Audibility of Initial Pitch Glides in String Instrument Sounds

Hanna Järveläinen, Vesa Välimäki

Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology
email: hanna.jarvelainen@hut.fi

Abstract

Listening experiments were made to measure the detection thresholds for initial pitch glides in string instrument sounds, where a rapid decline of pitch is caused by tension modulation during the attack. Realistic sounding synthetic tones were generated by additive synthesis. The frequency decay of the glide was defined through the overall decay rate of amplitude, simulating the behavior of real instruments. It was found that on the ERB frequency scale, the thresholds remained roughly constant at approximately 0.1 ERB with varying fundamental frequency. Thus, any pitch glide weaker than the given threshold remains inaudible for most listeners and could be left unimplemented in digital sound synthesis.

1 Introduction

High-quality sound synthesis is possible with the modern synthesis methods, such as physical modeling (Smith 1998) and sinusoidal modeling (Serra and Smith 1990). However, implementing all details of the sound is computationally costly. It would be desirable to leave such features, whose effects are not perceived by the listener, unimplemented.

A rapid descent of pitch during the attack is characteristic to many plucked and struck string instruments in forte playing. It can be detected for instance in the clavichord (Välimäki et al. 2000), the guitar (Tolonen et al. 2000), and the kantele – a traditional Finnish string instrument (Välimäki et al. 1999). The primary cause of the pitch descent is the varying string tension as a consequence of finite string displacement after plucking or striking the string (Legge and Fletcher 1984). In the clavichord, the effect is boosted by the mechanical aftertouch. The string tension can be directly controlled by the player through key pressure.

Fig. 1 shows a fundamental frequency estimate obtained from a recorded electric guitar tone by the autocorrelation method (Tolonen et al. 2000). The f_0 estimate decreases exponentially with time from 499 to 496 Hz, giving a glide extent of approximately 3 Hz.

The tension modulation can be implemented in physical modeling, for instance, by a special filter structure with signal-dependent fractional delay elements (Tolonen et al. 2000),

(Välimäki et al. 1998), but ignoring it would bring remarkable computational savings.

The detection and discrimination of frequency glides has been previously studied from a more theoretical viewpoint. Still the underlying mechanism remains unclear. It was suggested by Madden and Fire (1997) that the detection is based on changes in the low-frequency side of the excitation pattern. Moore and Sek (1998) argued that at least both sides of the excitation pattern should be compared, and that for low center frequencies the time-related cues, such as phase locking could have an effect as well. The studies agree that the detection and discrimination of glides is little affected by duration, center frequency, or direction (up or down). However, the previous results are of little help for the synthesis of instrument tones, since the range of center frequencies was from 0.5 to 6.0 kHz and the shape of the glide as a function of time was unnatural to real instruments.

The objective of this study is to set perceptually motivated guidelines for the need to implement the initial pitch glide in string instrument synthesis. The test tones were synthesized and the pitch glides were defined in a way typical of string instruments. The results of two listening experiments are reported.
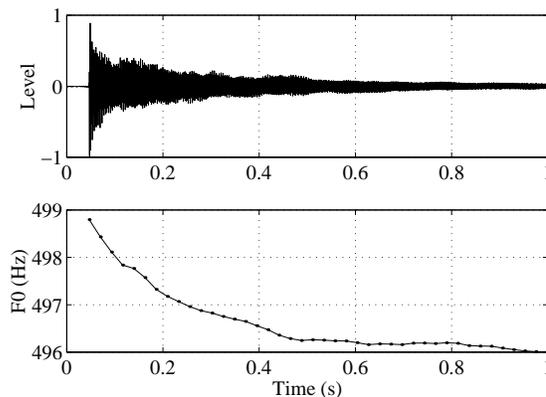

Figure 1: Waveform of a single tone played on the electric guitar (top) and its short-time fundamental frequency estimate, which shows a typical descent (bottom).

2 Listening tests

The detection thresholds for initial pitch glides were measured in two separate listening experiments. In both experiments, the independent variable was the transition span Δf, i.e., the extent of the pitch glide; in addition, fundamental frequency f_0 was used as a parameter in experiment I and decay time constant τ in experiment II.

2.1 Test sounds

The test sounds were generated using additive synthesis for easy control of the pitch glides. Each tone consisted of its 20 lowest harmonics and had a duration of 2.0 seconds. An exception was made in the highest tone $E5$ (659 Hz), for which only 16 harmonics could be generated before meeting the Nyquist limit of 11 kHz at the sampling rate of 22.05 kHz.

Realistic sounding test tones were created by copying the initial amplitudes and decay characteristics of real tones. The acoustic guitar was taken as a reference, for its behavior is well-known and plenty of analysis data are available. The initial phases of the harmonics were randomized. The amplitude decay was controlled by two parameters, the time constant τ of the overall decay and a frequency-dependent damping coefficient a, which corresponds to the feedback coefficient of the one-pole loop filter in a digital waveguide string model. The initial amplitudes as well as the decay parameters were chosen according to (Välimäki and Tolonen 1998).

The independent variable was the extent of the pitch glide in Hz. The pitch contour decreased exponentially with time from the highest value $f_0 + \Delta f$ towards the steady state fundamental frequency f_0. The time constant τ_f of the frequency descent was 50% of the overall time constant τ of amplitude decay (Legge and Fletcher 1984).

In experiment I, the thresholds were measured at four different fundamental frequencies to cover the whole pitch range of the acoustic guitar: E_5 (659.26 Hz), F_4 (349.23 Hz), G_3 (196 Hz), and $B\flat_2$ (116.5 Hz). The overall decay time constant τ and frequency-dependent damping coefficient a were kept constant.

The motivation for the second experiment is the known relation of the time constant τ of amplitude decay and time constant τ_f of pitch decay in string instrument tones (Legge and Fletcher 1984). The perceptual tolerances for amplitude decay have been published previously (Tolonen and Järveläinen 2000), giving the allowable deviation of the time constant from the reference value. However, varying the time constant even within this tolerance affects the time constant of the pitch decay and hence the duration of the pitch glide, which can cause a shift in the detection threshold. This was studied by fixing the fundamental frequency to 196 Hz and varying the time constant τ_f. Three values were used for the time constant according to the previous results on the perceptual tolerances: 100%, 80%, and 60% of the reference value 0.39
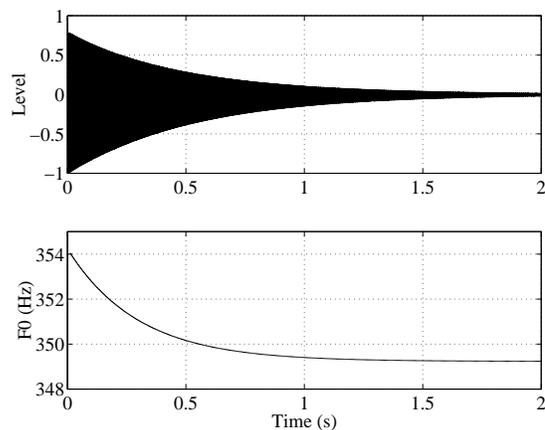

Figure 2: Waveform of a synthetic guitar tone F_4 (top) and its fundamental frequency with a 5-Hz pitch glide (bottom).

s, which is 50% of the measured time constant τ of the overall amplitude decay (0.77 s).

2.2 Subjects and test method

Five subjects participated in both experiments. They were 20-30 years old, and all had previous experience in psychoacoustic listening tests. None of them reported any hearing defects, and they were allowed to practise before the test. The sound samples were played through headphones from a computer.

A standard tone without a pitch glide and a stimulus tone with a pitch glide were presented to the subject sequentially in random order, and the task was to judge whether the sounds were same or different. Five values of glide extent Δf were used for each fundamental frequency in experiment I and four for each time constant in experiment II. Each trial was judged four times together with as many corresponding fake trials (two standard tones, no stimulus) in random order.

A detected difference was either a *hit* or a *false alarm*, depending on whether the trial actually included the stimulus tone or not. A measure of correct answers $P(C)$ was derived for each condition from the proportion of hits and false alarms as follows (Yost 1994):

$$P(C) = \frac{p(\text{hit}) + (1 - p(\text{false alarm}))}{2} \qquad (1)$$

The function has values between 0.50, which corresponds to chance level with equal proportions of hits and false alarms, and 1.0, which requires 100% hit proportion and no false alarms. The detection threshold was estimated by finding the midpoint (i.e., the 75% point) of this function. If the threshold was not directly evident in the data, it was interpolated between the nearest higher and lower scores.

3 Results

3.1 Effect of fundamental frequency

The effect of fundamental frequency on the detection of pitch glides was studied in experiment I. The detection thresholds increased monotonically with fundamental frequency. The mean thresholds were 3.1 Hz, 4.4 Hz, 5.4 Hz, and 11.7 Hz for $B\flat_2$, G_3, F_4, and E_5, respectively. The situation turns upside down when the thresholds are expressed on the logarithmic scale, see Fig. 3. For $B\flat_2$, the median of individual thresholds is 52 cents (1/100 of a semitone) – more than half of a semitone, while for the highest tone E_5 it is 30 cents.
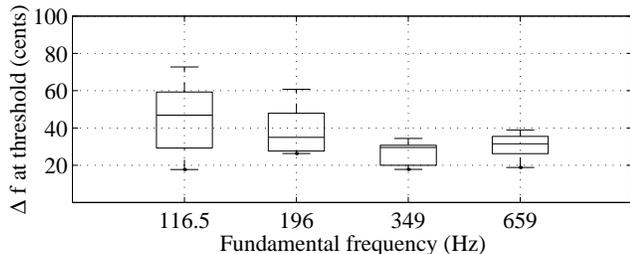
Figure 3: Listening test results with fundamental frequency as a parameter. Boxplot of Δf (cents) at threshold.

The results were roughly normally distributed, but the error variance between different tones was typically unequal. However, these differences were reasonably equalized by a transformation from the linear frequency scale to the auditorily motivated ERB (Equivalent Rectangular Bandwidth) scale as follows (Glasberg and Moore 1990):

$$\text{Number of ERBs} = 21.4 \log_{10}(4.37F + 1), \qquad (2)$$

where F is frequency in kHz. The analysis of variance (ANOVA) was now performed on the data presented on the ERB scale (Lehman 1991). The result was nonsignificant ($p = 0.25$), indicating that the threshold could well be the same for each f_0. The results are summarized in Fig. 4. The figure presents a boxplot of the results with the median and the 75% and 25% quartiles. The mean thresholds are between 0.083 and 0.122 ERB.

Since a true difference in the mean thresholds for each note is relatively unlikely, the data were collapsed across fundamental frequency. The sample mean across all subjects and fundamental frequencies is 0.10 ERB, which could be considered an estimate of the constant detection threshold of pitch glides within the pitch range of the acoustic guitar.

3.2 Effect of decay rate

The listening test data from experiment II were processed in the same way as the data from the first experiment. Again
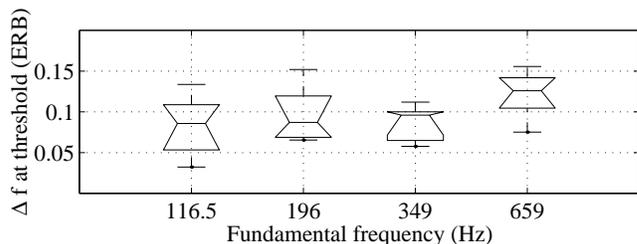
Figure 4: Listening test results with fundamental frequency as a parameter. Boxplot of Δf (ERB) at threshold.

the ANOVA revealed no significant differences among the means of the different conditions ($p = 0.85$). This suggests that relatively great variations in the time constant of the pitch glide have no effect on detecting the glide. The results are shown in Fig. 5. The boxplot is shown in the top figure and the mean thresholds on the ERB scale in the bottom figure.

Of course, it is likely that the glide duration would show some effect if the glides were short enough. Such behavior would be unnatural to string instruments, but it is an interesting task to extend this study to synthetic sounds with greater variation in the decay characteristics. A third, rather informal test was conducted to study extremely fast and slowly decaying sounds. The time constant τ_f was now varied with five linear steps between 20% and 180% of the original value of 0.39 s, i.e., between 0.08 s and 0.7 s. The procedure was the same as before, but only three subjects participated. The shortest τ_f caused a considerable increase in the thresholds while the other thresholds were similar to the previous results. The mean threshold for $\tau_f = 0.08$ s was 10.8 Hz. For the four greater values of τ_f the mean thresholds were between 5.6 Hz and 7.7 Hz.

4 Conclusions and future work

The thresholds for detecting initial pitch glides in string instrument tones were measured. It was found that when expressed on the ERB scale, the thresholds remain roughly constant with fundamental frequency within the pitch range of the acoustic guitar. Furthermore, no significant effect was detected for glide duration, which was controlled through the time constant of the pitch decay. This suggests that a constant threshold of 0.1 ERB could be proposed for string instrument sounds at least within the studied range. Thus any pitch glide weaker than this could be left unimplemented.

For instance, the electric guitar tone in Fig. 1 exhibits an initial pitch glide of 3 Hz at a fundamental frequency of 466 Hz, i.e., about 0.06 ERB. Thus it would probably remain inaudible to most listeners. On the other hand, a kantele tone presented in (Välimäki et al. 1999) shows a pitch glide of
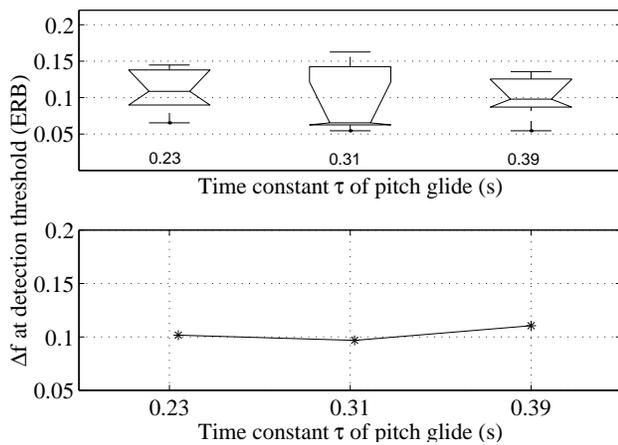
Figure 5: Listening test results with decay time constant as a parameter. Top: Boxplot with the median and 75% and 25% quartiles, bottom: mean thresholds on the ERB scale.

almost 0.3 ERB, which should be clearly audible.

Although the effect of glide duration was insignificant in the range typical of string instruments, the thresholds increased for very fast decaying sounds. For these short sounds, there was less time to listen to the steady pitch f_0. This suggests that the subjects were using end point detection, i.e., comparing the absolute frequencies at both ends of the glide, to detect the glides. Both Madden and Fire (1997) and Moore and Sek (1998) prevented their subjects from using such cues, which might partly explain why the thresholds from this study are lower than what they measured. On the other hand, in musical context the expectance of a certain pitch may well connect the detection of a glide to absolute frequency, and this is why our study enabled the end point cues.

The contradiction between glide detection for short and long tones may also concern real instrument sounds, not only synthetic sounds with unrealistic decay characteristics. Cutting off the steady part of the tone might have a similar effect on the detection threshold as shortening the decay time showed in this study. This calls for some more experiments.

The results from this study can be applied in digital sound synthesis, where computational savings can be achieved by ignoring the pitch glides whenever they are inaudible. Coding is another field of application. The new structured methods of sound representation (Vercoe et al. 1998) make it desirable to control the perceptual features of sounds separately.

Acknowledgments

References

Glasberg, B. and B. Moore (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Res. 47*, 103–138.

Legge, K. and N. Fletcher (1984). Nonlinear generation of missing modes on a vibrating string. *J. Acoust. Soc. Am. 76*(1), 5–12.

Lehman, R. S. (1991). *Statistics and Research Design in the Behavioral Sciences.* Belmont, California: Wadsworth Publishing Company.

Madden, J. and K. Fire (1997). Detection and discrimination of frequency glides as a function of direction, duration, frequency span, and center frequency. *J. Acoust. Soc. Am. 102*(5), 2920–2924.

Moore, B. C. and A. Sek (1998). Discrimination of frequency glides with superimposed random glides in level. *J. Acoust. Soc. Am. 104*(1), 411–421.

Serra, X. and J. Smith (1990). Spectral modeling synthesis: a sound analysis/synthesis system based on a deterministic plus stochastic decomposition. *Computer Music J. 14*(4), 12–24.

Smith, J. O. (1998). Principles of digital waveguide models of musical instruments. In M. Kahrs and K. Brandenburg (Eds.), *Applications of Digital Signal Processing to Audio and Acoustics*, Chapter 10, pp. 417–466. Kluwer.

Tolonen, T. and H. Järveläinen (2000). Perceptual study of decay parameters in plucked string synthesis. Preprint 5205, 109th Conv. Audio Eng. Soc., Los Angeles, California.

Tolonen, T., V. Välimäki, and M. Karjalainen (2000). Modeling of tension modulation nonlinearity in plucked strings. *IEEE Trans. Speech and Audio Processing 8*(3), 300–310.

Välimäki, V., M. Karjalainen, T. Tolonen, and C. Erkut (1999). Nonlinear modeling and synthesis of the kantele – a traditional Finnish string instrument. In *Proc. Int. Computer Music Conf.*, Beijing, China, pp. 220–223. Full paper and sound examples are available at http://www.acoustics.hut.fi/publications/.

Välimäki, V., M. Laurson, C. Erkut, and T. Tolonen (2000). Model-based synthesis of the clavichord. In *Proc. Int. Computer Music Conf.*, Berlin, Germany, pp. 50–53. Full paper and sound examples are available at http://www.acoustics.hut.fi/publications/.

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

Välimäki, V., T. Tolonen, and M. Karjalainen (1998). Signal-dependent nonlinearities for physical models using time-varying fractional delay filters. In *Proc. Int. Computer Music Conf.*, Ann Arbor, Michigan, pp. 264–267. Full paper and sound examples are available at http://www.acoustics.hut.fi/publications/.

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.

Yost, W. A. (1994). *Fundamentals of Hearing – An Introduction* (3rd ed.). New York: Academic Press.