EFFECTS OF VIBRATO ON PITCH PERCEPTION

Regis Verdin	Deepak Gopinath	Alan Zhang	Brian Lee	Roozbeh Khodambashi
GTCMT	GTCMT	CoC	ECE	GTCMT
rverdin6@gatech.edu	deg3@gatech.edu	dingtianzhang@gatech.edu	blee98@gatech.edu	khodambashi@gatech.edu

ABSTRACT

In this paper, we investigate the effects of vibrato on the perception of pitch. 22 subjects participated in the experiment; each reporting perceived central pitches when hearing tones with different vibrato parameters. Two vibrato parameters were used as independent variables: rate and extent. We hypothesized that high rate and low extent of vibrato would yield the smallest errors when reporting central frequency of the vibrato. Using ANOVA and correlation analysis, the hypothesis was partly confirmed: high extent was shown to increase the subject's error with a correlation coefficient of +0.5681. Rate had no significant effect on the subject's error.

1. INTRODUCTION

Vibrato is a common feature in music performance, used by many instruments capable of varying pitch continuously. The vibrating quality which vibrato produces is primarily the result of frequency modulation. In other words, the performer is regularly moving their pitch above and below the note they are playing/singing. Amplitude modulation also plays a small role in instrumental and vocal vibrato, but is usually thought of as a separate effect: tremolo. Musicians use vibrato for a variety of reasons: it helps to add warmth and expression to the tone; it helps to modify the radiation pattern of the sound produced, thereby projecting it better in a concert hall; it masks tuning inaccuracies, especially as used by string ensembles.

The two most important parameters of vibrato are the extent (or amplitude) and the rate (or frequency) of the modulator. The vibrato extent refers to the maximum deviation from the central pitch, whereas the vibrato rate refers to the number of times the pitch deviates above and below the central pitch in one second. Although theoretically, any values for extent and rate are allowed, when used in musical scenarios the amplitude is usually within a semitone or so above and below the pitch, and the rate is 5-10Hz. When the rate of vibrato becomes enters the realm of hearing (>20Hz), audible sidebands are produced. In our study, we focus on the vibrato rates typically found in musical performance. Vibrato is modeled by applying a low-frequency modulation signal to a carrier signal (which is essentially the melodic tone).

Our interest in vibrato relates to its use in operatic singing. This vocal style typically employs a high level of vibrato, which could have a negative effect on a listener's ability to accurately perceive the notes in a melody. This led us to the question: how much vibrato is too much? While initially this seems like a highly subjective question, and to some extent it is a matter of personal preference, we were interested in identifying quantifiable aspects of vibrato and designing an experiment that would treat these aspects as variables. Furthermore, are there limits on the vibrato parameters, beyond which the intelligibility of a melody significantly decreases?

First, we hypothesized that the rate and extent of the vibrato, defined respectively as the frequency of the modulator and the amplitude of the modulator, have an effect on the intelligibility of melody. We hypothesized that melody would be most intelligible with a relatively high vibrato rate and low vibrato extent. This was based on the idea that a faster rate would have more of a centering effect on the pitch: where the boundary pitches (the extreme pitches in the vibrato) produced by the frequency modulation would be perceived as their mean, rather than as the boundary pitches themselves. Similarly, the further the boundary pitches are from the center frequency, the greater the set of pitches that can be heard between them as the center pitch. Based on this observation, we hypothesized that melodic intelligibility would decrease as vibrato extent increased.

2. RELATED WORKS

Studies on vibrato parameters generally fall into one of three categories. Some studies consider measurement of vibrato parameters for acoustic instruments, while others consider the parameters of operatic vibrato. A third group of studies experiment with synthesized vibrato. The advantage of synthesized vibrato is that we have a complete control over parameters in a wide and continuous range and therefore the data set can be generated as needed.

Regardless of the method, the aforementioned studies are primarily interested in the pitch perception of vibrato by human subjects. Studies on vibrato pitch perception dates back to 1938 with the work of Seashore [1]. He studied vibrato from both performance and psychological perspectives. Regarding perception of central frequency, Seashore came to the conclusion that the pitch of a vibrato is perceived as the average of the two extremes. Brown [2] has summarized the results of previous studies of pitch perception of frequency modulated sounds. According to him, the perceived pitch is either a flat [3-4], a sharp [5], a sharp or flat depending on vibrato extend [6] and the mean pitch [7–9].

A study from 2002 found that the extent of the vibrato had significantly less effect on the perceived pitch than that of the rate [10]. This study showed that matching two vibrato samples was easiest when the samples had similar vibrato rates, while vibrato extent showed little matching correlation. Additionally, the same study notes that when using vibrato, not only is low-frequency FM created, but relatively small amounts of amplitude modulation is created as well. This occurs due to the instrument's natural resonances amplifying and attenuating various harmonics of the played note. Though this small amount of amplitude modulation was ultimately determined to have a greater impact on the tonal sound than the FM vibrato, in the present study we chose to restrict our samples to pure sinusoids with just vibrato. This was decided based on AM's high variability between different notes and instruments and the convention that vibrato is solely modelled as frequency modulation.

Knowledge of the vibrato parameters can help in synthesizing better vibrato sounds. Moreover, currently there is a lack of standard parameters for operatic vibrato and the vibrato is performed on a subjective basis depending highly on the vocalist. For instance, Seashore (1938) has defined a good vibrato as a "pulsation of pitch usually accompanied with synchronous pulsations of loudness and timbre, of such extent and rate as to give a pleasing flexibility, tenderness and richness to the tone". He also states that, "A good singer will have an average vibrato of from five to eight regular pulsations per second" [11]. It is not surprising that the vibrato is highly varying between different vocalists and from time to time. A study by Ferrante [12] on soprano voice recordings of the same tone sung by 75 artists over the last century shows a clear decrease of the mean vibrato rate by 1.86 +/-0.3 Hz/century and an increase of vibrato extent by 56.46 +/-0.3 cent/century.

In order to contribute to knowledge about successful use of vibrato, which can be used as a guide for operatic singers as well as for the audio synthesis, in this paper, an objective assessment of vibrato parameters has been performed across a number of subjects through a listening test. Results have been analyzed for their statistical significance.

3. EXPERIMENT DESIGN

The data acquisition system for this experiment is designed in *MaxMSP*. The interface is simple and user friendly. The independent variables for this experiment are: rate of vibrato, extent of vibrato and the central frequency around which the vibrato happens.

The range and values chosen for the independent variables and the justification for the same are as follows:

1. Rate - [2, 6,10,14] Hz - Typical vibrato rate for vocal and instrumental performances lie in the range of 5-10 Hz. This choice of rate parameters therefore covers the normal range as well as includes the extreme limits. A vibrato rate close to 20Hz will introduce audible sidebands due to frequency modulation type effects and is therefore avoided.
2. Extent - [30, 60, 90, 120] - Once again the typical extent in a vibrato is usually within a semitone above and below the central pitch. 120 cents corresponds to a little above a semitone and therefore tests the extreme limits.
3. Central Frequencies - [C2, D3, E4, F#5, G#5] – This frequency range has been chosen so that it is slightly less than the entire range of a piano, a common range in music. This variable is added as an exploratory independent variable so as to distribute the pitches on the frequency slider in a manner that will reduce subject learnability of the test. The hypothesis does not concern itself with the deviation as affected by the central frequencies.

22 subjects were recruited for the study, all of them students enrolled in the Music Perception and Cognition course at Georgia Tech. The subjects come from diverse musical and technical backgrounds.

3.1 Design considerations

A sine wave oscillator is chosen as the modulator, as this is the simplest possible modulator wave shape. *MaxMSP* provides a built-in sine wave generator. It should be noted that an equal amount of deviation from the central frequency in each direction in cents does not correspond to equal deviation in Hertz. For example, if a 30 cents extent about a central frequency corresponds to 'a' Hz above and 'b' Hz below a central frequency, then a sine wave with an amplitude of $\frac{(a+b)}{2}$, DC shifted by $\frac{(a-b)}{2}$ will have its positive peak at 'a' and negative peak at 'b'. This modified sine wave can then be used to modulate the frequency of another oscillator to produce the desired vibrato effect. Any possible influence on pitch perception due to timbral differences is avoided by choosing a sine wave as the carrier waveform.

Since the central frequencies cover a wide range, volume normalization is critical in order to avoid hearing discomfort. Fletcher Munson curve based volume normalization was performed on each one of the test samples. This is not entirely justified as the actual volume also depends on the output level of the sound card in the computer and may not represent the correct Fletcher Munson curve for the system level set by the user. Regardless, having some volume normalization is still far better than not having one.

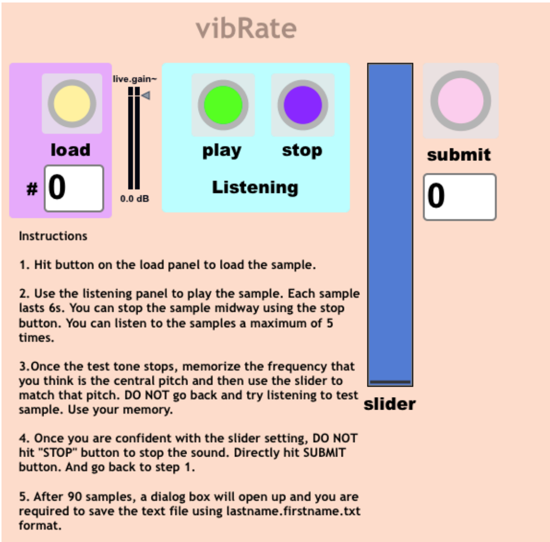

Figure 1. Application Interface

The application comes packaged with the test samples whose parameter values are pre-computed and stored in a text file. The order of the samples is randomized so that no two consecutive test samples have the same central frequency.

Figure 1 shows the application interface. A typical trial run would be as follows:

1. The subject uses the button in the left pink panel to load a test sample. This internally would set all the parameters (rate, extent, DC shift, volume) to the correct values.
2. The user gets to play a test sample a maximum of 3 times (each lasting 10s) using the play/stop buttons from the middle 'play panel'. The user is free to record his/her response whenever he/she wishes.
3. Once the stimulus stops, the objective is to memorize the frequency that the subject thinks is the central pitch and then use the slider to match that pitch. The subject is required not to go back and listen to the test sample but instead rely on his/her internal hearing and memory when making the decision.
4. Once the subject is confident, he/she will hit the 'submit' button on the right panel, which will record the frequency from the slider and writes into a text file.
5. Repeat 1-5, 80 times for different central frequency, rate and extent settings.

The first ten samples are pure sinusoids used for calibration purposes. The subject uses these samples to familiarize himself/herself with the experiment flow. Furthermore, this information is useful in some way to determine how accurate a subject's inherent pitch matching capabilities are. The average test time is approximately 20-25 minutes. The slider ranges are randomized for each trial run so that the correct pitch happens to be in different regions of the slider.

4. ANALYSIS

As a first step, we used ANOVA to test the significance of each independent variable. ANOVA assumes normal distribution of the dependent variable. **Figure 2** shows overall results, where deviations largely follow normal distribution, with a mean of 42.3844 cents, suggesting that people tend to report the perceived pitch slightly higher than the true pitch. This justifies the use of ANOVA the details of which are explained below.

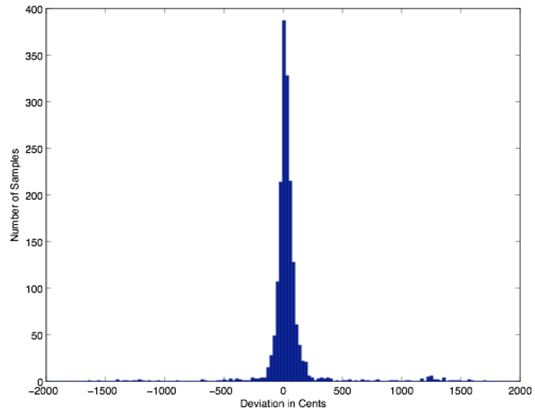

Figure 2. Histogram of deviation from central frequency

4.1 Analysis of Variance (ANOVA)

Analysis of Variance (ANOVA) is a method that allows us to compare the means of a continuous variable between a number of groups affected by discrete factors. It works by comparing between-group variances with within-group variances. The null hypothesis assumes that the continuous random variable is not affected by the discrete factors (all means are equal). By calculating a F-test statistic (p-value), ANOVA determines whether there are significant differences between the means of the different groups. A cut off point of <0.01 or <0.05 is used as a rule of thumb to determine whether or not the relationship is significant. Depending on the number of factors, there are one-way, two-way, and up to N-way ANOVA. N-way ANOVA is used to determine if the means in a set of data differ when grouped by multiple factors.

In our case, the continuous dependent value is the deviation of perceived central pitch from actual central pitch (in cents), while the discrete independent factors are rate, central frequency, and extent. We also did a test on different subjects. Thus a linear model of N-way ANOVA is employed to test the effect of each factor. We do not consider interactions between different factors (assuming they are independent of each other). The null hypotheses would be that the deviation has the same mean among different groups of one same factor. We use Matlab's '*anovan'* function to perform such analysis, and use a cut off point of less than 0.01. Since the F-test is a "global" test, we know that there is a significant difference but not where exactly the significance lies. We used Matlab's multiple comparison test, '*multcompare'* to select and compare one group with all the other groups on an interactive graph. A 95% confidence interval was used for all the comparison tests.

4.1.1 Rate

To compare the influences of different rates, the null hypothesis is that the means of errors are the same for all rates. The result of ANOVA shows $p = 0.3232 > 0.01$, which does not reject the null hypothesis.

4.1.2 Central frequency

To compare the influence of different central frequencies, the null hypothesis is that the means of errors are the same for all central frequencies. The result of ANOVA shows $p = 0.0002 < 0.01$, which rejects the null hypothesis.

ANOVA shows that deviations have a slight increasing trend as central frequency increases, with a sharp drop between 146.83 and 329.63 Hz. It also indicates that subjects are most accurate when central frequencies are around 329.63Hz to 739.99 Hz.

4.1.3 Extent

To compare the influence of different extents, the null hypothesis is that the means of errors are the same for all extents. The result of ANOVA shows $p = 0.0053 < 0.01$, which also rejects null hypothesis.

The analysis also reveals that the deviations have a slight decreasing trend, with a sharp rise between 90 and 120 cents. It also indicates that subjects are most accurate when extent is around 90 cents.

In conclusion, ANOVA shows us that central frequencies and extents have a significant influence on central pitch perception in a vibrato, while rates do not. The analysis also showed that different subjects report perceived pitch quite differently, and this can be attributed to multiple factors, like equipment, psychological effects, perceptual differences, etc.

4.2 Correlation Analysis

The results of the n-way ANOVA suggest that extent has a significant influence on the error in pitch perception of a vibrato. This motivates further inquiry into the strength and direction of this influence. Prior to the correlation analysis outlier pruning was performed. Any data point greater than a perfect fifth above or below the actual pitch was treated as an outlier and was not included in the correlation analysis. This hard threshold for pruning is justified because the maximum vibrato extent used in the experiment is only slightly greater than a semitone.

There are three independent variables under consideration (rate, extent, and central frequency). In order to isolate the influence of one of them, all others have to be kept constant. **Figure 3** shows extent vs. error plots for all combinations of central frequencies and vibrato rate (20 in total). Each one of the lines in the figure represents the linear best fit for the 4 data points, corresponding to the 4 mean errors for each one of the extent values. From **Figure 3** the overall trend is that when the extent increases the error increases as well. The correlation coefficient of each set of data points (20 sets) was computed and the average correlation coefficient was determined to be +0.5681. The positive correlation signifies that, when the extent increases, the error also increases and this supports the hypothesis, although the correlation is moderately weak.

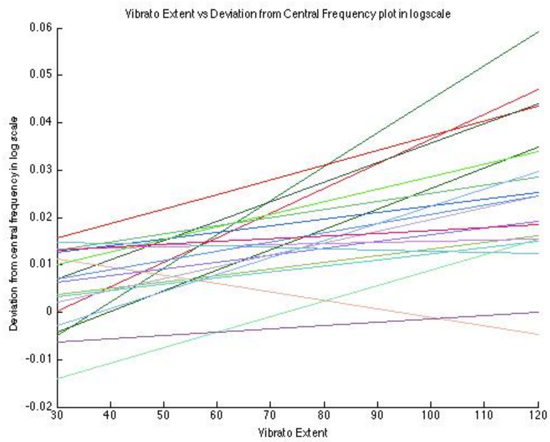

Figure 3. Extent vs. Error (log scale)

Even though n-way ANOVA suggested that vibrato rate does not have a significant effect on the error in pitch perception, it was still decided to investigate the correlation between rate and error. Once again, in order to investigate the effect of rate in isolation all other independent variables have to be held constant. **Figure 4** shows rate vs. error for all combinations of central frequencies and extents. Each line in the figure represents the linear best fit for the 4 data points, corresponding to the 4 mean errors for each one of the rate values. Unlike the extent vs. error plot, there is a lot more variation in the slope of the best-fit lines under different conditions. The correlation coefficient of each set of data points (20 sets) was computed and the average correlation coefficient was -0.1595. It is clear that the correlation is very small and therefore not very significant. An interesting observation is that, the negative sign of the correlation coefficient still suggests that as the rate increases, the error decreases which weakly supports the initial hypothesis. But no strong conclusions can be derived when the p-value is very high.

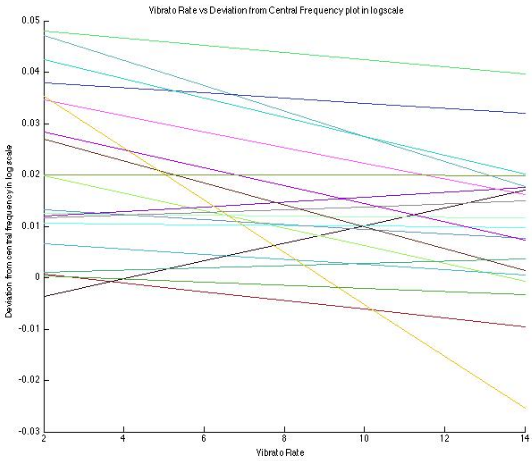

Figure 4. Rate vs. Error (log scale)

N-way ANOVA indicated that central frequency also has a significant effect on the error in perceived pitch.

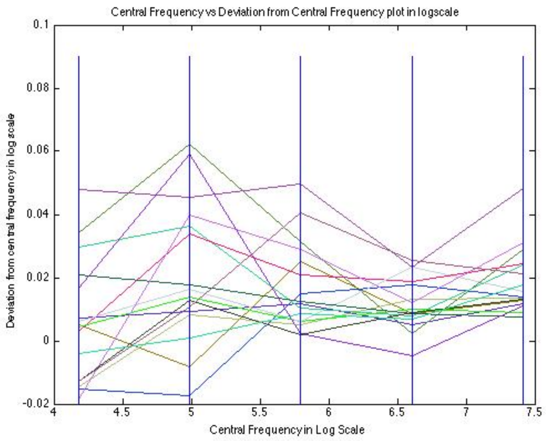

Figure 5. Central Frequency vs. Error (log scale)

The vertical blue lines in **Figure 5** correspond to the 5 central frequencies used in the experiment. Each of the piecewise linear curves (16 of them) correspond to a unique rate-extent combination. The y-axis denotes the deviation from central frequency on a log scale. It is clear from the figure that for a unique combination of rate and extent, as the central frequency is changed the error fluctuates quite significantly. Trying to do a linear best fit is not sensible in this case because of the high degree of fluctuation. There is no clear pattern to this fluctuation, although it can be observed the variance in the error is smaller for higher frequencies. This might have to do with the fact that the slider in the application allowed for better resolution at higher frequencies than at lower frequencies.

5. CONCLUSIONS

The effect of two vibrato parameters (rate and extent) on the perception of a central pitch has been studied. Our findings suggest that error of pitch perception increases as extent increases, while rate has no significant effect on pitch perception accuracy. This would suggest that operatic singers can successfully employ a wide variety of vibrato rates and still maintain melodic intelligibility, while the extent should be kept relatively small. This is in contrast with Järvaläinen's 2002 findings [10]. Greater resolution in our independent variables, that is, more combinations of rate and extent, could yield more statistically significant results. Future work could include a better data acquisition system in which the slider resolution is the same for all central frequencies.

6. REFERENCES

[1] C. E. Seashore: *Psychology of music*, Courier Dover Publications, 1938.

[2] J. C. Brown and K. V. Vaughn: "Pitch center of stringed instrument vibrato tones," *The Journal of the Acoustical Society of America*, vol. 100, no. 3, pp. 1728–1735, 1996.

[3] J. Tiffin: "Some aspects of the psychophysics of the vibrato," *Psychological Monographs: General and Applied*, vol. 41, no. 4, pp. 153–200, 1931.

[4] H. Fletcher and L. C. Sanders: "Quality of violin vibrato tones," *The Journal of the Acoustical Society of America*, vol. 41, no. 6, pp. 1534–1544, 1967.

[5] H. Fletcher, E. D. Blackham, and O. N. Geertsen: "Quality of Violin, Viola, 'Cello, and Bass‐Viol Tones. I," *The Journal of the Acoustical Society of America*, vol. 37, no. 5, pp. 851–863, 1965.

[6] K. Hirose: "An experimental study on the principal pitch in the vibrato," *Japanese Journal of Psychology*, vol. 9, pp. 793–845, 1934.

[7] S. Iwamiya, K. Kosugi and O. Kitamura: "Perceived principal pitch of vibrato tones," *Journal of the Acoustical Society of Japan (E)*, vol. 4, no. 2, pp. 73–82, 1983.

[8] J. I. Shonle and K. E. Horan: "The pitch of vibrato tones," *The Journal of the Acoustical Society of America*, vol. 67, no. 1, pp. 246–252, 1980.

[9] J. Sundberg: "Effects of the vibrato and the 'singing formant'on pitch," *Journal of Research in Singing*, vol. 5, no. 2, pp. 5–17, 1978.

[10] H. Järveläinen: "Perception-based control of vibrato parameters in string instrument synthesis." *Proc. International Computer Music Conference, Sweden*, 2002.

[11] M. Bunch: *Dynamics of the singing voice*, Springer, 1982.

[12] I. Ferrante: "Vibrato rate and extent in soprano voice: A survey on one century of singing," *The Journal of the Acoustical Society of America*, vol. 130, no. 3, pp. 1683–1688, 2011.