Publication P2

H. Järveläinen and T. Tolonen. Perceptual tolerances for the decaying parameters in string instrument synthesis. In *Journal of the Audio Engineering Society*, 49(11), pages 1049–1059, 2001.

Perceptual Tolerances for Decay Parameters in Plucked String Synthesis*

HANNA JÄRVELÄINEN, *AES Member*, **AND TERO TOLONEN**[1]

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

The audibility of variations of decay parameters in plucked string synthesis was studied by a listening test. The decay was described by an overall time constant and a frequency-dependent parameter. Two fundamental frequencies, tone durations, and types of excitations were used for both parameters. The results indicate that variations between 25 and 40% in the time constant are inaudible. This suggests large perceptual tolerances for the decay parameters. The results are applied in model-based audio processing.

0 INTRODUCTION

With the development of interactive multimedia terminals and increasing bandwidth in both fixed and wireless networks, multimedia communication becomes an increasingly important concept. Until now, audio and musical content has typically been stored and transmitted as sampled signals, possibly encoded with an auditorily motivated method. Recently the MPEG-4 multimedia standard included structured methods for the representation of synthetic audio and effects as parametric models and control data [1]–[4]. This object-based approach enables novel interactive solutions as well as applications where high-quality content is required to be delivered in a low-bandwidth channel, such as in mobile multimedia services.

The perception of timbre has been an active field of research for several decades (see [5], [6] for overviews and references). However, the research into perceptual aspects of model-based sound synthesis has been limited. The perception of inharmonicity in piano tones was studied from a synthesis viewpoint in [7], [8]. Another work on the perception of inharmonicity with a model-based synthesis motivation was presented in [9], [10]. The perception of the vibrato of violin tones was investigated in [11].

Similarly to natural audio coding [12], significant improvements to model-based synthesis can be expected when the human auditory system is taken into account. The knowledge in human perception can be exploited in the parameterization of the models, designing coding schemes for the control data, and developing auditorily motivated analysis methods for the calibration of synthesis models.

In this work we investigate the classical acoustic guitar and its parameterization as a computational model that can be used for the generation of high-quality synthetic tones. One of the crucial perceptual features of plucked string tones is the decay. Even when the pluck and the body response are captured well, the tone is perceived as unnatural if the decay is inaccurate. This paper describes a listening experiment that was conducted for the perception of variations of the overall decay and the frequency-dependent decay of a plucked string instrument tone.

The synthesis model is based on the digital waveguide approach [13]–[15], and it uses the commuted waveguide synthesis (CWS) technique [16], [17]. The model is computationally efficient and is suited well for applications where high-quality object-based music representation and synthesis are required. The decay of a tone is determined by a loop filter with two parameters—a loop gain parameter that controls the overall decay and a loop pole parameter for the frequency-dependent decay. Typically when the model is used for sound synthesis, the parameters are obtained by time–frequency analysis of recorded tones, preferably played in an anechoic chamber [18]–[20].

The objective of the listening experiment is to estimate thresholds for detecting a variation in the decay pattern of a plucked string tone. Our approach is very closely related to the particular synthesis model chosen: rather than attempting to obtain results that would be generalizable for a wide set of exponentially decay tones, we concentrated on the present model and its two decay parameters. This approach was motivated from a model-based analysis/synthesis viewpoint, as explained in Section 4.

* Presented at the 109th Convention of the Audio Engineering Society, Los Angeles, CA, 2000 September 21–25, under the title "Perceptual Study of Decay Parameters in Plucked String Synthesis"; revised 2000 December 8.

[1]Now with Luxxon Corporation, Mountain View, CA 94043, USA.

The paper is organized as follows. The CWS model used for the synthesis of the tones is reviewed in Section 1. Section 2 describes the listening experiments, including experiment methods, subjects, stimuli, and variations of the investigated model parameters. The results of the experiments are analyzed in Section 3, and they are applied in model-based audio processing in Section 4. Section 5 concludes the paper and proposes future directions for research in model-based and perceptual sound source modeling.[2]

1 PLUCKED STRING MODEL

The block diagram of the string model is presented in Fig. 1. The model is derived from a bidirectional digital waveguide [13]–[15], and it uses the method of commuted synthesis [16], [17]. The derivation of this model of Fig. 1 from a digital waveguide model is presented in [21].

The transfer function for the string is

$$S(z) = \frac{1}{1 - z^{L_1} F(z) H_1(z)} \tag{1}$$

where L_1 is the length of the delay line,

$$H_1(z) = \frac{g(1-a)}{1 - az^{-1}} \tag{2}$$

is the one-pole low-pass loop filter which determines the decay of the tone, and $F(z)$ is a fractional delay filter modeling the noninteger part of the string length [22], [23]. The use of a fractional delay filter allows for the fine-tuning of the pitch. Since we wish to study the decay of the tone caused by the loop filter, we have chosen to use an all-pass filter with maximally flat phase delay for the loop filter. With an all-pass filter as the fractional delay implementation, the only component producing losses in the model of Eq. (1) is the loop filter $H(z)$. The string transfer function $S(z)$ is fully described by the string length L in the samples, the loop gain g, and the loop filter cutoff parameter a.

The model of Eq. (1) can be used for the synthesis of high-quality tones when the commuted synthesis technique is employed. In commuted synthesis, the string model parameters are calibrated based on the analysis of recorded tones [18]–[20]. After parameter calibration, the inverse of the model in Eq. (1) is used to inverse-filter the recorded tones. If the calibration is done properly, the residual of the inverse-filtering is a relatively short signal, which consists of the contributions of the pluck and the body response. When this excitation is used in synthesis, an identical copy to the original is obtained. The excitation signals are typically windowed into a length of approximately several hundreds of milliseconds in order to save memory. Other methods of reducing the length of the excitation signal include modeling of the signal with a digital filter and the use of separate parametric models for the most prominent body resonances [19], [24], [25].[2]

[2]Sound examples of test signals and synthetic guitar tones are available at http://www.acoustics.hut.fi/publications/papers/aes109-decay.

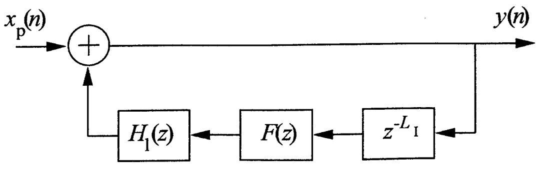


Fig. 1. Block diagram of string model [18].

2 LISTENING TESTS

The thresholds for detecting a change in the decay were measured by listening experiments. Two separate experiments were conducted, one for detecting a change in the overall decay (parameter g) and one for the frequency-dependent decay (parameter a).

Two different fundamental frequencies, tone durations, and types of excitation of the CWS model were used, totaling eight test sets for both parameters. Each of the sets consisted of nine test signals, including one signal that was equal to the reference signal. Four of the signals exhibited longer decay and the remaining four shorter decay than the reference tone.

The selected tones were G_3 (196.0 Hz) played on the fifth fret of the D string, and F_4 (349.2 Hz) played on the first fret of the high E string. The tones were selected so that one of them is played on a nylon string and one on a wound string.

The durations of the signals were 0.6 and 2.0 s. Fig. 2 shows the amplitude envelope of a test signal. The signal is attenuated after the specified duration, using a linear ramp with a length of 100 ms. This is perceived somewhat similar to damping of the tone. The durations were selected so that the short tones corresponded to a length typically found in music while the long tones allowed a more accurate perception of change in the timbre of the tone.

Natural sounding tones were generated with an excitation signal obtained by analysis of recorded tones. An impulse was used in half the sets so that basically the impulse response of the string model of Eq. (1) was perceived. The bandwidth of the impulse response is wider compared to the natural tone, which is typically of more low-pass nature.

2.1 Test Signals

The test signals were generated using the model of Eq. (1). The parameters for the synthesis models were obtained using methods presented in [20]. Table 1 shows the estimated synthesis parameters for the reference tones in the two cases.

The equalized residual signals that were used for excitation in half the test signals were computed using the technique presented in [26], [19]. In the method a sinusoidal model of the tone is computed and subtracted from the original signal to yield a residual signal. The residual signal is equalized using the inverse of the model with estimated model parameters and shortened to a desired length using time-domain windowing.

Preliminary listening experiments were performed in order to find a suitable range for the parameters to be tested. In the g parameter test, the time constant of the overall decay of the tone was computed using

$$\tau = -\frac{L}{f_s \ln(g)} \tag{3}$$

where τ is the time constant in seconds and L is the string length in samples f_s/f_0. The time constant was varied in a systematic way in the listening experiment, as explained in the following subsection. One of the motivations for using the time constant of the overall decay instead of using the g parameter directly is that since the value of g is typically very close to 1, a relatively small change in the parameter value can result in a drastic change in the overall decay.

The a parameter is related to frequency-dependent decay. The results of the preliminary listening experiments suggested that the a parameter behaves sufficiently well in a meaningful range for detecting the threshold. Thus in the listening experiments we varied the a parameter directly.

Fig. 3 shows an example of how the magnitude responses of the loop filters change when the a and g parameters are varied. Since the loop filter is the only component in the loop causing attenuation, the plotted magnitude responses define the attenuation of the tone in each case. Notice that the dc gain is constant at the a parameter test [Fig. 3(a)] while the frequency tilt varies. In the g parameter test [Fig. 3(b)], the shape of the magnitude responses is approximately constant while the overall level varies.

In the g parameter test the time constants of the test signals were varied linearly on both sides of the reference time constant. However, the results of preliminary experiments suggested that the relative difference in time constants should be different for the time constants that are larger or smaller than the reference time constant. In addition, different time constant ranges were selected for different fundamental frequencies and different durations.

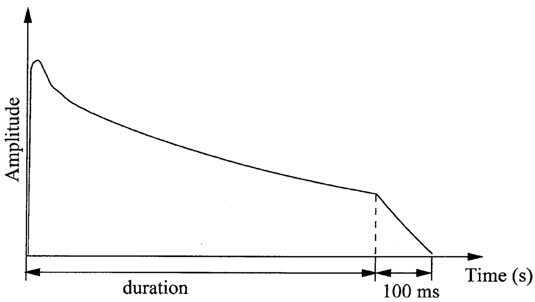


Fig. 2. Example of amplitude envelope of test signals.

Table. 1 Synthesis model parameters for reference tones.

	G_3	F_4
L	112.7515	62.8568
g	0.9934	0.9952
a	0.2219	0.0771

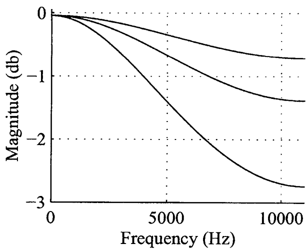
Also the a parameter was varied linearly on both sides of the reference value. Again, the preliminary experiments suggested that the relative difference in a parameter values should be different on the two sides. This time, however, different parameter value ranges were selected only for different durations.

The test sets and the corresponding parameters of the two experiments are presented in Tables 2 and 3. Sets 1–4 correspond to long test signals (duration 2.0 s) and sets 5–8 to short signals (0.6 s). The signals are paired according to the tones so that sets 1–2 and 5–6 correspond to tone F_4 and sets 3–4 and 7–8 to G_3. Every pair consists of signals obtained with equalized residuals and an impulse as an excitation signal.

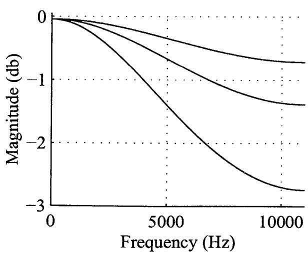
a_{ref} and g_{ref} are the a and g parameter values of the reference tone, and τ_{ref} is the time constant corresponding to g_{ref}. In Table 2 the last two rows show the ratios of the minimum and maximum time constants to τ_{ref}. The time constants of decay of the test signals are distributed linearly between these extrema and τ_{ref}. In Table 3 the last two rows show the ratios of the minimum and maximum values of the a parameters to a_{ref} The values of the a parameters of the test signals are distributed linearly between these extrema and a_{ref}.

Fig. 4 shows the amplitude envelopes of the impulse responses in the g parameter test sets 1–2, 3–4, 5–6, and 7–8 [Fig. 4(a)–(d)]. The middle (fifth) curve of each plot corresponds to the reference tone. In the short signals, the variations of the time constants are quite large, although the amplitude envelopes plot almost atop of each other (see Table 2).

Fig. 5 depicts the magnitude responses of the loop filters $H(z)$ of the a parameter experiment sets 1–2, 3–4, 5–6, and 7–8 [Fig. 5(a)–(d)]. Again, the middle (fifth) curve corresponds to the loop filter of the reference tone.



(a)



(b)

Fig. 3. Loop filter transfer functions when a (a) and g (b) parameters are varied.

Notice that the actual difference in magnitude responses varies between the G_3 tone case [Fig. 5(a), (c)] and the F_4 tone case [Fig. 5(b), (d)] although the relative differences in the pole locations are almost equal (see Table 3).

2.1.1 Additional Test Sets

Additional tests were designed to study the thresholds as a function of fundamental frequency. For this purpose, two new test sets were generated for both g and a parameter experiments to cover the whole pitch range of the acoustic guitar (see Tables 4 and 5). In this way the thresholds could be measured at four fundamental frequency points, $B\flat_2$, G_3, F_4, and E_5. In these limited experiments, the duration of each sound was 2.0 s, and inverse-filtered excitation was used in the synthesis model.

2.2 Subjects and Test Methods

Five experienced subjects with normal hearing were selected, two of whom were the authors. The listeners were personnel of the HUT Acoustics Laboratory and post- and undergraduate students with a musical background. The experiments were conducted in a listening room, one subject at a time. The sounds were played from an SGI O2 computer through Sennheiser HD 580 earphones at an average sound pressure level of 78 dB. The level of the individual test sounds differed from the aver-

Table 2. Synthesis model parameters for g parameter test.

Set	1	2	3	4	5	6	7	8
Excitation	Impulse	Inverse-filtered	Impulse	Inverse-filtered	Impulse	Inverse-filtered	Impulse	Inverse-filtered
Tone	F_4	F_4	G_3	G_3	F_4	F_4	G_3	G_3
Duration (s)	2.0	2.0	2.0	2.0	0.6	0.6	0.6	0.6
a_{ref}	0.0771	0.0771	0.2219	0.2219	0.0771	0.0771	0.2219	0.2219
g_{ref}	0.9952	0.9952	0.9934	0.9934	0.9952	0.9952	0.9934	0.9934
τ_{ref} (s)	0.60	0.60	0.77	0.77	0.60	0.60	0.77	0.77
L	62.86	62.86	112.75	112.75	62.86	62.86	112.75	112.75
τ_{min}/τ_{ref}	0.6	0.6	0.45	0.45	0.45	0.45	0.45	0.45
τ_{max}/τ_{ref}	2.15	2.15	3.25	3.25	9.0	9.0	30.0	30.0

Table 3. Synthesis model parameters for a parameter test.

Set	1	2	3	4	5	6	7	8
Excitation	Impulse	Inverse-filtered	Impulse	Inverse-filtered	Impulse	Inverse-filtered	Impulse	Inverse-filtered
Tone	F_4	F_4	G_3	G_3	F_4	F_4	G_3	G_3
Duration (s)	2.0	2.0	2.0	2.0	0.6	0.6	0.6	0.6
a_{ref}	0.0771	0.0771	0.2219	0.2219	0.0771	0.0771	0.2219	0.2219
g_{ref}	0.9952	0.9952	0.9934	0.9934	0.9952	0.9952	0.9934	0.9934
L	62.86	62.86	112.75	112.75	62.86	62.86	112.75	112.75
a_{min}/a_{ref}	0.45	0.45	0.45	0.45	0.5	0.5	0.5	0.5
a_{max}/a_{ref}	2.15	2.15	2.15	2.15	2.0	2.0	2.0	2.0

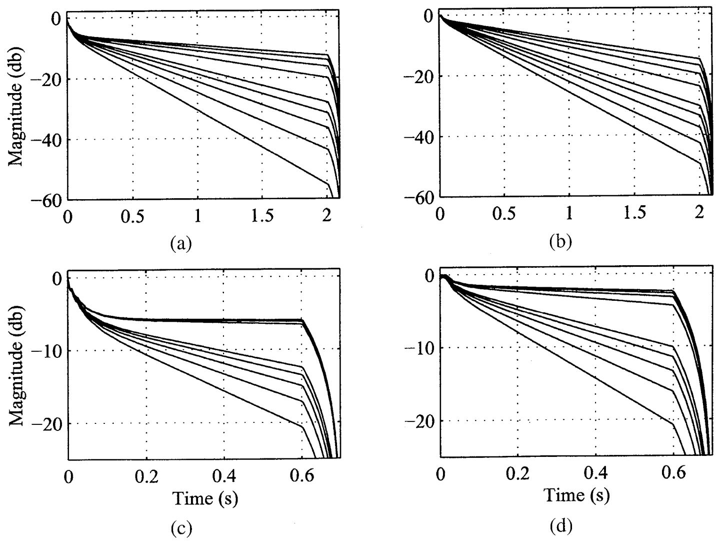


Fig. 4. Amplitude envelopes of string model impulse responses corresponding to g parameter experiment. (a) Sets 1–2. (b) Sets 3–4. (c) Sets 5–6. (d) Sets 7–8.

age, but since this was due to the natural behavior of the CWS model, the differences were not equalized. The GuineaPig2 software [27] was used for the control of playback and for recording the results.

Two separate tests were designed, one for each parameter. Each test signal was compared to its reference, including the reference itself. With eight different kinds of signals (treatments) and nine test signals (conditions) in each set, this results in 72 different test pairs per experiment. Each pair was played 25 times. Both experiments were divided into five one-hour sessions. The 72 test pairs were played five times per session, and each subject was only allowed to participate in one session per day. The first session of each experiment was regarded as practice and excluded from the analysis. The order of playback was randomized as well as the order of the reference and test signals in each pair.

The subjects were forced to judge each test pair as either equal or different. The thresholds for detecting a difference in the decay pattern were measured separately for decay times longer and shorter than the reference value. The method of constant stimuli was used [28]. As an example, the judgments of one of the subjects concerning the shorter decay times of test set 2 of the g parameter test are shown in Fig. 6. The 100% level of "different" judgments was obtained with τ_{min}, and the 0% level with τ_{ref}.

The judgment data were used to approximate the psychometric function, and the threshold of audibility was obtained by estimating the 50% point of the function. When the proportion of "different" judgments is higher than that, it is expected that the subject perceives a difference. The estimation was made by normal interpolation [28]. The method assumes that the psychometric function relating the judgments to the parameter values of the test signals is a cumulative normal curve. The judgment proportions are transformed into corresponding standard measure values z. The 50% point now corresponds to $z = 0$, that is, the mean of the noncumulative distribution, which is estimated by interpolating between the nearest positive and negative values of the measure. The thresholds were estimated for each of the subjects in all cases in a similar manner.

3 RESULTS

3.1 Data Analysis

Because the number of available listeners was limited, the test followed a factorial within-subjects design: each

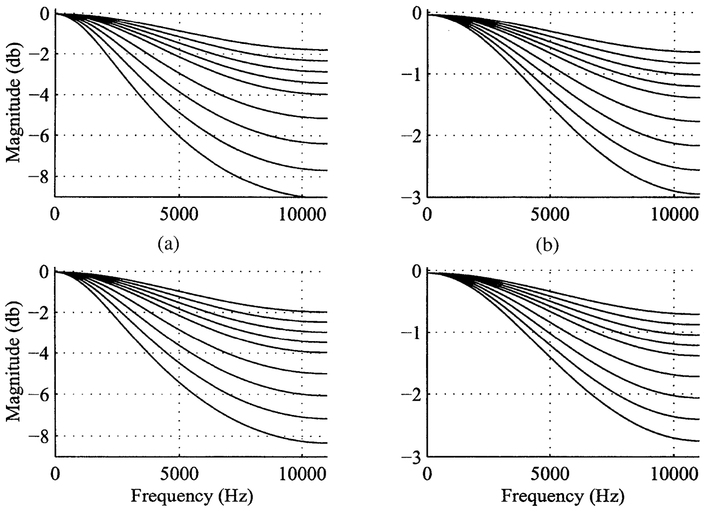


Fig. 5. Loop filter magnitude responses for a parameter experiment. (a) Sets 1–2. (b) Sets 3–4. (c) Sets 5–6. (d) Sets 7-8.

Table 4. Synthesis model parameters for additional g parameter test sounds.

Set	1	2
Excitation	Inverse-filtered	Inverse-filtered
Tone	Bb_2	E_5
Duration (s)	2.0	2.0
a_{ref}	0.4031	0.0484
g_{ref}	0.9924	0.9952
L	188.68	33.28
τ_{min}/τ_{ref}	0.45	0.6
τ_{max}/τ_{ref}	3.25	2.15

Table 5. Synthesis model parameters for additional a parameter test sounds.

Set	1	2
Excitation	Inverse-filtered	Inverse-filtered
Tone	Bb_2	E_5
Duration (s)	2.0	2.0
a_{ref}	0.4031	0.0484
g_{ref}	0.9924	0.9952
L	188.68	33.28
a_{min}/a_{ref}	0.7	0.7
a_{max}/a_{ref}	1.6	1.6

subject received each of the eight treatments (test sets) [29]. The results were roughly normally distributed within each treatment level, but the error variance within levels was typically unequal. The different ranges of the g and a parameters on both sides of the reference values suggest that the thresholds are proportionally rather than linearly symmetric around the reference. This was also seen by a quick examination of the results. It was therefore decided to make a 10-base logarithmic transform to the results in the analysis phase. This way the error variance between treatments was reasonably equalized to fulfill the requirements of the analysis of variance.

Analysis of variance (ANOVA) [29] was performed on the threshold data to detect a significant difference between the mean thresholds of the five subjects. After a significant p value, pairwise follow-up tests were conducted to make inferences about the significance of some particular characteristics of the sounds.

The Tukey honestly significant difference (HSD) test is appropriate for exploring differences in pairs of means after a significant results from ANOVA [29]. It gives a value for the smallest possible significant difference between two-condition means. Any difference greater than that can be considered significant.

3.2 *g* Parameter Experiment Results

In the gain parameter test, the thresholds varied most distinctly with the sound duration. For the long sounds they remained roughly the same, regardless of other variables. The upper thresholds were about 40% higher and the lower thresholds about 25% lower than the reference value of the time constant of decay. However, with short sounds the upper thresholds increased drastically. The lower thresholds decreased correspondingly, but more weakly.

The upper and lower thresholds (corresponding to decay times longer and shorter than the reference value, respectively) are shown in Figs. 7 and 8. The mean thresholds over the subjects and the corresponding standard deviations are shown in Table 6.

The ANOVA results were highly significant for both upper and lower thresholds ($p = 1.1896 \times 10^{-9}$ and $p = 3.475 \times 10^{-8}$, respectively). This suggests that there are actual differences between the mean thresholds of the test sets.

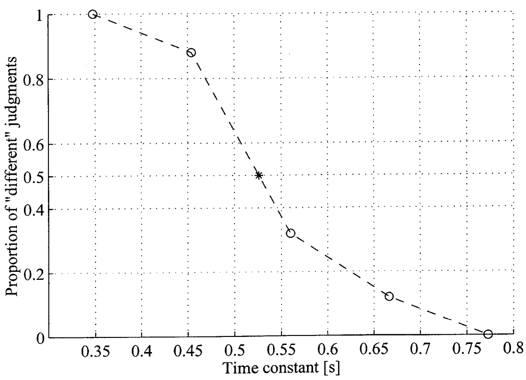


Fig. 6. Estimating lower 50% threshold for sound set 4 in g parameter test.

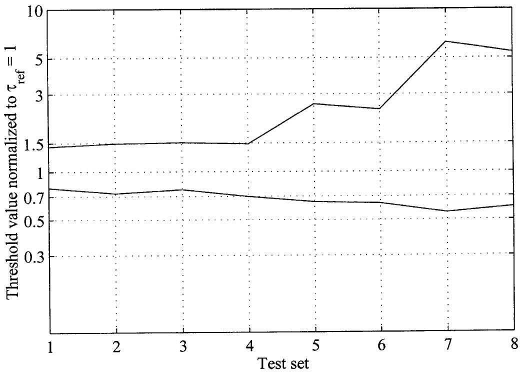


Fig. 7. Mean upper and lower thresholds of audibility in g parameter experiment. Values have been normalized according to $\tau_{ref} = 1$.

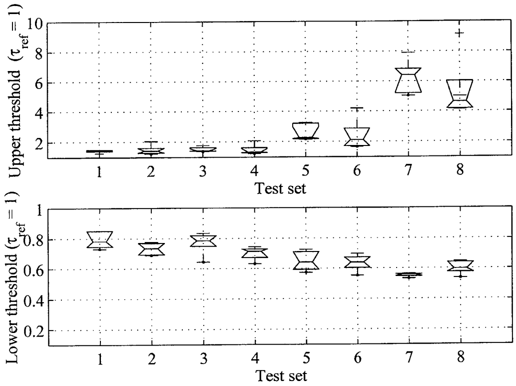


Fig. 8. Box plot of thresholds for individual listeners in g parameter test.

Table 6. Sample means μ presented as τ/τ_{ref} and corresponding standard deviations σ^2 of g parameter thresholds.

Set	Upper μ (τ/τ_{ref}) Upper σ^2	Lower μ Lower σ^2
1	1.4309	0.7923
	0.0984	0.0558
2	1.4725	0.7317
	0.3254	0.0405
3	1.4904	0.7781
	0.1541	0.0627
4	1.4951	0.6998
	0.3366	0.0435
5	2.5235	0.6502
	0.6471	0.0654
6	2.4369	0.6351
	1.0314	0.0552
7	5.6375	0.5559
	0.7444	0.0143
8	5.4107	0.6018
	2.1142	0.0454

A set of post-hoc tests (Tukey HSD) followed. A pairwise comparison was made between test sets that differed only by one parameter value. For instance, test sets 1 and 5 are identical except the sounds of test set 1 are long and those of set 5 are short. Others that differ only by duration are sets 2 and 6, 3 and 7, and 4 and 8. Similar pair comparisons were made for matching sets that differ only by the fundamental frequency or the type of excitation. A significant difference was detected for both upper and lower thresholds by practically all comparisons of sets that differ by duration. The lower threshold data showed a significant effect of the fundamental frequency, but only for short sounds. No other comparison was significant.

3.3 *a* Parameter Experiment Results

The results of the *a* parameter experiment were different from the *g* experiment in at least one respect. The duration of the sounds had no significant effect on the thresholds. The thresholds are shown in Figs. 9 and 10. The mean values of the *a* parameter and the corresponding standard deviations are shown in Table 7.

The ANOVA was significant for both lower and upper threshold, but only on an $\alpha = 0.05$ error probability level ($p = 0.0318$ and $p = 1.5286 \times 10^{-4}$, respectively). This time the follow-up tests did not reveal any significant effects, except for the type of excitation in two cases. At the lower threshold, a significant effect was detected between test sets 5 and 6, and at the upper threshold between sets 7 and 8. A rough examination of the results suggests that the type of excitation may explain the variation of the results in other cases as well. In all cases the thresholds were nearer to the reference value when impulse excitation was used. This could be due to the greater bandwidth of the impulse excitation compared to the inverse-filtered one.

A group comparison test [29] was made between all the sets that used impulse excitation and all those that used inverse-filtered excitation. A comparison variable was computed by subtracting the thresholds of all impulse excitation samples from the thresholds of inverse filtered samples. A student's *t* test was made on the mean of the comparison variable with Scheffe's adjustment [29]. The results were highly significant for both upper and lower thresholds. We can conclude that the type of excitation affected the detection thresholds in the *a* parameter tests, but other significant effects were not found.

3.4 Results of Additional Tests

Since the effect of fundamental frequency remained unclear in both experiments, additional experiments were made to cover the pitch range of the guitar. Two additional fundamental frequencies were chosen. The test was limited to only long sounds with inverse-filtered excitations. The corresponding measurements (test sets 2 and 4) from the first experiments were combined with the new ones. In this way the thresholds could be studied in four frequency points with the fundamental frequency as the only independent variable. The frequencies were 116.9 Hz, 196.0 Hz, 349.2 Hz, and 662.6 Hz, corresponding to $B^{\flat}_2$, G_3, F_4, and E_5, respectively.

The results of the additional tests are seen in Figs. 11 and 12 and tabulated in Tables 8 and 9 for the *g* and *a* parameter tests, respectively. To complete the analysis, a

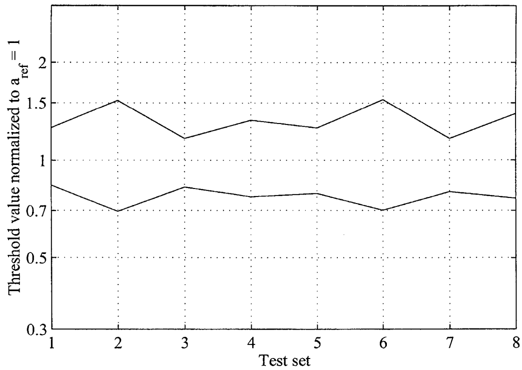


Fig. 9. Mean upper and lower thresholds of audibility in *a* parameter experiment. Values have been normalized according to $a_{ref} = 1$.

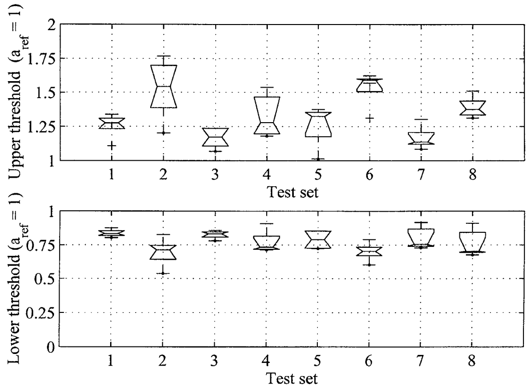


Fig. 10. Box plot of thresholds for individual listeners in *a* parameter test.

Table 7. Sample means μ presented as a/a_{ref} and corresponding standard deviations σ^2 of *a* parameter thresholds.

Set	Upper μ (a/a_{ref}) Upper σ^2	Lower μ Lower σ^2
1	1.2585	0.8379
	0.0878	0.0263
2	1.5268	0.6958
	0.2193	0.1045
3	1.1667	0.8273
	0.0740	0.0290
4	1.3281	0.7720
	0.1558	0.0818
5	1.2575	0.7904
	0.1470	0.0640
6	1.5389	0.7014
	0.1281	0.0673
7	1.1668	0.7999
	0.0822	0.0817
8	1.3902	0.7632
	0.0769	0.1003

logarithmic transformation was again made of the results. According to the ANOVA, the effect of the fundamental frequency was not significant in the *g* parameter test (p = 0.1221 for the lower and p = 0.8049 for the upper thresholds). The *a* parameter results were significant on the α = 0.05 level, but not on the α = 0.01 level (p = 0.0046 for the lower and p = 0.0342 for the upper thresholds). In the *a* parameter test, the mean thresholds of the lowest fundamental frequency differed significantly from the other three frequency points, but other significant effects were not found. In either case, no clearly monotonic effect was detected as a function of increasing or decreasing fundamental frequency.

3.5 Discussion of Results

It can be concluded that the thresholds for detecting differences in the decay pattern are fairly robust against changes in parameter values. The exception was that the thresholds increased strongly with decreasing duration in the *g* parameter experiment. In the *a* parameter experiment this was not observed. This is natural, since the overall decay time varied in the *g* parameter test, whereas the *a* parameter affected the tone mainly immediately after the attack. The change in the beginning of the tone is audible with short sounds as well as with long ones, but it is very hard to detect differences in the overall decay time based on only the beginning of the sound.

Instead of duration, the *a* parameter results were affected by the type of excitation signal used in the synthesis model. The thresholds decreased with impulse excitation. This is probably due to the larger bandwidth of these test signals compared to those with inverse-filtered excitation.

No other significant effects were detected. The thresholds remained roughly constant as a function of the fundamental frequency. This suggests that a constant minimum tolerance could be recommended for the deviation of the decay parameters. From a perceptual viewpoint, relatively large deviations in the decay parameters can be accepted. The test results indicate that a variation of the time constant between about 75 and 140% of the reference value can be allowed in most cases. With short sounds the tolerance is even greater. For the *a* parameter, the average acceptable range of deviation is between 83 and 116% of the reference value. The large perceptual range suggests that the results can be effectively applied in model-based audio processing, as described in the following section.

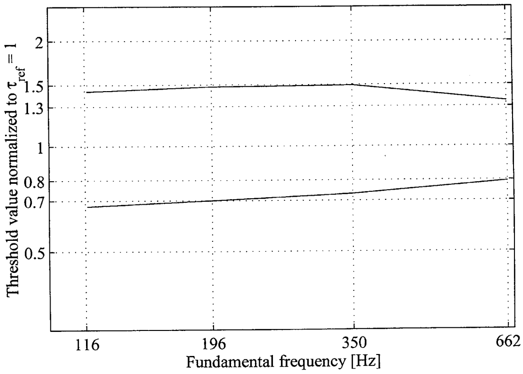


Fig. 11. Mean upper and lower thresholds as a function of fundamental frequency in additional *g* parameter test. Values have been normalized according to $\tau_{ref} = 1$.

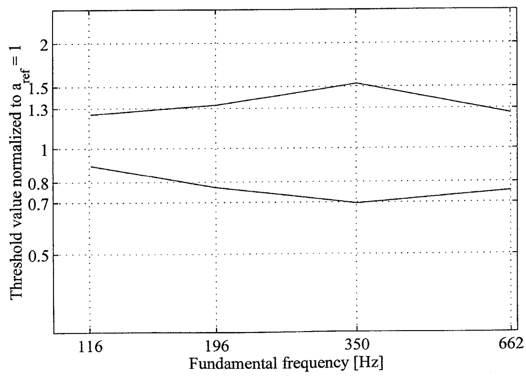


Fig. 12. Mean upper and lower thresholds as a function of fundamental frequency in additional *a* parameter experiment. Values have been normalized according to $a_{ref} = 1$.

4 APPLICATION OF RESULTS IN MODEL-BASED AUDIO PROCESSING

The results of the listening experiments indicate the range of deviation in overall and frequency-dependent

Table 8. Sample means μ presented as τ/τ_{ref} and corresponding standard deviations σ^2 of *g* parameter thresholds as a function of fundamental frequency.

F_0 (Hz)	Upper μ (τ/τ_{ref}) Upper σ^2	Lower μ Lower σ^2
116.86	1.4366	0.6742
	0.1487	0.1226
196.0	1.4806	0.6998
	0.3355	0.0435
350.8	1.4922	0.7318
	0.3196	0.0406
662.6	1.3462	0.7944
	0.2306	0.0769

Table 9. Sample means μ presented as a/a_{ref} and corresponding standard deviations σ^2 of *a* parameter thresholds as a function of fundamental frequency.

F_0 (Hz)	Upper μ (a/a_{ref}) Upper σ^2	Lower μ Lower σ^2
116.86	1.2499	0.8928
	0.0850	0.0210
196.0	1.3281	0.7720
	0.1558	0.0818
350.8	1.5268	0.6958
	0.2193	0.1045
662.6	1.2591	0.7545
	0.0658	0.0118

decay that can be tolerated from a perceptual viewpoint. The perceptual tolerance can be used in several applications of model-based processing. On the analysis side, the perceptual thresholds provide a means for assessing the performance of an analysis system that estimates the parameters from recorded tones. In a model-based representation, the thresholds give guidelines on how the decay of a tone is optimally represented. Figs. 13 and 14 show examples of how the results may be interpreted from a more general viewpoint. This approach is elaborated in the two subsections that follow.

Fig. 13 illustrates the audibility thresholds of the g parameter test set 1. The amplitude envelopes corresponding to tones with values of g at the upper and lower thresholds are plotted with solid lines. The dashed line depicts the amplitude envelope of the reference tone. The horizontal dash–dotted line shows the amplitude level corresponding to $1/e$ of the maximum. The vertical lines indicate the time constants of the tones in the three cases, that is, the time instants where the tone has decayed to $1/e$ of the maximum value. The tones with overall decay between the solid lines are perceptually indistinguishable from the reference tone.

The audibility thresholds corresponding to the a parameter test set 1 are depicted in Fig. 14. In this case the solid lines indicate the frequency envelopes corresponding to the upper and lower thresholds, and the dashed line depicts the frequency envelope of the reference tone. Fig. 14(a) shows the thresholds up to 10 kHz. Fig. 14(b) is a closeup of the low-frequency band with the horizontal dash–dotted line indicating the −6-dB level. The vertical dash–dotted lines show the −6-dB cutoff frequencies of the three tones. Again, tones with frequency envelopes between the solid lines are perceptually indistinguishable from the reference tone.

4.1 Model Parameterization

When a model of Eq. (1) is used for synthesis, the most straightforward parameterization is to deal with the values of g and a directly. However, although we are investigating a specific model here, it is useful to have its parameterization as more generic parameters so that other synthesis methods may also be supported. In that case it is particularly advantageous to have the boundaries for perceptually acceptable deviations from the target values.

The g parameter determines approximately the overall decay of the tone. The time constants of the overall decay of tones $B^{\flat}_2$, G_3, F_4, and E_5 were 1.21, 0.77, 0.60, and 0.31 s, respectively. The corresponding values of the g parameter were 0.9924, 0.9952, 0.9934, and 0.9952. The time constant parameterization is generic in that it can be used with other synthesis methods, and it gives a clear picture of the decay of each tone with boundaries for perceptually acceptable deviation, compared to the application-specific direct parameterization.

In the listening tests the a parameter values were varied directly. Typically, the a parameter behaves better compared to the g parameter and sufficiently well for many applications. However, the parameter is not descriptive in that it does not readily give an idea about the frequency-dependent decay character. A frequency-domain approach may help to give a better insight into frequency-dependent decay. An example is presented in Fig. 14, where the −6-dB cutoff frequencies of the reference tone and of the tones at audibility thresholds are plotted. Naturally, the frequency envelope depends not only on the string model but also on the excitation signal used.

The range between the thresholds is relatively broad in both examples of Figs. 13 and 14. This provides a starting point for the generation of coding schemes for model-based music representation.

4.2 Model Parameter Analysis

An iterative parameter extraction algorithm for the loop filter parameters of the model of Eq. (1) is presented in [20]. The algorithm first optimizes the parameters based on detected amplitude envelopes of the partials, as described in [30], [19]. A synthetic tone is computed using the estimated parameters, and its amplitude envelope is

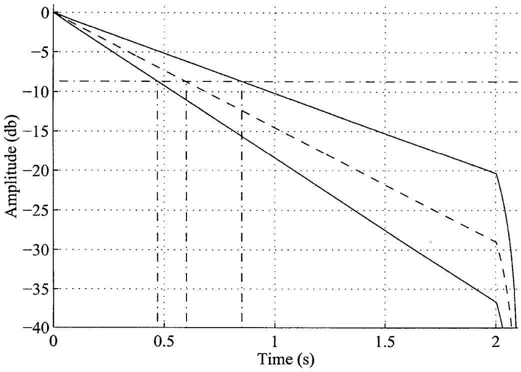


Fig. 13. Amplitude envelopes of tones with g parameter variation detection for test set 1. — upper and lower thresholds; – – – reference tone; horizontal; – · – · – $1/e$ level; vertical time constants in three cases.

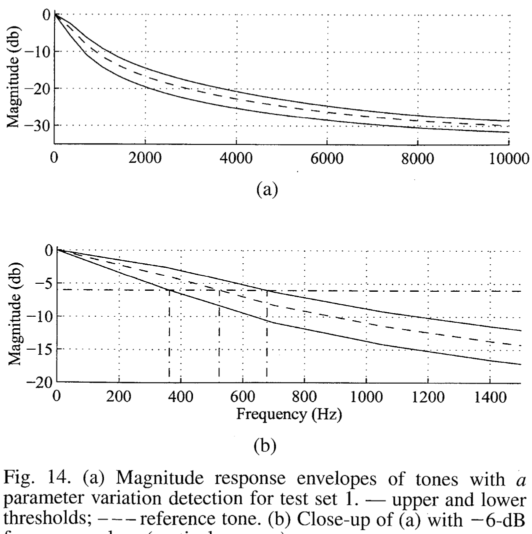


Fig. 14. (a) Magnitude response envelopes of tones with a parameter variation detection for test set 1. — upper and lower thresholds; – – – reference tone. (b) Close-up of (a) with −6-dB frequency values (vertical – · – · –).

compared to that of the original tone. If there is a sufficient discrepancy between the decay of the envelopes of the original and the synthetic tones, an iterative optimization algorithm is used to detect the optimal loop-filter parameters.

The results of the g parameter test can be used in such an iterative algorithm. First the results provide a perceptually motivated threshold for deciding whether the iterative algorithm should be used. If the initial parameter estimates produce an overall decay that cannot be perceptually distinguished from the decay of the original tone, the parameters can be readily used in synthesis applications. In addition, the perceptual thresholds provide thresholds for the iterative optimization algorithm: when the difference between the time constants of decay of original and synthetic tones is imperceptible, the iteration may be finished.

Besides the comparison of the overall decay, the frequency-dependent decay may also be included in such an iterative parameter optimization procedure. In this case the frequency envelopes of the original and the synthetic tones are compared. Note that the frequency characteristic of the excitation signal needs to be taken into account.

5 CONCLUSIONS AND FUTURE DIRECTIONS

We have reported a listening experiment on detecting a change in the decay of plucked string instrument tones. The results provide audibility thresholds for variations of the overall and frequency-dependent decay with a specific sound synthesis model. The results were applied in model-based audio processing.

The presented experiment gives good insight into the perception of decay variations in this specific application, although the experiment was forced to be limited to a rather small set of test signals. The research will continue by conducting experiments with other plucked string instruments, to other aspects of plucked string tones, and with other sound sources. At this point, model-based audio processing faces a huge unexplored field of research in perceptual sound source modeling.

Another path for future work is to develop the analysis system discussed. Most likely this will also give directions for designing new perceptual studies and listening experiments.

This study supports that model-based audio and music processing can gain significant benefits by taking into account the human auditory system. This will in turn help to make the model-based approach even more attractive in future audio and music applications.

6 ACKNOWLEDGMENT

The authors wish to thank Prof. Matti Karjalainen for many fruitful discussions and support throughout this work. The financial support of the GETA and Pythagoras graduate schools, Nokia Research Center, Jenny ja Antti Wihurin rahasto (Jenny and Antti Wihuri Foundation), Tekniikan edistämissäätiö, and Nokia Foundation is gratefully acknowledged.

7 REFERENCES

[1] B. R. Vercoe, W. G. Gardner, and E. D. Scheirer, "Structured Audio: Creation, Transmission, and Rendering of Parametric Sound Representations," *Proc. IEEE*, vol. 86, no. 5 (1998).

[2] "ISO/IES IS 14496-3, "Information Technology—Coding of Audiovisual Objects, Part 3: Audio," International Standards Organization, Geneva, Switzerland (1999).

[3] E. D. Scheirer, Y. Lee, and J. W. Wang, "Synthetic and SNHC Audio in MPEG-4," *Signal Process.: Image Commun.*, vol. 15, pp. 445–461 (2000).

[4] J. Lazzaro and J. Wawrzynek, "The MPEG-4 Structured Audio Book," http://ww.cs.berkeley.edu/~lazzaro/sa/book/index.html.

[5] J. M. Hajda, R. A. Kendall, and E. C. Carterette, "Methodological Issues in Timbre Research," in *Perception and Cognition of Music*, I. Deliège and J. Sloboda, Eds. (Psychology Press, Hove, UK, 1997), pp. 253–306.

[6] S. McAdams, "Recognition of Auditory Sound Sources and Events," in *Thinking in Sound: The Cognitive Psychology of Human Audition* (Oxford University Press, New York, 1993).

[7] D. Rocchesso and F. Scalcon, "Bandwidth of Perceived Inharmonicity for Musical Modeling of Dispersive Strings," *IEEE Trans. Speech Audio Process.*, vol. 7, pp. 597–601 (1999).

[8] F. Scalcon, D. Rocchesso, and G. Borin, "Subjective Evaluation of the Inharmonicity of Synthetic Piano Tones," in *Proc. Int. Computer Music Conf.* (1998), pp. 53–56.

[9] H. Järveläinen, V. Välimäki, and M. Karjalainen, "Audibility of Inharmonicity in String Instrument Sounds, and Implications to Digital Sound Synthesis," in *Proc. Int. Computer Music Conf.* (Beijing, China, 1999), pp. 359–362.

[10] H. Järveläinen, T. Verma, and V. Välimäki, "The Effect of Inharmonicity on Pitch in String Instrument Sounds," in *Proc. Int. Computer Music Conf.* (Berlin, Germany, 2000), pp. 237–240.

[11] M. Mellody and G. H. Wakefield, "The Time–Frequency Characteristic of Violin Vibrato: Modal Distribution Analysis and Synthesis," *J. Acoust. Soc. Am.*, vol. 107, pp. 598–611 (2000).

[12] N. Jayant, J. Johnston, and R. Safranek, "Signal Compression Based on Models of Human Perception," *Proc. IEEE*, vol. 81, pp. 1385–1422 (1993).

[13] J. O. Smith, "Music Applications of Digital Waveguides," Tech. Rep. STAN-M-39, CCRMA, Department of Music, Stanford University, Stanford, CA (1987).

[14] J. O. Smith, "Physical Modeling Using Digital Waveguides," *Computer Music J.*, vol. 16, no. 4, pp. 74–91 (1992).

[15] J. O. Smith, "Acoustic Modeling Using Digital Waveguides," in *Musical Signal Processing*, C. Roads, S. Pope, A. Piccialli, and G. De Poli, Eds. (Swets and Zeitlinger, Lisse, The Netherlands, 1997), ch. 7, pp.

221–264.

[16] J. O. Smith, "Efficient Synthesis of Stringed Musical Instruments," in *Proc. Int. Computer Music Conf.* (Tokyo, Japan, 1993), pp. 64–71.

[17] M. Karjalainen, V. Välimäki, and Z. Jánosy, "Towards High-Quality Sound Synthesis of the Guitar and String Instruments," in *Proc. Int. Computer Music Conf.* (Tokyo, Japan, 1993), pp. 56–63.

[18] V. Välimäki, J. Huopaniemi, M. Karjalainen, and Z. Jánosy, "Physical Modeling of Plucked String Instruments with Application to Real-Time Sound Synthesis," *J. Audio Eng. Soc.*, vol. 44, pp. 331–353 (1996 May).

[19] T. Tolonen, "Model-Based Analysis and Resynthesis of Acoustic Guitar Tones," Rep. 46, Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology, Espoo, Finland (1998).

[20] C. Erkut, V. Välimäki, M. Karjalainen, and M. Laurson, "Extraction of Physical and Expressive Parameters for Model-Based Sound Synthesis of the Classical Guitar," presented at the 108th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 48, p. 354 (2000 Apr.), preprint 5114.

[21] M. Karjalainen, V. Välimäki, and T. Tolonen, "Plucked-String Models: From Karplus–Strong Algorithm to Digital Waveguides and Beyond," *Computer Music J.*, vol. 22, no. 3, pp. 17–32 (1998).

[22] V. Välimäki, "Discrete-Time Modeling of Acoustic Tubes Using Fractional Delay Filters," Ph.D. thesis, Helsinki University of Technology, Espoo, Finland (1995).

[23] T. I. Laakso, V. Välimäki, M. Karjalainen, and U. K. Laine, "Splitting the Unit Delay—Tools for Fractional Filter Design," *IEEE Signal Process. Mag.*, vol. 13, pp. 30–60 (1996).

[24] M. Karjalainen and J. O. Smith, "Body Modeling Techniques for String Instrument Synthesis," in *Proc. Int. Computer Music Conf.* (Hong Kong, 1996), pp. 232–239.

[25] V. Välimäki, M. Karjalainen, T. Tolonen, and C. Erkut, "Nonlinear Modeling and Synthesis of the Kantele—A Traditional Finnish String Instrument," in *Proc. Int. Computer Music Conf.* (Beijing, China, 1999 Oct.), pp. 220–223.

[26] T. Tolonen and V. Välimäki, "Analysis and Synthesis of Guitar Tones Using Signal Processing Methods," in *Proc. Finnish Signal Processing Symp.* (Pori, Finland, 1997), pp. 1–5.

[27] J. Hynninen and N. Zacharov, "GuineaPig—A Generic Subjective Test System for Multichannel Audio," presented at the 106th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 47, p. 513 (1999 June), preprint 4871.

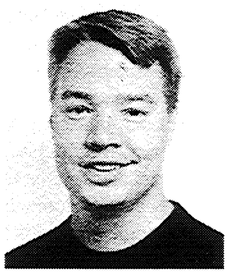
[28] J. P. Guilford, *Psychometric Methods* (McGraw-Hill, New York, 1954).

[29] R. Lehman, *Statistics and Research Design in the Behavioral Sciences* (Wadsworth Publi., Belmont, California, 1991).

[30] T. Tolonen and V. Välimäki, "Automated Parameter Extraction for Plucked-String Synthesis," in *Proc. Inst. Acoust.*, vol. 19, pp. 245–250 (1997).

THE AUTHORS

H. Järveläinen



T. Tolonen

Hanna Järveläinen received the M.Sc.(EE) degree from Helsinki University of Technology (HUT), Finland, in 1997. Since 1996 she has been with the HUT Laboratory of Acoustics and Audio Signal Processing, conducting research in psychoacoustics and perception. Starting with tractor noise and cellular phone buzzer sounds, she has recently moved to the more pleasant world of musical instrument sounds.

•

Tero Tolonen majored in acoustics and audio signal processing and received M.Sc.(EE), Lic.Sc.(Tech), and D.Sc.(Tech) degrees from Helsinki University of Technology (HUT) in 1998, 1999, and 2000, respectively. The topic of his doctoral thesis was object-based sound source modeling. His research interests include audio coding, model-based audio representations, physical modeling of musical instruments, and digital signal processing. From 1996 to 2000 he conducted research at Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing. Currently, Dr. Tolonen is with Luxxon Corporation, Mountain View, CA, USA.

