Perception-based control of vibrato parameters in string instrument synthesis

Hanna Järveläinen

DEI – University of Padova, Italy

Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, Finland

email: hanna.jarvelainen@hut.fi

Abstract

Perceptual knowledge provides a tool for making sound synthesis more efficient. Also for the parametric representations of sound, it is necessary to understand the perception of individual features. This paper reports four listening experiments exploring the perception of vibrato, aiming at guidelines for the control of vibrato parameters in string instrument synthesis. The results suggest that accurate control of the vibrato rate is much more important than control of vibrato extent. JND's for vibrato rate were found to be around 6 %, while for the extent the perceptual tolerances were much wider. Additional experiments on the perception of pitch and musical consonance of vibrato tones did not support the importance of vibrato extent either.

1 Introduction

With the development of the 'Transmit less, receive more' philosophy, communication technologies face new challenges. For instance, the structured methods for representing synthetic audio (ISO/IEC 1999) allow transmitting high-quality content in a low-bitrate channel. However, if the transmission consists of parametric models and control data instead of the sampled signal waveform, the conventional data reduction methods (ISO/IEC 1993), (Vercoe, Gardner, and Scheirer 1998), (Scheirer and Yang 2000) become useless. For more efficiency in transmission and synthesis, we should be able to simplify the models and reduce the amount of control data.

Studying perception offers tools for both. Knowing which are the perceptually prominent features of the sound, we can skip synthesizing the others that would remain inaudible. Knowledge of the perceptual effects of changes in the control parameters can help us design coding-schemes for them.

The current study on the perception of vibrato is connected to others exploring the perception of different features of string instrument sounds (Järveläinen et al. 2001), (Järveläinen and Tolonen 2001). It aims at perceptual-based rules for synthesizing high-quality vibrato sounds and adjusting the synthesis parameters.

1.1 Perception of vibrato

Vibrato is created by the motion of the player's finger back and forth on the finger board. The variable string length causes a constant frequency modulation. Vibrato is used because it gives the sound more depth and sustain. Another objective is to make the vibrato sounds stand out from the rest of the sound space.

Figure 1 presents the frequency modulation patterns analyzed from recorded classical guitar tones played by a professional guitar player (Erkut et al. 2000). The pitch of the tones is estimated by the autocorrelation method. It is seen from the analysis of these and two other tones that the modulation rate is typically around 5 Hz, while the total variation of pitch is between 0.7 Hz . . . 3 Hz.

Although the player creates mainly frequency modulation, it results in changes in amplitude that are crucial for the perception of vibrato (Mellody and Wakefield 2000). The moving harmonics are boosted and depressed according to the resonances of the instrument body. This poses problems for the systematic study of the perception of vibrato, since the body resonance characteristics vary from instrument to instrument, and the amplitude modulation changes for each note as a function of the depth of the frequency modulation. Mellody and Wakefield (2000) found that even though triggered by the sinusoidal frequency modulation, the amplitude changes were more complex in nature and the amplitude envelopes of individual harmonics had little or no correlation between each other. They also found that the absence of frequency modulation had little effect on the perceptual quality of synthesized vibrato sounds, whereas the absence of the amplitude modulation caused significant effects in sound quality.

Fig. 2 presents the pitch variation as well as the amplitude envelopes of the three lowest harmonics of a synthetic guitar sound, where the amplitude modulations were produced using a filter representing the resonances of the instrument body (Penttinen et al. 2001). The amplitude modulation varies significantly between harmonics, being strongest for the second harmonic which exhibits coherent modulations with the pitch contour.
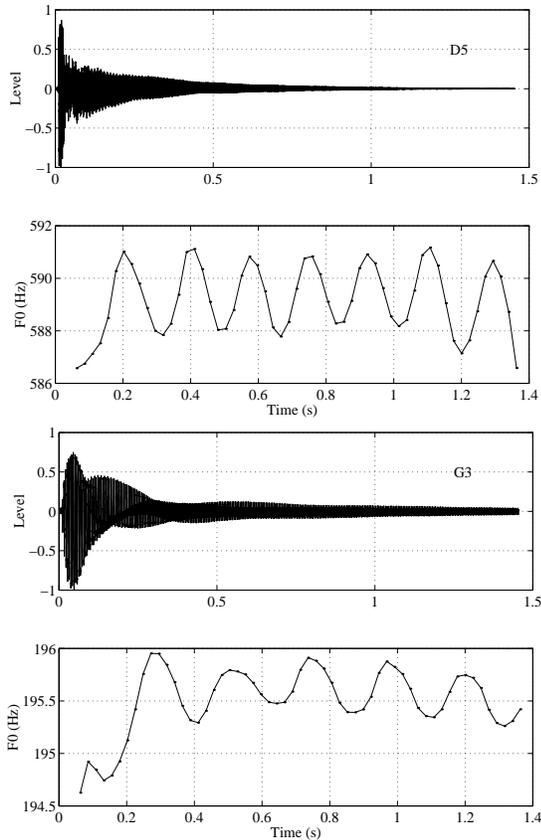
Figure 1: Waveform of a single vibrato tone played on the classicalguitar (top), and a pitch estimate showing a typical frequency modulation pattern (bottom) – (a) D5, (b) G3.
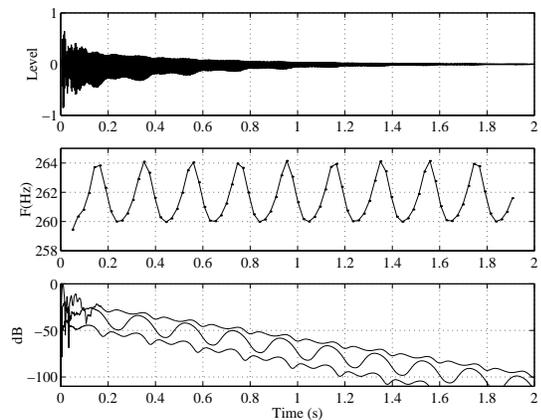

Figure 2: Top to bottom: Signal waveform, pitch estimate, and amplitude envelopes of the lowest three harmonics of a synthesized guitar vibrato sound with center frequency 262 Hz.

The origin and nature of vibrato in instrument sounds and especially voice is well-known (Dejonckere et al. 1995), (Prame

1994), (Prame 1997), (Gleiser and Friberg 2000). Research into the perception of vibrato concerns mainly emotional expression (Jansens et al. 1997) or the pitch center of vibrato tones, which is still subject to ongoing discussion (d'Alessandro and Castellengo 1994), (Brown and Vaughn 1996), (Yoo et al. 1998).

Because of the complex relations of the frequency and amplitude modulation in vibrato, it is obvious that the theoretical knowledge on the perception and detection of FM or AM (Plack and Carlyon 1997), (Hartmann 1997) is hard to apply to musical sounds. However, from the synthesis viewpoint the main interest is exactly this – to gain quantitative knowledge of our ability to discriminate different vibrato patterns, and to find the criteria for high-quality synthesis of vibrato. Two questions are addressed in this paper:

- How accurately should the original vibrato pattern be captured so that the difference would remain inaudible in synthesized sounds?

- Can the presence of vibrato "mask" the effects of some other features, such as consonance or intonation problems caused by inharmonicity?

These questions were studied in four listening experiments. The first one was a similarity rating test for different vibrato patterns, which were generated by varying both the vibrato rate and extent of synthetic guitar sounds. The results are reported in section 2.2. The second test (section 2.3) explored more thoroughly the thresholds for detecting changes in vibrato rate, which was found to be the primary factor of perceived similarity. The other two experiments studied the importance of vibrato extent and are discussed in sections 2.4 and 2.5. The third test concerned the accuracy of pitch perception of vibrato sounds, and the last one concerned the effect of vibrato on the perceived consonance of a musical interval.

2 Listening tests

2.1 Synthesis of test tones

Even though the amplitude modulations are considered perceptually more prominent than the frequency modulations in vibrato sounds, in many synthesis techniques all vibrato effects are controlled by frequency modulation. This is the case for instance in physical modeling (Jaffe and Smith 1983), (Karjalainen, Välimäki, and Janosy 1993), where the controlled variations of pitch automatically cause the AM effect which is due to the resonances of the instrument body. This is the reason for using the control parameters for frequency modulation as independent variables throughout the study.

The test tones were created by additive synthesis, generating all harmonics up to the Nyquist frequency at 11.025 kHz. Realistic decay characteristics and initial amplitudes,

presented in (Välimäki and Tolonen 1998) and (Erkut, Välimäki, Karjalainen, and Laurson 2000), were used along with a guitar body filter, obtained by modeling the body response of an acoustic guitar (Penttinen et al. 2001). The frequency response of the body filter is presented in Fig. 3.

Since sinusoidal modulation is close enough to the real vibrato pattern (see Fig. 1 and (Erkut et al. 2000)), the frequency modulation of the tones was controlled by two parameters – vibrato rate f_{mod} and vibrato extent Δf. The harmonic relations of the partials were maintained so that the extent of the modulation was greater for the higher partials. Reference values for the parameters were approximated from the recorded tones, and they represent thus an individual player's typical vibrato, not an average of several players. Fig. 4 presents the synthesized reference tones corresponding to the real tones in Fig. 1. The duration of each tone is 2.0 s.
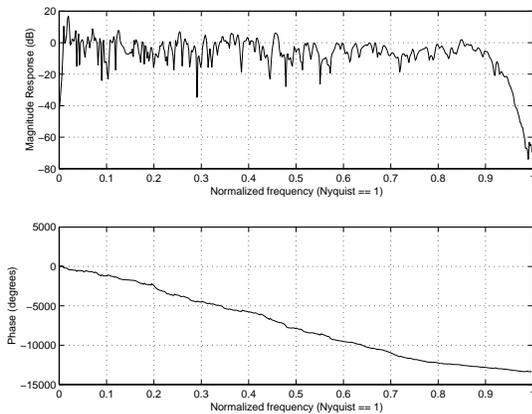

Figure 3: Frequency response of the guitar body filter used to generate the AM effect for the test tones.

2.2 Similarity rating experiment

The first experiment studied the perceived similarity of two vibrato patterns as a function of vibrato rate and vibrato extent. The objective was to find out, how much inaccurate control of the vibrato parameters degrades the perceived naturalness of vibrato, and which of the synthesis parameters is more crucial in this sense. Four tones were studied, spanning a major part of the pitch range of the guitar: D5, A4, C4, and G3.

Five subjects with normal hearing and former experience in psychoacoustic testing participated in the experiment. The task was to compare the standard tone whose vibrato parameters were fixed close to the measured references to a test tone whose parameters were varied. A variation range was determined on both sides of the reference values, such that the differences were clearly audible for the tones with the greatest deviation from the reference parameter values. The parameter values are summarized in Table 1.
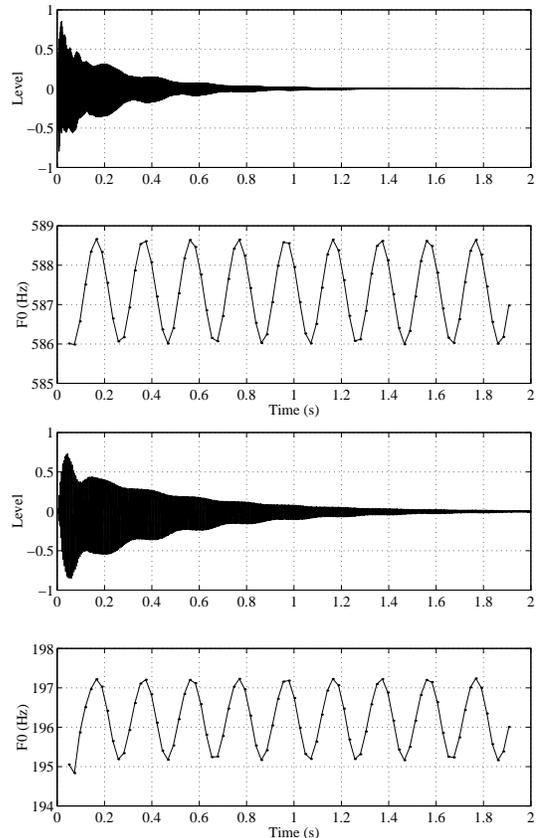

Figure 4: Waveform of a single synthetic vibrato tone (top), and a pitch estimate showing its frequency modulation pattern (bottom) – (a) D5, (b) G3.

All combinations of the parameter values were tested, including the standard-standard pairs, resulting in 25 tone pairs to judge for each f_0. The subjects were asked, how similar the vibrato pattern of the test tone sounded compared to the standard tone. The tones were rated on a scale from 0 for 'Very different' to 10 for 'Very similar'. Before the test there was a rehearse cycle presenting part of the material, including the most similar and most different cases, to give the subjects a chance to decide on their criteria and preferences.

f_0	Δf / Hz	f_{mod} / Hz
D5 (587 Hz)	[0.5 ... **1.5** ... 2.5]	[3 4 **5** 6 7]
A4 (440 Hz)	[0.3 ... **1.0** ... 2.0]	[3 4 **5** 6 7]
C4 (262 Hz)	[0.3 ... **1.0** ... 2.0]	[3 4 **5** 6 7]
G3 (196 Hz)	[0.1 ... **0.5** ... 1.3]	[1.5 3 **4.5** 6 7.5]

Table 1: Modulation parameter values in the similarity rating experiment. Reference values are marked bold.

All similarity rating results are presented in Fig. 8 for D5, A4, C4, and G3 from top to bottom. The left column presents

the similarity ratings as a function of vibrato extent Δf with vibrato rate as parameter. The five values of f_{mod}, given in Table 1, are presented as square, circle, solid line (reference value) , star, and triangle. The right column presents the ratings as a function of vibrato rate f_{mod} with vibrato extent Δf as parameter. The symbols for Δf are the same as for f_{mod} in the left column.

The figures show that vibrato rate is more crucial for perceived similarity than vibrato extent. Whenever f_{mod} is similar to that of the reference tone, both tones are perceived rather similar regardless of vibrato extent. But when the modulation is either too slow or too fast, the similarity ratings get significantly worse even if the vibrato extent is identical to the reference. The trend is seen clearly in both left and right columns. In the left column figures, the solid line shows that the best ratings were obtained using the reference value of f_{mod}. In the right column, all ratings follow the same trend as a function of f_{mod} regardless of Δf.

The results for G3 are unclear for the lowest two values of Δf, shown by squares and circles in the right column, which show better perceived similarity than the reference value itself, shown by solid line. These can be explained by badly chosen reference values. The vibrato in the recorded reference tone was very weak, almost inaudible, so that the test tones with even less vibrato fall in the same category.

The two-way analysis of variance (ANOVA) was performed on the data to find out significant differences and possible interactions in the ratings. The results were significant in all cases for both vibrato rate and vibrato extent, as well as their interaction, i.e., it is likely that both parameters individually as well as their combinations have an effect on the results. The note D5 was exceptional in this respect, because a significant interaction was observed only for vibrato rate ($P < 0.01$ for f_{mod}; P=0.0583 for Δf and P=0.2483 for interaction). The interaction was also insignificant for G3 (P=0.0579).

The results for the tones with either f_{mod} or Δf fixed to the reference were analyzed further to find, how much deviation is allowed for both parameters individually until the similarity ratings fall by 25 %. When Δf was fixed to the reference value, the range within 25 % was between 70 % and 115 % of the reference for f_{mod}. When f_{mod} was fixed to the reference and the -25 % range was estimated for Δf, it was typically between 45 % and 167 % of the reference. However, an estimate could only be obtained for some of the tones; for D5 the ratings were always better than 75 % of the reference, and for C4 and G3 the -25 % point was never reached at least on one side of the reference. This is a further indication of the perceptually wide tolerances for vibrato extent.

2.3 JND's for vibrato rate

To gain more specific knowledge about the perception of changes in the vibrato rate, the just noticeable differences were measured for four fundamental frequencies. The vibrato rate of the standard tone was fixed to 5.0 Hz in each case, and for the stimuli the rate was varied in steps of 0.2 Hz on both sides of this reference. The vibrato depth was fixed to correspond to the measured reference value for each note, being 1.5 Hz for D5 and A4, 1.0 Hz for C4, and 0.8 Hz for G3. The value for G3 was slightly increased from the measured reference ($\Delta f = 0.5$ Hz), which produced only a weak vibrato in the first experiment.

Six subjects participated in the experiment. It was made using the same-different procedure, i.e. the task was to detect a difference between two sounds in a trial. Since 50 % of the trials were standard-standard pairs, a detected difference was either a hit or a false alarm, depending on whether a stimulus was present or not. Each stimulus was presented four times. A measure of sensitivity d' was estimated from the data and a threshold was computed, expressed as the difference in vibrato rate required for 75 % area under the ROC curve (Green and Swets 1988).

The results are shown in Fig. 5. The mean upper and lower thresholds are presented relative to the reference vibrato rate 5.0 Hz that corresponds to 1.0 on the vertical axis. The whiskers show one standard deviation upwards and downwards.

The upper and lower thresholds are symmetrical about the reference and rather independent of fundamental frequency. However, for C4, the thresholds are very close to the reference. This may be due to nonmonotonous data: since the number of repetitions was low, it is probable that some of the subjects have correctly detected a very small difference once or twice even though they were not able to detect a greater difference. But because the lowering of thresholds is symmetrical, it is possible that some detail in the test sounds makes it easier to distinguish between the standard and the stimuli for C4. However, none of the subjects reported anything exceptional.

The significance of the differences was again evaluated by ANOVA, even though equal variance could hardly be assumed. The differences between the upper thresholds were insignificant (P = 0.16), while for the lower thresholds a true difference in means seemed more probable. However, when C4 was excluded from the analysis, the differences between between G3, A4, and D5 were clearly insignificant for both thresholds. a mean over G3, A4, and D5 gave 6.5 % for the lower and 6.1 % for the upper thresholds. When the problematic C4 was included, the mean became 5.3 % for both sides.

2.4 Pitch matching experiment

The previous experiments showed that vibrato rate dominates the perceived similarity of vibrato patterns. But vibrato might have indirect effects which depend on vibrato extent. Intuitively the effects of vibrato on pitch perception and mu-
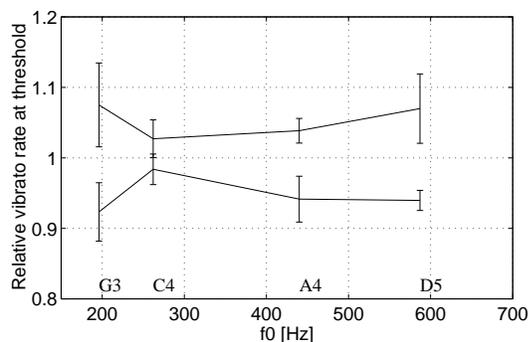
Figure 5: Upper and lower JND's for detecting a difference in vibrato rate for G3, C4, A4, and D5. The results are given relative to the reference vibrato rate of 5 Hz corresponding to 1 in the figure.

sical consonance, if any, should also depend on the extent of the pitch variation and not only rate.

It is an interesting question, whether the presence of vibrato affects our ability to judge the actual pitch of the sound. If the accuracy of pitch perception decreases as a result of vibrato, many features of string instrument sounds with minor effects on pitch could be left unimplemented.

The pitch center of vibrato sounds has been subject to many studies previously. A recent study along with a summary on the previous ones is presented by Brown and Vaughn (1996). A common result is that the perceived pitch of vibrato sounds is the mean over the vibrato cycle, although some performers are convinced that either the sharp or the flat extreme of the vibrato cycle is perceived as the overall pitch. Yoo et al. (1998) observed that more time is required for determining the pitch relationship of successive sounds when vibrato is present. However, many of the previous studies were made using tones that are unlike any real instrument. Brown and Vaughn used recorded viola tones played by a virtuoso violist, but they had no way of adjusting the rate or depth of the vibrato pattern.

In the current study, the perceived pitch of vibrato sounds with variable vibrato rate and depth was measured by a listening experiment. The rate was either 5 Hz, corresponding to a typical vibrato pattern, or 8 Hz, which is clearly faster. The vibrato depth was varied in five linear steps between 0 Hz and 8 Hz. A4 (440 Hz) was chosen as the fundamental frequency. The vibrato sounds were synthesized both with and without the body filter that produces the amplitude modulation. The combinations of these parameters resulted in 20 different vibrato sounds to judge.

The task was to adjust the pitch of a pure tone until it matched the pitch of the vibrato sound. Six listeners participated in the experiment. They were allowed to switch freely between the vibrato tone and the adjustable pure tone, until they were satisfied with the pitches. The pure tone could be adjusted in steps of 0.5 Hz.

The experiment revealed no effect of either vibrato rate or vibrato depth on pitch perception. Furthermore, the synthesis method (with or without body filtering) made no difference. For both synthesis methods and vibrato rates, the ANOVA was insignificant for Δf. The results from different synthesis methods and modulation rates were tested against each other, but no significant differences were found. Furthermore, the pitch judgments for non-vibrato tones ($\Delta f = 0$ Hz) did not differ from the other tones significantly.

The data was collapsed over vibrato rate and synthesis method. The mean judgments for all Δf were practically equal. However, for some reason they were more than 1 Hz flat; the nominal fundamental frequency of the test tones was 440 Hz, which was verified by estimating the pitch of the test tones by the autocorrelation method. It was found that two of the subjects had judged all pitches clearly flat. When their judgments were removed from the total results, the mean judgments were closer to 440 Hz in all cases. The resulting box plot is seen in Fig. 6.

A conclusion of the pitch matching test is that the presence of vibrato was not found to interfere with pitch perception. The perceived pitch did not vary with increasing vibrato depth, which suggests that the pitch of vibrato sounds corresponds to the mean accross the vibrato cycle. The variability of the judgments was not significantly greater for vibrato sounds than non-vibrato sounds. This indicates that the accuracy of the pitch judgments was not impaired by vibrato. The results are consistent with the previous studies of Brown and Vaughn (1996) and d'Alessandro and Castellengo (1994). However, the fundamental frequency was exactly same for all test tones. It would be worthwhile to rerun the test with fundamental frequency randomized over a small range. This would prevent the subjects from memorizing the actual f_0, which might affect the results.
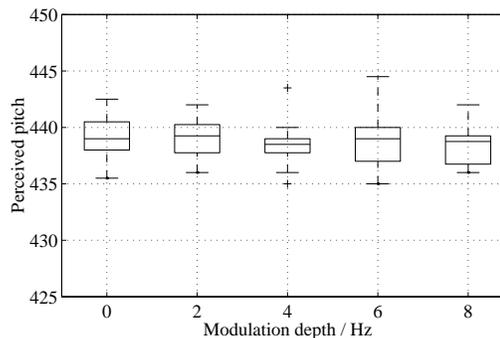

Figure 6: Results of the pitch matching test.

2.5 Vibrato and musical consonance

Another effect, which could reveal the importance of Δf after all, is the perception of musical consonance. This was studied by an experiment, where the pitch of the lower tone D♭4 and the vibrato depth of the higher tone F4 of a major third were adjusted. There was no vibrato in the lower tone. The vibrato rate of the higher tone was fixed to 5.0 Hz, while the vibrato depth was either 0, 1, or 2 Hz. The pitch of the lower tone was varied in 8-cent steps within ± 24 cents of equal temperament tuning. The subjects graded the consonance of each condition on a scale from 0 for 'Very dissonant' to 10 for 'Very consonant'. The measurements were repeated three times.

The mean consonance ratings are shown in Fig. 7. As was expected, the best general grades were given when the lower tone was tuned 16 cents sharp, because it is closest to the pure major third whose frequency ratio is 5:4. However, the presence of vibrato caused very little variation in the results. The intervals without vibrato (marked with a star) were judged generally as consonant as those with $\Delta f = 1$ Hz (circle) and $\Delta f = 2$ Hz (square). The only exceptions are made by the pure major third (+16 cents), where the interval without vibrato was judged more consonant than the others, and the equal major third (0 cents), which was judged slightly more consonant when vibrato was present. All other differences were insignificant.
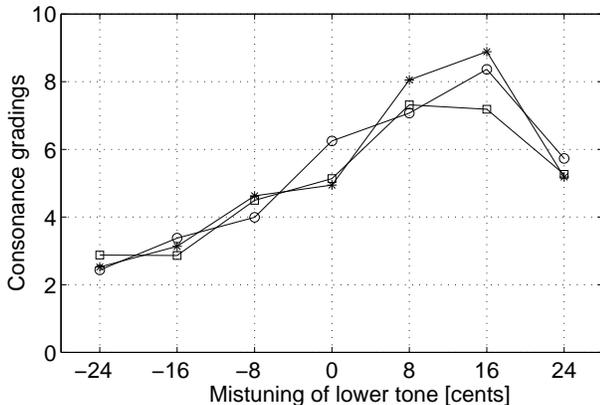

Figure 7: Results of the consonance rating task. $\star$ for no vibrato, o for $\Delta f = 1$ Hz, $\diamond$ for $\Delta f = 2$ Hz

extent had only a minor effect on the results. However, the joint effect of both parameters can not be totally neglected.

The thresholds for detecting changes in vibrato rate were measured more carefully in the second experiment. This was done in order to gain practical quantitative information for synthesis applications. Deviations smaller than about 6 % at both sides of the reference value were generally inaudible in this experiment, even though the thresholds for one of the tested pitches, C4, were lower. Otherwise, no systematic effect of fundamental frequency was detected.

The other two experiments studied effects were vibrato extent might have some importance. If the accuracy of pitch perception or the perceived consonance of an interval were affected by vibrato extent, it would indicate that accuracy is needed in the control of vibrato extent as well as rate. However, the results showed no significant effects of vibrato on either pitch or consonance.

The main implication on digital sound synthesis is that vibrato in string instruments can essentially be controlled through the rate parameter. However, this should be captured quite accurately, with less than 6 % deviation from the target value. Another point is that vibrato does not interfere with the perception of pitch or musical consonance. Thus the pitch or timbre effects of inharmonicity, for instance, should not be ignored in the presence of vibrato.

The perceptual guidelines for vibrato as well as other features of musical sounds share the practical aim of making sound synthesis more efficient. Even though it seems an impossible task to model the perception of musical sounds completely, these kind of perceptual models are likely to find many applications in future. With the development of parametric coding schemes and object-based representations of sound, the perception-based control of individual features will become even more attractive.

3 Conclusions

Different aspects of vibrato perception were studied in four listening experiments. The first experiment concerned the similarity of vibrato patterns with variable vibrato rate and extent. It was found that accurate control of vibrato rate is more crucial to the quality of synthetic vibrato, while the
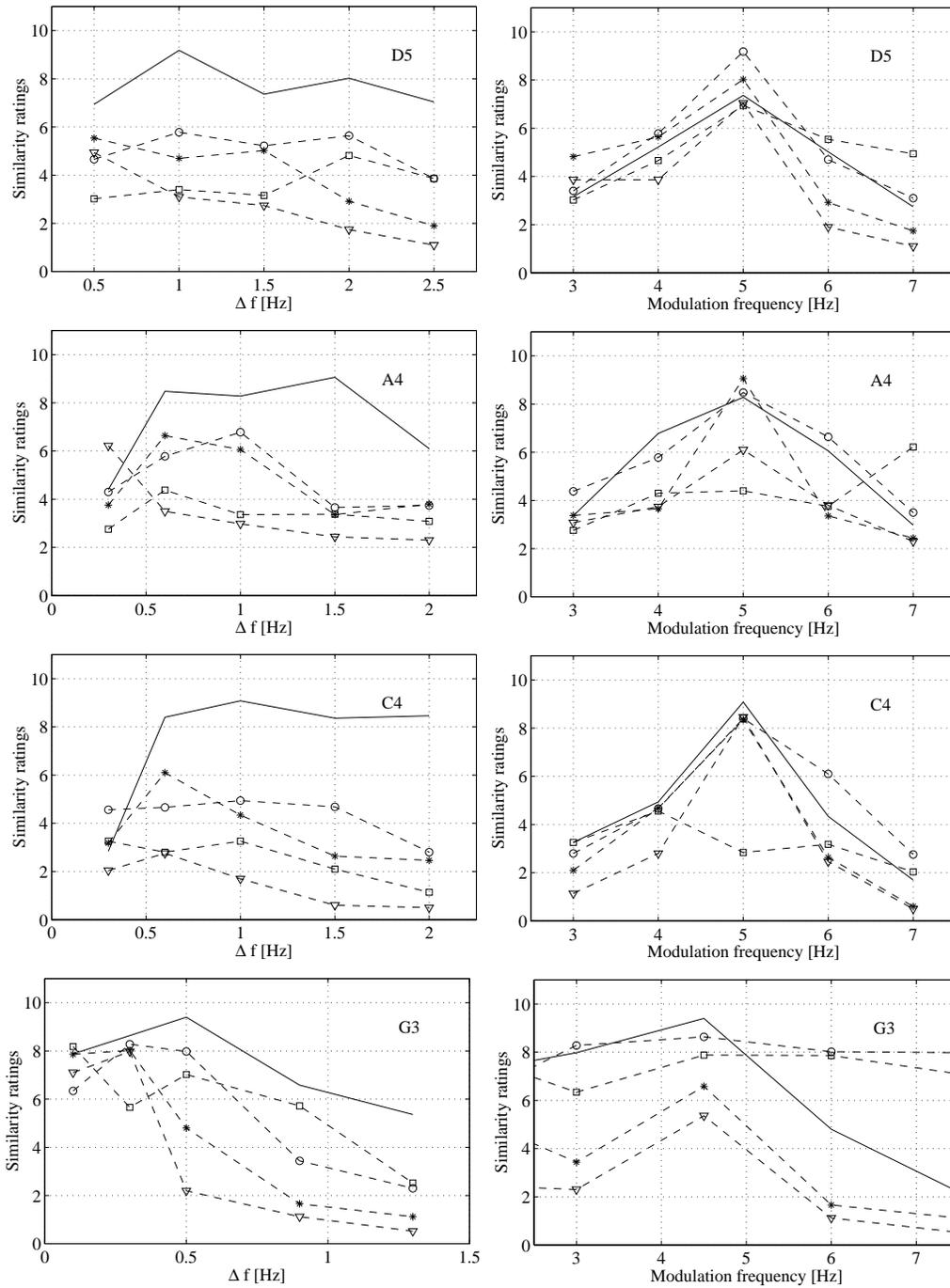
Figure 8: Top to bottom: Average results of the similarity rating test for D5, A4, C4, and G3. Left column: Similarity ratings as a function of vibrato extent Δf with vibrato rate as parameter. The five values of f_{mod}, given in Table 1, are presented as square, circle, solid line (reference value), star, and triangle. Right column: Similarity ratings as a function of vibrato rate f_{mod} with vibrato extent Δf as parameter. Symbols for Δf values same as in the left column.

Acknowledgments

References

Brown, J. and K. Vaughn (1996). Pitch center of stringed instrument vibrato tones. *J. Acoust. Soc. Am. 100*(3), 1728–1735.

d'Alessandro, C. and M. Castellengo (1994). The pitch of short-duration vibrato tones. *J. Acoust. Soc. Am. 95*(3), 1617–1630.

Dejonckere, H., M. Hirano, and J. Sundberg (Eds.) (1995). *Vibrato*. London: Singular Publishing Group.

Erkut, C., V. Välimäki, M. Karjalainen, and M. Laurson (2000). Extraction of physical and expressive parameters fot model-based sound synthesis of the classical guitar. In *Proc. 108th AES Convention*. Preprint 5114.

Gleiser, J. and A. Friberg (2000). Vibrato rate and extent in violin performance. In *Proc. 6th Int. Conf. Music Perception and Cognition*, Keele University, United Kingdom.

Green, D. M. and J. A. Swets (1988). *Signal Detection Theory and Psychophysics*. Los Altos, California: Peninsula Publishing.

Hartmann, W. (1997). *Signals, Sound, and Sensation*. New York: AIP Press.

ISO/IEC (1993). ISO/IEC IS 11172-3 Information Technology – Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s – Part 3: Audio.

ISO/IEC (1999). ISO/IEC IS 14496-3 Information Technology – Coding of Audiovisual Objects, Part 3: Audio.

Jaffe, D. A. and J. O. Smith (1983). Extensions of the Karplus-Strong plucked-string algorithm. *Computer Music J. 7*(2), 56–69.

Jansens, S., G. Bloothooft, and G. de Krom (1997). Perception and acoustics of emotions in singing. In *Proc. 5th Eurospeech*, Volume 4, Rhodes, Greece, pp. 2155–2158.

Järveläinen, H. and T. Tolonen (2001). Perceptual tolerances for decaying parameters in plucked string synthesis. *J. Audio Eng. Soc.*. Accepted for publication in the Journal of the Audio Engineering Society.

Järveläinen, H., V. Välimäki, and M. Karjalainen (2001). Audibility of the timbral effects of inharmonicity in stringed instrument tones. *Acoustics Research Letters Online 2*(3), 79–84. http://ojps.aip.org/arlo/.

Karjalainen, M., V. Välimäki, and Z. Janosy (1993). Towards high-quality sound synthesis of the guitar and string instruments. In *Proc. Int. Computer Music Conf.*, pp. 56–63.

Mellody, M. and G. Wakefield (2000). The time-frequency characteristic of violin vibrato: modal distribution analysis and synthesis. *J. Acoust Soc. Am. 107*, 598–611.

Penttinen, H., M. Karjalainen, T. Paatero, and H. Järveläinen (2001). New techniques to model reverberant instrument body responses. In *Proc. Int. Computer Music Conf.*, Havana, Cuba.

Plack, C. and R. Carlyon (1997). The detection of differences in the depth of frequency modulation. *J. Acoust. Soc. Am. 96*(1), 115–125.

Prame, E. (1994). Measurements of the vibrato rate of ten singers. *J. Acoust. Soc. Am. 96*(4), 1979–1984.

Prame, E. (1997). Vibrato extent and intonation in professional western lyric singing. *J. Acoust. Soc. Am. 102*(1), 616–621.

Scheirer, E. and J.-W. Yang (2000). Synthetic and SNHC audio in MPEG-4. *Signal processing: Image communication 15*, 445–461.

Välimäki, V. and T. Tolonen (1998). Development and calibration of a guitar synthesizer. *J. Audio Eng. Soc. 46*(9), 766–778.

Vercoe, B., W. G. Gardner, and E. D. Scheirer (1998). Structured audio: Creation, transmission, and rendering of parametric sound representations. *Proc. IEEE 86*(5), 922–940.

Yoo, L., S. Moore, D. Sullivan, and I. Fujinaga (1998, August). The effect of vibrato on response time in determining the pitch relationship of violin tones. In *Proc. Int. Conf. Music Perception and Cognition*, Seoul, South Korea.