhibiting vibrato as naturally performed, the synthetic vibrato parameters were set to produce outputs matching such naturally performed notes. Besides rate and depth of pitch modulation, however, the focus of the two experiments was upon the influence on vibrato character afforded by the manipulation of a third synthetic vibrato parameter designated by a Q value that controlled the amount of irregularity in the quasi-periodic nature of resulting simulated vibrato. The results of the first experiment revealed the extent to which changes in Q value influenced the perceived fluctuation strength of the synthetic vibrato. The results of the second experiment showed how detectably different the synthetic vibrato sounded as the Q value was varied. An important implication of the current study's results is that different fractal sequences can produce greater differences in perceived '*fluctuation strength*' than the differences introduced by varying the Q value of the filter applied to the modulation signal, and this strong dependence on differences between fractal sequences is a factor that must be taken into account in future studies using the three-parameter synthetic vibrato processing that was employed here.

Considering the effort that has been made to design of musical sound synthesizers to produce natural-sounding vibrato, and the continued development of vibrato effects processors for use in creating popular guitar sounds [5], it is remarkable that more research into quasi-periodic pitch vibrato has not been reported. As the desire to enable the generation of more natural-sounding synthetic vibrato was one of the primary motivations for the current study, there is still a need to connect the current efforts to the more applied research and development that could bridge the gap between laboratory studies and product-driven investigations. Therefore, additional analysis of natural vibrato performance is underway, and will involve blind tests of natural versus delay-modulation-based vibrato effects. In addition to planned comparisons between naturally produced vibrato and the synthetic vibrato investigated in the current study, future work will address potential concerns regarding the musically useful ranges of the Q values for a representative sample of delay-modulation-based synthetic vibrato rates and depths, using methods such as those taught in Martens and Marui [6].

References

- [1] Terhardt, E. "Über akustische Rauigkeit und Schwankungsstärke" (On acoustic roughness and fluctuation strength), *Acustica*, **20**, 215-224, (1968).
- [2] Fastl H. and Zwicker, E. "Fluctuation strength", *Psychoacoustics – Facts and Models*, Springer, Berlin, pp. 247-256, 2007.
- [3] H. Fastl, "Fluctuation strength of modulated tones and broadband noise", *Hearing – Physiological Bases and Psychophysics*, Springer, Berlin, pp. 282-286, 1983.
- [4] Evangelista, G. "Fractal modulation effects", *Proceedings of the 9th International Conference on Digital Audio Effects (DAFx-06)*, Montreal, Canada, 18-20 September 2006.
- [5] Dattorro, J. "Effect design, Part 2: Delay-line modulation and chorus", *Journal of the Audio Engineering Society*, **45**(10), 764-788, (1997).
- [6] Martens W.L. and Atsushi, M. "categories of perception for vibrato, flange, and stereo chorus: mapping out the musically useful ranges of modulation rate and depth for delay-based effects", *Proceedings of the 9th International Conference on Digital Audio Effects (DAFx-06)*, Montreal, Canada, 18-20 September 2006.