CATEGORIES OF PERCEPTION FOR VIBRATO, FLANGE, AND STEREO CHORUS: MAPPING OUT THE MUSICALLY USEFUL RANGES OF MODULATION RATE AND DEPTH FOR DELAY-BASED EFFECTS

William L. Martens and Marui, Atsushi

Sound Recording Area
Schulich School of Music, McGill University
Montreal, Canada
wlm@music.mcgill.ca

ABSTRACT

Vibrato, Flange, and Stereo Chorus are perhaps the three most often used digital audio effects that are created by smoothly modulating the duration of a delay line at typically sub-audio rates. Common practice is to use a periodic or quasi-periodic modulation control signal with frequency roughly between 2 and 9 Hz, and both the rate and depth of delay modulation are typically adjusted according to the aesthetic criteria of a performer or by an audio production engineer. In order to establish norms for the musically useful range of modulation rate and depth for such delay-based effects, 25 listeners were asked to make categorical judgments regarding their perception of vibrato, flange, and stereo chorus effects. The results map out for these two modulation parameters three perceptual regions for these three related effects: the region in which modulation is too subtle for effective use, the parameter ranges that seem most musically useful, and the region in which it is too extreme for most musical applications. Of particular interest is the observed commonality between these perceptual regions for vibrato, flange, and stereo chorus effects.

1. INTRODUCTION

Among the most popular digital audio effects (DAFx) that are used in modern music production, there are three effects that are very frequently used, and which share a common underlying factor: They are created by smoothly modulating the duration of a delay line at typically sub-audio rates. The three DAFx are often identified as vibrato, flange, and stereo chorus [1][2][3, Chapter 3]. Although these three algorithms are well understood, there are no clear norms for the musically useful range of modulation rate and depth for delay lines underlying these effects, which must be based upon the perceptual attributes associated with the application of the effects. It should be clear that answering this question from the perspective of signal processing *per se* cannot be entirely satisfying. So the motivation for the study reported here was to map out for these two modulation parameters the distinct perceptual regions describing where they might be used most effectively. The method by which this was done was to ask a group of listeners to make categorical judgments regarding the perception of each effect as follows: the perceived modulation should be indicated to be either too subtle for effective use, quite suitable for musical use, or too extreme for most musical applications.

There is some precedent in the literature for this type of study, which should not be regarded as a study of just noticeable differences in frequency variation [4]; rather, this study is more closely related to studies of fluctuation strength, which is one of the major psychoacoustic variables considered in sound-quality evaluation. The authors know of only one study, conducted by Wickelmaier and Ellermeier [5], in which the dependence of fluctuation strength on modulation rate and modulation depth was tested in a factorial design, as it was here. Of course, the current study did not investigate fluctuation strength directly, but only asked for categorical distinctions with regard to a variety of temporal fluctuations that are most definitely multidimensional in perceptual character, hence the use of the different terms vibrato, flange, and stereo chorus.

In order to regularize the signal processing used in stimulus preparation for this study, a particular structure was chosen as the basis for these three DAFx, which indeed may be described as linear, time-variant systems that are based upon similar algorithmic structures. But a few simple differences in structure result in the three DAFx having qualitatively different perceptual characteristics, which are well identified by their names. In the simplest of implementations, the three types of delay modulation effects can be created using only a single interpolated delay line as the core component. The only difference is in the presence and/or destination of the feed-forward signal path that may be added to the core component. (A fourth effect that has been popular in guitar effects processing, termed phase shifting, has been omitted from the current study primarily because it does not match the strict delay-modulation structure on which the other three are based [6].)

Vibrato, as implemented here, is a primarily a pitch-related effect that is based upon a simple delay modulation with no feed-forward signal. Flange is primarily a timbral effect that results from the zeros of a comb-filter sweeping through frequency due to the summation of the a feed-forward signal with the vibrato's delay modulation signal. Under some circumstances, the pitch variation can still be perceived here, but it is usually overshadowed by the timbral effect. The final option to be described here is when the feed-forward signal is sent to a different audio output channel than the vibrato's delay modulation signal. Again, the pitch variation virtually disappears, and what is left is perceived as a spatial effect that has been identified as a stereo chorus effect. In it's most common implementation, stereo chorus effects are implemented with more than one delay line, but for the current investigation, only the simplest implementation has been considered in order to maximize similarity with the other two algorithms. In fact, feedback signals are also often included in the implementation of both flange and stereo chorus effects, and frequently play a role in the musical "sweetening" of reverberation algorithms [7].

Of course, by choosing only very simple implementations for these three effects, there is a chance that the results of the study will not be so readily generalized to other implementations. It is preferred to begin with these simple algorithms to establish a baseline for further work. It should also be noted that the stimuli may be characterized by the outputs of the effects processing, rather than in terms of the processing itself. This provides a basis for comparisons with other implementations that transcend the particular structures employed here. Other studies of these effects might include more complete characterization of their possibilities, such as that described in the review of different types of vibrato by Verfaille et al. [8]. They explained that the acoustical effects of vibrato could be categorized into amplitude modulation, frequency modulation, and spectral envelope modulation. For the current investigation, pitch modulation alone was intended for the vibrato stimulus, in order to keep it as distinct as possible from the timbral effects of flange outputs that are associated with spectral envelope modulation, within the limitation of the simplistic model. Of course, vibrato effects that are created by simple delay modulation, as were those created for the current study, also exhibit a limited amount of modulation in spectral energy distribution. Nonetheless, for the guitar-sound inputs processed herein, which were relatively smooth spectrally, the slight variation in tone color (subtle brightening and darkening) was negligible, especially since the maximum depth of delay modulation never exceeded 1 ms (see below for details). The choice to limit the vibrato production to simple delay modulation (with all-pass interpolation for fractional-sample delay values) also served to keep the vibrato directly comparable to the other delay-modulation-based effects of interest here, most notably in terms of its simplistic signal processing implementation. It is also true that the output vibrato character that was introduced into the guitar performance that was used as the input to these effects was quite similar to that which can be produced by a human performer, with the exception that the modulation followed a perfect sinusoidal pattern.

The vibrato discussed in this paper is an instance of coherent vibrato [9] which means all the harmonic components in the output signal are modulated at the same rate and depth. (In the case of non-coherent vibrato, a particular frequency stands out. from the rest). In the current study, output signals of the vibrato process exhibit some dynamic frequency deviation relative to the input signal due to the same phenomenon as the Doppler shift (as if the input signal were "moving closer and farther" relative to the listener's position). Frequency deviation of the output signal from the carrier frequency can be expressed as a function of modulation rate and modulation depth parameters. According to Dutilleux and Zölzer [3, Chapter 4], a *resampling factor* for sinusoidal modulation can be regarded as a pitch change ratio and equal to the ratio of the instantaneous output frequency divided by the carrier frequency. The pitch change ratio is a function of time, and it correspond to the frequency deviation, as follows:

$$\alpha(t) = 1 - 1/1000000 \cdot \mathrm{DEPTH} \cdot \mathrm{RATE} \cdot 2\pi \cdot \cos(\mathrm{RATE} \cdot 2\pi t) \quad (1)$$

where DEPTH is the modulation depth in μs (thus $1/1000000$ in front of it to match the unit) and RATE is modulation rate in Hz (cycles/second). The minimum and the maximum values of the resampling factor are the ratios of the peak deviation of the instantaneous frequency from the carrier frequency, and the reach these extremes when the above cosine term reaches 1 and -1, respectively.

$$1 \mp 1/1000000 \cdot \mathrm{DEPTH} \cdot \mathrm{RATE} \cdot 2\pi \quad (2)$$

Since this equation is a ratio of the instantaneous frequency divided by the carrier frequency, the peak deviation frequency can be obtained by multiplying the ratio times carrier frequency value.

2. METHOD

2.1. Stimulus Generation

A single short excerpt of an electric guitar performance was digitized at 44.1 kHz in 16-bit precision, and stored in a file for subsequent processing. The performance was a single note played on a electric guitar (the note G produced by hammering the string with the left-hand just below the 5th fret on the D string). The analogue signal from the guitar was submitted to mild compression and and overdrive effects to achieve a sound character similar to that used in popular music.

This signal was then submitted to each of the three delay-modulation-based effects processing algorithms, employing a factorial combination of delay modulation depths and rates to complete a two-dimensional palette of output character for each. Modulation rate was set to each of five frequencies: 2, 3, 4, 6, and 9 Hz. Modulation depth at peak was set to five different delay-line durations spaced logarithmically between 0.04 and 1.0 ms. The character of the output sound was naturally dependent on the shape of the modulation signal. Although a more musically pleasing result can be obtained through quasi-periodic modulation, a simple, gated sinusoidal modulation signal was employed in this study. The stimulus duration was 1.4 s, and the modulation signal was gated by a raised-cosign window of 200 ms rise and fall, leaving the modulation at its maximum for exactly 1 s.

For the flange and vibrato, the average delay value about which the delay modulated was 1 ms, which is of no consequence for pitch vibrato, but matters very much for the timbral modulation resulting for the flange stimuli. In the case of the flange effect, this means that the timbral variation was the most extreme possible (since at maximum modulation depth, the signals begin added exhibited very short delays). In the case of the stereo chorus effect, the non-modulated signal was delayed by .5 ms, which means that the delay modulated signal fluctuated between leading and lagging the non-modulated signal in time of arrival. This means that the spatial variation was like that associated with with the variation in Interaural Time Delay (ITD) for a sound source taking a circular path around the head, since maximum modulation depths for ITD don't often exceed .6 to .7 ms.

The stimuli were presented over either one or two JBL Control 8SR studio monitors in a large anechoic chamber, at a level producing 80 dB SPL at the listening position. For the vibrato and flange stimuli, subjects faced one of the two monitors. When listening to the stereo chorus stimuli, subjects oriented themselves to position the two monitors 30° to the left and right of the median plane.

2.2. Categorical Judgment Task

As a straightforward means for operationally defining the musical useful range of modulation rate and depth for delay-based effects, subjects were asked to make categorical judgments regarding their perception of vibrato, flange, and stereo chorus effects. The following three categories for response were explained prior to the experiment, the perceived modulation:

- would be too subtle for effective use,

- seems quite musically useful,
- would be too extreme for most musical applications.

While subjects were familiarized with the stimuli, it was explained that their responses should be limited to the three categories listed above. The three categorical judgment tasks for the three types of delay-modulation-based effects were completed in succession on a single day, but not in a randomized order. All subjects completed the task for the vibrato case first, followed by the flange case, and finally the stereo chorus case.

2.3. Subjects

Twenty-five subjects participated in the rating task. All 25 subjects were either undergraduate or master's level computer science students at the University of Aizu (Japan), with interest in audio effects processing but without strong musical background. Their age ranged between 18 and 24, and all reported that they enjoyed listening to popular music on a daily basis.

3. RESULTS AND ANALYSIS

In order to summarize the relative proportion of responses within each of the three response categories, logistic regression analysis (LRA) was employed to fit a curve to the observed categorical choice data summed across all 25 subjects. In commonly used linear regression models, continuous data are submitted for both independent variables and dependent variables. Although non-continuous variables can be used, evaluation of the linear regression results given such data have to be made with caution. LRA is an instance of nonlinear regression method suitable for analyzing a binary dependent variable against continuous independent variable [10, Chapter 13][11, Chapter 11]. An example of using LRA is fitting a probability curve on examination outcomes on test scores which the result of an exam is often binary (pass or fail) and predictor variable (examination score) is continuous. In LRA, a sigmoid (i.e., "S"-shaped) function called *logit* function shown in Equation 3 is used to fit to data.

$$\hat{Y} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots)}} \tag{3}$$

where X_i denotes a continuous independent variable, and $\hat{Y}$ denotes a predicted dependent variable taking values between zero and one.

In the current investigation of delay modulation based effects, parameter values for modulation depth and modulation rate were submitted to the analysis as X and binary value of whether each rating category was chosen as Y. Since there were three response options, the analysis was performed first for the response of "too weak effect" versus other responses. Then the analysis was performed for the response of "too much effect" versus other responses. The curve for the proportion of "musically useful" responses was derived by subtracting the other two curves from unity, since the three response portions were constrained to sum to unity.

Results are shown in Figure 1 for the vibrato case, and results for the other two effects, flange and chorus, looked remarkably similar (and are not presented in this paper in order to reduce its length). The figure presents a plot showing the logit functions fit to the response-proportion distributions obtained from the three-category response task for one of tested stimulus sets. The case shows results of analyzing responses to vibrato stimuli presented at a fixed modulation rate of 6 Hz, and with modulation depth varying between 40 and 1000 μs. Under the left-most curve, the term 'sub-' is written to indicate that for this range of modulation depth values, the resulting vibrato was perceived to be sub-useful. For the right-most curve, the term 'supra-' is written to indicate that these modulation depth values produce too strong a resulting vibrato to be judged as musically useful. The middle curve plots the remaining proportion of the responses that define the useful range of modulation depth values at this modulation rate. The green vertical dashed line marks the transition from sub-useful to useful, and the red vertical dashed line marks the transition from useful to supra-useful modulation depths.

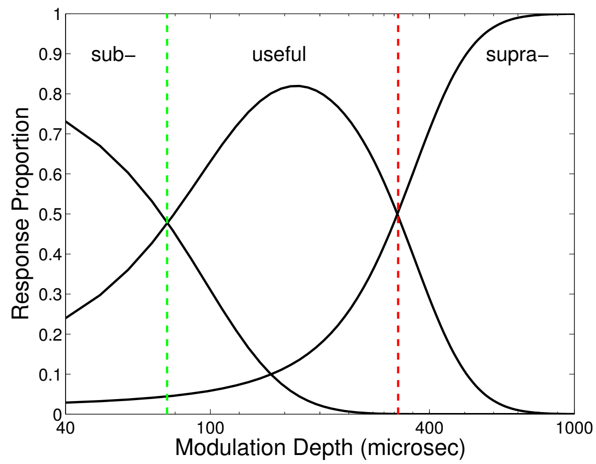

Figure 1: *A response proportion plot showing the logit functions fit to the distributions obtained from the three-category response task, in this case given in response to vibrato stimuli with modulation depth varying between 40 and 1000 μs, all presented at a fixed modulation rate of 6 Hz. Under the left-most curve, the term 'sub-' is written to indicate that for this range of modulation depth values, the resulting vibrato was perceived to be sub-useful. For the right-most curve, the term 'supra-' is written to indicate that these modulation depth values produce too strong a resulting vibrato to be judged as musically useful. The middle curve plots the remaining proportion of the responses that define the useful range of modulation depth values at this modulation rate. The green vertical dashed line marks the transition from sub-useful to useful, and the red vertical dashed line marks the transition from useful to supra-useful modulation depths.*

The fact that a very similar pattern of results was observed for all three effects is underscored in the summary plot shown in Figure 2. In the figure, the boundaries between three perceptual regions are shown by a single common curve for three digital audio effects tested. These boundaries are plotting in the same manner as those in the contour plot shown in Figure 1. Transition points in the control space for the lower boundary of the musically useful region are marked using downward-pointing triangles as plotting symbols. Transition points for the upper boundary of the useful region are marked using upward-pointing triangles. Due to the curvilinear relation between modulation rate and the boundary values of modulation depth defining the useful region of modulation-based effects, the depth values were regressed upon modulation period (in seconds), the inverse of modulation rate (in Hz). Submitting the inverse data to regression effectively linearized the curve, and

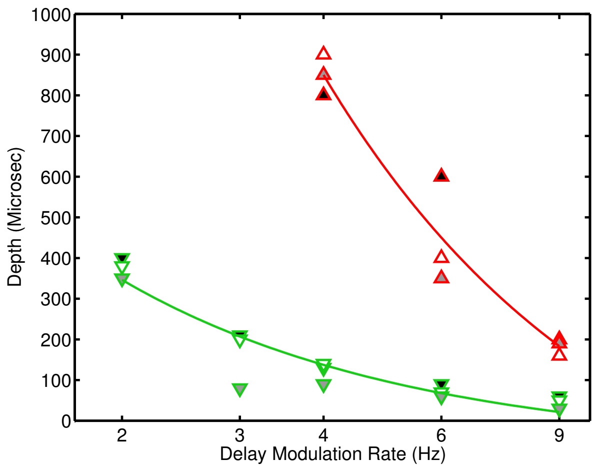

Figure 2: *Plot showing curves fit to the transition points between adjacent perceptual response regions within the control space defined by two delay-modulation parameters, modulation rate and modulation depth. Points associated with the lower boundary of the musically useful region are plotted using downward-pointing triangles, and points associated with the upper boundary of the useful region are marked using upward-pointing triangles. Three different symbol colors denote effect types (black: vibrato, white: flange, and grey: chorus). For the modulation depth data derived for all three types of effects, a simple linear regression was executed to fit a prediction equation for the lower boundary of the useful region using modulation period as the independent variable. The resulting equation accounted for 85% of the variance in modulation-depth, with a slope of 814, and an intercept of* -66 *(these are in dimensionless units, but large because the independent variable, modulation period, was measured in seconds, and the dependent variable, modulation depth, was measured in* μ*s). The F statistic for the lower boundary regression was 71.52, significant at* $p < .01$*. The equation fit to the upper boundary accounted for 94% of the variance, but could only be fit for modulation rates 4, 6, and 9 Hz, since at lower rates the modulation-depths at which these effects were tested always gave usable results (hence no boundary could be observed at 2 and 3 Hz). At least within this restricted range, the equation fit the derived data with a slope of 4800, and an intercept of* -350*. The F statistic for the upper boundary regression was 140.37, again significant at* $p < .01$.

decreased the residual error by a factor of 2 and 3, for upper and lower boundaries, respectively). To depict the curvilinear relations, however, the figure shows the regression results plotted on modulation rate. The regression analysis is summarized in the figure caption.

4. CONCLUSION

The results map out, for two delay-modulation parameters, three perceptual regions for three related digital audio effects. Most importantly, the results indicate the parameter ranges that seem most musically useful for a representative group of young adult listeners. While the generality of the results may be somewhat limited by the narrow scope of the stimuli presented here, the success of the methods used here implies that similar tests might well answer questions about optimal parameter ranges for other applications and implementations. An important caveat to make in this regard is that the musical usefulness of a particular delay-modulation-based effect must depend not only upon the processing parameters (e.g., depth and rate), but also upon input signal characteristics. This is a factor not addressed adequately in the current study, and calls for further investigation of the how to predict musical usefulness of delay-modulation-based effects from an analysis of the output signal itself. It is likely, however, that generalization is possible for guitar-sound inputs in regard to specifying the region of parameter values that would likely be found to produce a result judged too extreme for most musical applications. The fact that the same limits are observed for all three types of effects implies that most implementations of these effects will have similar limits, since many implementation will be less different from the three examples tested here than those examples are from one another. This observed commonality between vibrato, flange, and stereo chorus effects may be particularly helpful in predicting where limits will be for novel DAFx that are based upon modulated delay lines.

5. ACKNOWLEDGMENTS

This work was partially supported by McGill University's Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT), and by the National Sciences and Engineering Research Council of Canada (NSERC).

6. REFERENCES

[1] C. Roads, *Computer Music Tutorial*. MIT Press, 1998.

[2] J. Dattorro, "Effect design, part 2: Delay-line modulation and chorus," *J. Audio Eng. Soc.*, vol. 45, no. 10, pp. 764–788, Oct. 1997.

[3] U. Zölzer, *DAFx—Digital Audio Effects*. John Wiley & Sons, 2002.

[4] E. Zwicker and H. Fastl, *Psychoacoustics*, 2nd ed., ser. Springer Series in Information Sciences. Springer-Verlag, 1999, vol. 22.

[5] F. Wickelmaier and W. Ellermeier, "Scaling the perceived fluctuation strength of frequency-modulated tones," *Ac. Soc. of Am.*, vol. 115, no. 5, p. 2600, May 2004.

[6] W. M. Hartmann, "Flanging and phasers," *J. Audio Eng. Soc.*, vol. 26, no. 6, pp. 439–443, June 1978.

[7] J. Dattorro, "Effect design, part 1: Reverberator and other filters," *J. Audio Eng. Soc.*, vol. 45, no. 9, pp. 660–684, Sept. 1997.

[8] V. Verfaille, C. Guastavino, and P. Depalle, "Perceptual evaluation of vibrato models," in *Proc. Colloq. Interdisciplinary Musicology (CIM'05)*, Montreal, Quebec, Canada, Mar. 2005.

[9] W. M. Hartmann, *Signals, Sound, and Sensation*. Woodbury, NY: AIP Press, 1997.

[10] J. Neter, M. H. Kutner, C. J. Nachtcheim, and W. Wasserman, *Applied Linear Statistical Models*, 4th ed. WCB/McGraw-Hill, 1996.

[11] J. L. Fleiss, B. Levin, and M. C. Paik, *Statistical Methods for Rates and Proportions*, 3rd ed. John Wiley & Sons, 2003.