Audibility of the timbral effects of inharmonicity in stringed instrument tones

Hanna Järveläinen, Vesa Välimäki, Matti Karjalainen
Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology
P.O. Box 3000, 02015 HUT, Espoo, Finland
hanna.jarvelainen@hut.fi, vesa.valimaki@hut.fi, matti.karjalainen@hut.fi

Abstract: Listening tests were conducted to find the audibility of inharmonicity in musical sounds produced by stringed instruments, such as the piano or the guitar. The audibility threshold of inharmonicity was measured at five fundamental frequencies. Results show that the detection of inharmonicity is strongly dependent on the fundamental frequency f_0. A simple model is presented for estimating the threshold as a function of f_0. The need to implement inharmonicity in digital sound synthesis is discussed.

PACS numbers: 43.75.Cd, 43.75.Mn
Date Received: January 11, 2001 **Date Accepted:** March 18,2001

1. Introduction

The frequencies of the partials of string instrument sounds are not exactly harmonic. This is caused by stiffness of real strings, which contributes to the restoring force of string displacement together with string tension. The strings are dispersive: the velocity of transversal wave propagation is dependent on frequency. If the string parameters are known, the frequencies of the stretched partials can be calculated in the following way (Fletcher et al., 1962):

$$f_n = nf_0\sqrt{1+Bn^2} \tag{1}$$

where n is the partial number, f_0 is the fundamental frequency of the string completely without stiffness, and B is the inharmonicity coefficient whose value depends on string design and parameters. For an unwound string, the value of B can be calculated as follows (Fletcher et al., 1962):

$$B = \frac{\pi^3 Q d^4}{64 l^2 T} \tag{2}$$

where Q is Young's modulus, d is the diameter, l is the length, and T is the tension of the string.

Inharmonicity is not necessarily unpleasant. Fletcher et al. (1962) pointed out that a slightly inharmonic spectrum added certain warmth into the sound. They found that synthesized piano tones sounded more natural when the partials below middle C were inharmonic.

The effect of mistuning one spectral component in an otherwise harmonic complex is well known. Moore et al. (1985) reported that the thresholds for detecting mistuning decreased progressively with increasing harmonic number and increasing fundamental frequency. In their experiment, the test tones were complex tones with 12 harmonics at equal levels, and mistuning was expressed as a percentage of the harmonic frequency.

Moore's group also showed that mistuning is heard in different ways depending on the harmonic number (Moore et al., 1985). Shortening the stimulus duration produced a large impairment in performance for the higher harmonics, whereas it had only little effect on the performance for the lower harmonics. It was reasoned that, particularly for long durations, beats provide an effective cue but, for short durations, many cycles of beats cannot be heard. For the lower harmonics, beats were generally inaudible, and the detection of mistuning appeared to be based on hearing the mistuned component stand out from the complex. The thresholds varied only weakly with duration.

Rocchesso and Scalcon (1999) studied the bandwidth of correct positioning of the partials of synthesized piano tones. They found cutoff frequencies above which it was unnecessary to synthesize inharmonicity. For low tones, the relevant bandwidth was smaller than for higher tones, but on the other hand, many more partials were included in the frequency range for low tones. They also stated that the effect of inharmonicity was unimportant for the highest part of the keyboard.

There are few studies with sounds that exhibit systematic inharmonicity like string instrument tones. In the current study, the audibility of inharmonicity was investigated as a function of fundamental frequency with sound duration as parameter. The aim was to find general rules for the need to implement inharmonicity in digital sound synthesis. If inharmonicity were ignored, computational savings could be achieved. For instance in digital waveguide synthesis, an additional high-order allpass filter is needed to implement inharmonicity (Jaffe and Smith, 1983), (Paladin and Rocchesso, 1992), (Van Duyne and Smith, 1994). Preliminary results of this study were published in (Järveläinen et al., 1999).

2. Listening experiments

Subjects were required to distinguish between a complex tone whose partials were exactly harmonic and an otherwise identical complex tone whose partials were mistuned. The threshold for detection of inharmonicity, expressed as B required for 75% area under the ROC (Receiver Operating Characteristics) curve, was studied at fundamental frequencies 55 Hz (A_1), 82.4 Hz (E_2), 220 Hz (A_3), 392 Hz (G_4), and 1108.7 Hz ($C\#_6$). The effect of duration on the detection of inharmonicity was also studied.

2.1. Test sounds

The test signals were generated using additive synthesis to enable accurate control of the frequency and amplitude of each partial. The sampling rate was 22.05 kHz. The spectrum of the tones had a lowpass character of the form 1/frequency, which is similar to spectra of many string instruments. The decay of all partials was exponential with a time constant $\tau = 0.5$ seconds. The initial phase of each partial was chosen randomly.

The tones contained all partials of the fundamental frequency f_0 up to 10 kHz. A constant cutoff frequency was chosen because it was impossible to use the same number of partials for every fundamental frequency. Up to 50 partials are important to the perception of inharmonicity for low bass tones (Rocchesso and Scalcon, 1999). However, fewer than ten partials could be generated for the highest tone before meeting the Nyquist limit. Realizing that the variable number of partials might affect the audibility of inharmonicity, we reasoned that a constant cutoff would still be the most practical solution. A constant upper limit was also needed to keep the spectral width of all test sounds equal. Galembo and Cuddy (1997) showed that without control of the spectral width, inharmonicity changes the balance between high and low frequencies and creates an impression of sharpening.

A range of B was found through initial listening to cover the probable threshold of audibility in each case. The values of B were uniformly spaced between zero and the chosen B_{max}.

A pitch increase due to inharmonicity was heard for the highest tone. The subjects might listen to pitch differences and ignore the timbral effect of inharmonicity if the pitch change were audible. To prevent this, the pitches of the harmonic references were matched to those of the inharmonic tones. The pitch corrections were based on an earlier study (Järveläinen et al., 2000), in which it was found that the pitch judgment of inharmonic tones is sufficiently modeled by one of the higher partials. The order of the dominant partial decreases with increasing fundamental frequency. The pitch estimates were derived from the sixth partial for A_1

and E_2, from the third partial for A_3 and G_4, and from the second partial for $C\#_6$. The pitch compensation was negligible for all notes except $C\#_6$, for which the maximal compensation was 6.2 Hz.

One problem remained. The subjects could detect inharmonicity by detecting the pitch difference between test pairs that included an inharmonic tone and those that did not. For this reason, the f_0 was randomized over a range of 1/4 of a semitone from trial to trial.

The effect of duration was studied by two tone lengths, 1.5 s and 300 ms. The short samples were generated by fading out the longer sounds with a 5 ms ramp. The time constant remained the same for both durations. The longer sounds were used in experiment 1 and the shorter ones in experiment 2. In both experiments, the sounds were played with a 0.3 s time gap between them.

2.2. Subjects and test method

Seven subjects participated in the first experiment and six in the second one. The listeners were 20-30 years old, and all had musical training on some stringed instrument, either the piano or the guitar. None of them reported any hearing defects. One of the listeners was the author HJ. The sound samples were played through headphones from a Silicon Graphics workstation using the GuineaPig software (Hynninen and Zacharov, 1999). Before the test, the subjects were allowed to practice until they made consistent judgments.

The subjects heard pairwise a perfectly harmonic reference sound and a possibly inharmonic sound, and the task was to decide whether they sounded the same or different. Eight values of B (including $B = 0$) were used for each fundamental frequency. An inharmonic tone was included in half of the test pairs, and each inharmonic pair was judged four times. The playback order of the sound pairs was randomized within conditions, and the harmonic reference was the first sound within a pair twice and the second one twice (Guilford, 1956).

A reported difference was either a *hit* or a *false alarm*, depending on whether an inharmonic sound was included in the test pair. A measure of sensitivity d' was derived for each condition from the proportion of hits ($p(\mathrm{H})$) and false alarms ($p(\mathrm{FA})$) (Green and Swets, 1988) as follows:

$$d' = z(p(\mathrm{H})) - z(p(\mathrm{FA})), \tag{3}$$

where z is the inverse of the normal cumulative distribution function. The advantage of this measure is that it eliminates the effects of response bias that is caused by favoring either same or different. The hit and false alarm proportions were used to construct ROC curves by assuming underlying equal-variance normal distributions of responses for “same” and “different” trials. The area under the curve is between 0.5, which corresponds to chance level, and 1.0, which corresponds to 100% performance. The detection threshold was found by estimating the B required to reach the midpoint (i.e., the 75% point) of the area under the ROC curve, which corresponds to $d' \approx 1$ (Yost, 1994).

3. Test results

The audibility thresholds are shown as a function of fundamental frequency for each subject in Fig. 1. The sample duration was 1.5 seconds. The mean thresholds are shown in Table 1. Inharmonicity was most detectable for the lowest note and least for the highest note: the mean threshold was more than 1,000 times higher for $C\#_6$ than for A_1. A cause for this might be the measure of inharmonicity that was used. The lowest sounds simply have more partials in the 10-kHz band, and the highest partials are also mistuned by a greater percentage than the lowest ones.

The thresholds were roughly normally distributed, but the error variance was unequal for each note. The differences in variance were reasonably equalized by a logarithmic transform before further statistical analysis. Analysis of variance (ANOVA) was then performed on the test data (Lehman, 1991). The result was highly significant ($p < 0.001$), indicating that there is a significant difference between the average results for at least two of the notes. However, by a follow-up test made by the Tukey procedure (Lehman, 1991), it was found that the differences between $A1$ and $E2$ and between A_3 and G_4 were not significant at the $\alpha = 0.05$ level.

Table 1. Mean thresholds and standard deviations

f_0	B at mean threshold	Standard deviation σ
A_1	0.00000014	0.00000010
E_2	0.00000024	0.00000023
A_3	0.000010	0.0000093
G_4	0.000018	0.0000081
$C\#_6$	0.00021	0.00017

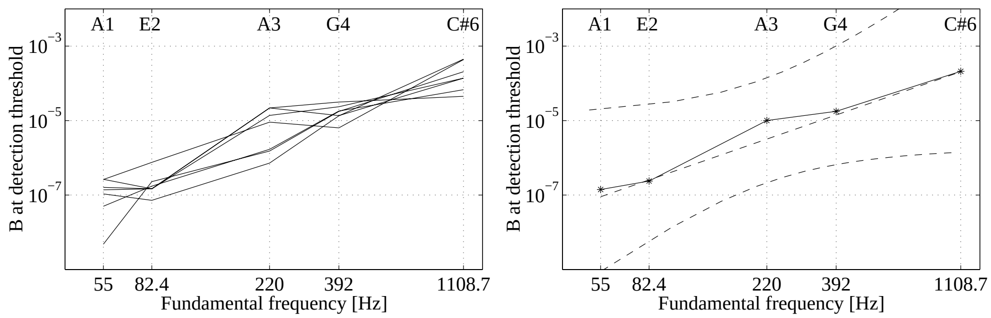

Fig. 1. Left: The individual thresholds for the seven listeners of the first experiment at A_1, E_2, A_3, G_4, and $C\#_6$. Right: The linear fit from Eq. (4) with a 90% confidence band (dashed lines) and the mean thresholds over all subjects (solid line). Sample duration was 1.5 seconds (long).

The thresholds show a strong linear trend when a logarithmic scale is used both for frequency and B, as seen in Fig. 1. A straight line was fitted to the mean threshold values in the least-squares sense. This way a simple formula was derived that could be used to model the audibility threshold as a function of fundamental frequency:

$$\ln B = 2.57 \ln f_0 - 26.5 \tag{4}$$

The fitted line is illustrated in Fig. 1. The mean threshold fits well to the estimated line (see Fig. 1). Some of the individual thresholds differ from the line at A_3, where a few subjects had higher thresholds than at G_4. The nonmonotonic behavior is not surprising, because the subjects reported that they used several different cues to detect inharmonicity. The performance can depend on the existence of certain cues at different fundamental frequencies and the subject's sensitivity to a particular cue, such as beating.

3.1. Effect of duration

The study was continued by a further experiment to find out the effect of sound duration. The test was repeated by using sound samples of 300 ms duration with a time gap of 300 ms between them in the test cycle. The results are presented in Fig. 2. The thresholds show the same kind of trend as in the long duration test; however, they are clearly higher for the lowest two notes.

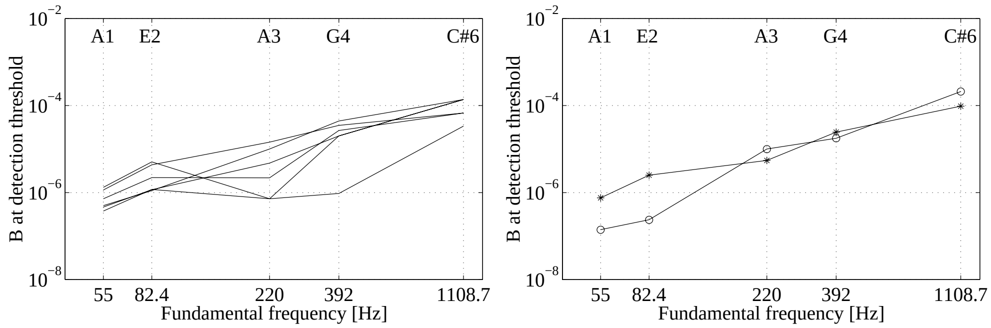

Fig. 2. Left: Individual thresholds for the six listeners of the second experiment at A_1, E_2, A_3, G_4, and $C\#_6$. Sample duration was 300 ms (short). Right: Mean thresholds from both experiments – long samples (o) and short samples (*).

One subject complained that it was harder to judge the short tones because the second tone in a test pair started so soon after the first one. To eliminate the possible effect of this, the test was repeated, and the time gap between the tones in the test cycle was increased to 1.5 s, which made the starting time of the second tone equal to that in the long-sample test.

It was verified by statistical t-tests that the differences between short tones with and without the extra time gap were insignificant. However, the differences between long and short tones were significant for E_2 and nearly significant for A_1 at the $\alpha = 0.1$ level. This suggests that there is a true effect of sound duration so that the detection of inharmonicity becomes more difficult with decreasing duration for low tones. For high tones, inharmonicity seems to affect the timbre immediately after the attack, while for low tones, the effect is detected after some time of listening. The mean thresholds from both experiments are compared in Fig. 2.

4. Discussion

Typical values of B for piano strings lie roughly between 0.00005 ... 0.0005 for bass tones and 0.0007 ... 0.015 for treble tones (Conklin, 1999). Though the sounds are less inharmonic in the bass range, inharmonicity is detected more easily than for the highest tones. Based on thresholds measured in the current study, inharmonicity would be clearly audible at low fundamental frequencies. In the treble range, it seems that inharmonicity could also be detected, although the threshold values of B are close to the possible values of B in real instruments. Whether inharmonicity is audible would thus depend on the quality of the instrument.

To reduce the computational load in digital sound synthesis, it would be desirable to ignore inharmonicity whenever it is inaudible. The results suggest that implementing the effect of inharmonicity would be necessary at least in the bass range. In the treble range, inharmonicity might in some cases remain inaudible, and computational savings could be achieved by omitting the allpass filter responsible for the effect of inharmonicity.

There can be several causes for the better performance at lower frequencies. The subjects mentioned that they were using beats as a cue. Beats were mostly audible at low fundamental frequencies. When most of the decay phase was cut off, the thresholds increased in the bass range. In this respect our results agree with those of Moore et al. (1985), who found that the thresholds increased for higher order partials with decreasing duration. In our experiments, the performance weakened in the bass range where the proportion of higher order partials was greater than in the treble range.

On the other hand, their results showed a decrease of thresholds with increasing fundamental frequency, whereas ours showed the opposite. The contradiction could be caused by different measuring variables. Moore et al. (1985) expressed the thresholds as percent mistuning for each partial individually and used a fixed number of partials. In the current study, the bandwidth was fixed instead of the number of partials, resulting in a larger number of partials at low fundamental frequencies. Thus, in the bass range, very small values of B produced enough mistuning in the higher partials to be detected. In the treble, a higher B was needed to produce audible mistuning in lower order partials.

Questions that remain open for future research are the relation of duration and the number of partials, the effect of spectral width, relative level of the partials, and different decay rates between them. The digital synthesis of stringed instruments calls for a practical viewpoint in the sense that inharmonicity should be considered relative to sound production in acoustic instruments. The definitive objective is to find a set of general computational models for evaluating the need to implement inharmonicity.

Acknowledgments

This work was financially supported by the Academy of Finland, the Pythagoras Graduate School, and Nokia Research Center. The authors would like to thank the test subjects and also Alexander Galembo, who kindly provided his publications.

References and links

Conklin, H. A., (**1999**), "Generation of partials due to nonlinear mixing in stringed instruments," *J. Acoust. Soc. Am.* **105**(1), pp. 536–545.

Fletcher, H., E. D. Blackham, and R. Stratton, (**1962**), "Quality of piano tones," *J. Acoust. Soc. Am.* **34**(6), pp. 749–761.

Galembo, A. and L. Cuddy, (**1997**), "String inharmonicity and the timbral quality of piano bass tones: Fletcher, Blackham, and Stratton (1962) revisited," Report to the 3rd US Conf. Music Perception and Cognition.

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

Guilford, J. P., (**1956**), *Psychometric Methods* (McGraw-Hill, New York).

Hynninen, J. and N. Zacharov, (**1999**), "GuineaPig – a generic subjective test system for multichannel audio," Presented at the Audio Engineering Society 106th Conv., May 1999, Munich, Germany, preprint no. 4871.

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

Järveläinen, H., T. Verma, and V. Välimäki, (**2000**), "The effect of inharmonicity on pitch in string instrument sounds," in *Proc. Int. Computer Music Conf.*, pp. 237–240, Berlin, Germany.

Järveläinen, H., V. Välimäki, and M. Karjalainen, (**1999**), "Audibility of inharmonicity in string instrument sounds, and implications to digital sound synthesis," in *Proc. Int. Computer Music Conf.*, pp. 359–362, Beijing, China.

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

Moore, B. C. J., R. W. Peters, and B. R. Glasberg, (**1985**), "Thresholds for the detection of inharmonicity in complex tones," *J. Acoust. Soc. Am.* **77**(5), pp. 1861–1867.

Paladin, A. and D. Rocchesso, (**1992**), "A dispersive resonator in real time on MARS workstation," in *Proc. Int. Computer Music Conf.*, pp. 146–149, San Jose, CA.

Rocchesso, D. and F. Scalcon, (**1999**), "Bandwidth of perceived inharmonicity for physical modeling of dispersive strings," *IEEE Trans. Speech and Audio Processing* **7**(5), pp. 597–601.

Van Duyne, S. and J. O. Smith, (**1994**), "A simplified approach to modeling dispersion caused by stiffness in strings and plates," in *Proc. Int. Computer Mus. Conf.*, pp. 407–410, Aarhus, Denmark.

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

