Restoration and Enhancement of Solo Guitar Recordings Based on Sound Source Modeling*

PAULO A. A. ESQUEF,[1] VESA VÄLIMÄKI,[2,1] *AES Member*, AND MATTI KARJALAINEN,[1] *AES Fellow*

[1]*Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, FIN-02015 HUT, Espoo, Finland*

[2]*Tampere University of Technology, Pori School of Technology and Economics, FIN-28101 Pori, Finland*

New propositions to audio restoration and enhancement based on sound source modeling are presented. A case based on the commuted waveguide synthesis algorithm for plucked-string tones is described. The main motivation is to take advantage of prior information of generative models of sound sources when restoring or enhancing musical signals.

0 INTRODUCTION

Audio enhancement is a wide concept and is closely related to audio restoration. An intuitive idea of audio enhancement is associated with any audio processing that is able to improve the perceptual quality of an audio signal. The goal of the digital audio restoration field [1] is, ideally, to improve the quality of audio signals extracted from old recordings, such as wax cylinders, 78 rpm, long-playing records, magnetic tape, and even digital media matrices. The usual approach consists of finding the best way to capture and transfer the recorded sound from the original matrices to a digital medium and, after that, applying digital signal-processing techniques to remove any disturbance or noise produced by the recording and reproducing system.

The most common tasks of audio restoration algorithms are to remove impulsive noise and reduce broad-band noise from the degraded audio sources. Whereas localized disturbances, at least those of short duration, are relatively easy to treat, dealing with global types of degradation is still a challenging task. In particular, in the broad-band noise-reduction problem the goal is to find better tradeoffs between effective noise reduction and signal distortion [1], [2]. Although the perceptual quality of the restored signals plays an important role in this matter, only recently have psychoacoustic criteria been proposed for audio enhancement purposes [3], [4], still bounded by the lack of an observable clean reference signal.

Usually audio restoration algorithms employ signal modeling techniques, which deal with the information available in the surface presentation of audio signals, that is, the attempt to model the waveform representation of the audio signal. In sound source modeling (SSM) techniques, however, the goal is to model the phenomenon that has generated the waveform. As a natural consequence, a structured audio representation [5] is required in SSM.

In addition to SSM, models for the propagation medium and the receptor characteristics have been increasingly employed in audio signal processing. In [6] a general framework for audio and musical signal processing is described. It shows the hierarchical scales and relationships among several levels of audio representations. In fact, actual challenging audio signal-processing applications seem to move toward the incorporation of higher representation levels of audio signals, such as the object- and content-based ones. Among those applications it is possible to cite sound source recognition [7], sound source separation [8], music retrieval, automatic transcription of music [9], object-based sound source modeling [10], and sound synthesis [11].

Due to the requirement of a structured audio representation when using SSM, its practical use for the analysis and synthesis of audio signals is still limited to specific cases. It is easy to see that for general cases, such as analysis and synthesis of polyphonic music, the SSM-based system faces difficult tasks. The analysis part requires

* An earlier version of this paper was presented at the 110th Convention of the Audio Engineering Society, Amsterdam, The Netherlands, 2001 May 12–15, under the title, “Restoration and Enhancement of Instrumental Recordings Based on Sound Source Modeling.” Manuscript received 2001 June 25; revised 2001 December 10.

instrument recognition, sound source separation, detection of musical events, and extraction of their features, among others. For the synthesis part it is required that the types of instruments present in the signal be known and synthesis models be available for them. Furthermore one must take into account that the signal analysis performance may be degraded when dealing with real recordings, which may contain spurious background noise, nonlinear distortions, and reverberation.

Nevertheless, SSM has been employed in more restrained situations. For instance, in [10] an object-based SSM system for analysis and synthesis of the acoustic guitar is presented. The system, which deals with two-voice polyphony examples, is able to analyze the signal, isolate the voices, and recreate the signal again using a guitar synthesizer. The analysis part involves signal modeling techniques, such as sinusoidal modeling [12]–[14], combined with auditory modeling to pitch determination and signal separation. The synthesis part employs the physical modeling approach, for example, the digital waveguide method [15], which has been used successfully to synthesis realistic instrument sounds by taking into account physical properties associated with the instruments and their particular playing techniques [16].

A similar approach as that adopted in [10] can be used for audio restoration purposes if both the analysis and the synthesis parts can be made robust to the presence of noise in the signal. In addition SSM allows taking advantage of previous knowledge of the model parameters associated with a high-quality instrument sound. This information can be useful when attempting to enhance the sound quality of a poorly recorded instrument. For instance, in this paper it is shown that it is possible to reconstruct missing high-frequency harmonics of guitar tones, since high-quality synthesis models for plucked-string tones are available, providing prior knowledge of what would be their frequency content.

Based on the considerations previously outlined one must note that the restoration or enhancement of audio signals within the SSM framework is still restricted to simple cases. Therefore in this paper, only single and isolated acoustic guitar notes are considered. This choice obviously simplifies both the analysis and the synthesis stages involved in the method. For the SSM of plucked strings, the computed waveguide synthesis (CWS) algorithm [17], [18] is used. This choice allows obtaining the model parameters by analyzing recorded tones [19]. The study presented here is divided basically into two parts: a proposition to extend the bandwidth of originally band-limited guitar tones, and a denoising scheme for guitar tones which mixes a traditional spectral-based dehissing method and SSM.

The paper is organized as follows. In Section 1 the CWS algorithm for plucked-string instruments used in this work is reviewed. The SSM-based method to extend the bandwidth of guitar tones is proposed in Section 2. In Section 3 the dehissing of guitar tones is discussed. Experimental results are described in Section 4. Discussion, conclusions, and directions to future works are given in Section 5.

1 CWS METHOD FOR PLUCKED-STRING INSTRUMENTS

Physical modeling techniques for digital sound synthesis of musical instruments have become a popular approach in recent years. In particular, the digital waveguide synthesis, first introduced by Smith [15], and its further improvements and extensions [20] have proved to be well suited to high-quality synthesis of string instruments.

In the case of plucked-string instruments, a natural structure of a physical-based synthesizer system would consist of an impulsive excitation signal as the input of a plucking event model cascaded with a string model and a body model of the instrument. If the plucking, string, and body models are considered linear and time-invariant systems, it is possible to commute them and combine the plucking and body responses into only one input signal. This is the basic principle of the CWS method [17], [18] for plucked strings. For more detailed information on the development of a guitar synthesizer based on the CWS method, see [21].

1.1 String Model

The function of the vibrating string model is to simulate the generation of string modes after the plucking event. Considering an isolated string, its behavior can be efficiently simulated by the string model illustrated in Fig. 1, whose transfer function is given by

$$S(z) = \frac{1}{1 - z^{-L_i} F(z) H(z)} \tag{1}$$

where z^{-L_i} and $F(z)$ are, respectively, the integer and fractional parts of the delay line associated with the length of the string L. Transfer function $H(z)$ is called the loop filter, and it is in charge of simulating the frequency-dependent losses of the harmonic modes.

In this work the loop filter is implemented as a one-pole low-pass filter with the transfer function given by

$$H(z) = g \, \frac{1 + a}{1 + az^{-1}} \, . \tag{2}$$

The magnitude response of $H(z)$ must not exceed unity in order to guarantee the stability of $S(z)$. This constraint imposes that $0 < g < 1$ and $-1 < a < 0$.

The presence of the fractional delay filter $F(z)$ is intended to provide a fine-tuning of the fundamental frequency by precisely adjusting the length of the string. Here it is implemented as a fourth-order Lagrange interpolator FIR filter [22]. In this configuration the string-model transfer function $S(z)$ is completely defined by the length of the loop delay L and the loop filter parameters

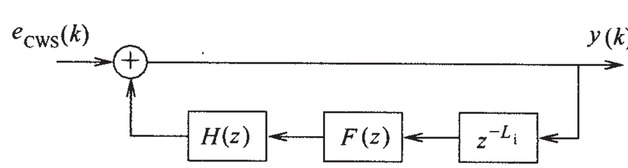

Fig. 1. Block diagram of string model.

g and a. In fact, these string-model parameters depend on the fundamental frequency and the fret number. Therefore they must be estimated for each tone to be synthesized.

1.2 Estimation of String-Model Parameters

In this section the estimation of the string parameters in the CWS model is discussed. The estimation procedure can be performed automatically by analyzing recorded tones, as shown in [19], [23].

The first step consists in estimating the fundamental frequency of the tone, for instance, through the autocorrelation method. In this case the analysis is performed over a signal excerpt taken after the attack part of the tone, since the value of the fundamental frequency of plucked-string tones takes some time to stabilize after the plucking instant. Then, given an estimate of the fundamental frequency $\hat{f}_0$ the length of the delay line in samples is obtained as

$$L = \frac{f_s}{\hat{f}_0} \tag{3}$$

where f_s is the sampling rate of the analyzed signal.

The next step consists in estimating the loop filter parameters. The magnitude response of the loop filter actually defines how the energy of the vibrating string modes decays as a function of time. However, the string model defined by $S(z)$ can only simulate the exponentially decaying behavior of the ideal string modes. In this case the time constants of the decaying exponentials have a direct relationship with the magnitude response of the loop filter.

The estimation of the loop filter parameters is carried out in three basic steps. First the decaying envelope of each harmonic is obtained through a pitch-synchronized STFT analysis, followed by a magnitude peak picking algorithm. Then linear curves are fitted to the envelopes on a logarithmic scale. The resulting set of slopes defines what would be the values of the loop gains [or the magnitude of $H(z)$] at the harmonic frequencies. Finally $H(z)$ is designed via a weighted least-squares procedure in which the error between its magnitude and the previously estimated values of the loop gains is minimized. A detailed description of the procedures used to estimate the string-model parameters is found in [19].

2 BANDWIDTH EXTENSION OF GUITAR TONES

In this section the problem of reconstructing missing spectral information in guitar tones is addressed within the SSM approach. The connections between bandwidth extension and audio restoration appear in two cases: to overcome the intrinsic bandwidth limitations of old recording systems in capturing the audio source and, even more interestingly, to reconstruct the spectral information lost during a denoising procedure. The latter case will be discussed in Section 3.

The test signal used in this study is a single guitar tone which was low-pass filtered in order to remove the high-frequency harmonics while preserving the fundamental frequency as well as a few harmonics.

The first step of the bandwidth extension procedure is to estimate the string-model parameters as described in Section 1.

Estimating the fundamental frequency is not problematic, assuming that it was preserved in the low-pass-filtered tone. The loop filter design is more critical, since there are no harmonics available above a certain frequency to have their decay rate estimated. Nevertheless, considering the simplicity of the loop filter employed in the string model and the fact that for this type of string model, variations between 25 and 40% in the time constant of the decay are not perceived [24], it is acceptable to estimate the decay rate of the missing harmonics by analyzing a similar full-band guitar tone.

As seen in Section 1, the string model is basically a comb filter tuned at the fundamental frequency and its harmonics. Thus the main effect of inverse filtering the guitar tone through the string model $S(z)$ is to attenuate the string modes. The resulting excitation $e_{CWS}(k)$ usually has a large number of resonances associated with all other information except that of the vibrating string, such as nonlinearities associated with the plucking event, body resonances, and coupling between strings.

In the bandwidth extension problem, the analyzed signal is already low-pass filtered, resulting in an excitation with a low-pass characteristic as well. Therefore it is only able to excite the string modes corresponding to the harmonic frequencies originally present in the analyzed signal. However, if an extra amount of energy is added to the excitation in a proper way, it is possible to excite all the modes of the string model. Thus by altering the excitation signal it is possible to resynthesize a new tone whose bandwidth is greater than that of the analyzed one.

Typically the frequency response of acoustic guitar bodies exhibits a few slowly decaying resonance modes in the low-frequency range [25]. Toward higher frequencies the number of resonance modes increases, but their decay time decreases. This characteristic motivates the use of exponentially decaying white noise to efficiently model the high-frequency response of guitar bodies [26].

In this sense a possible strategy to fulfill the information that is missing in the low-pass-filtered excitation would consist of adding an artificially generated noise burst $e_{pluck}(k)$ directly to the string model, triggered with the attack part of $e_{CWS}(k)$, as illustrated in Fig. 2. This noise burst, which can be considered either a rough model for the missing high-frequency modes of the guitar body or an extra plucking event, must have enough energy within the entire frequency range in order to fully excite

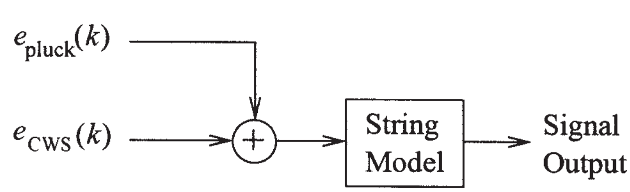

Fig. 2. Bandwidth extension scheme.

the string model.

If the artificial plucking can really excite the string modes, the resynthesized tone will exhibit harmonics in the full frequency range, although the decay rates of the previously nonexistent modes will be defined only by the string-model characteristics.

The simplest choice for the artificially generated plucking signal would be an impulse. However, it is known that the finger–string interaction is not really impulsive, and a better option is to generate an impulsive noise burst, for instance, by windowing a zero-mean Gaussian white-noise sequence. A noise burst of about 10 ms seems a reasonable choice to simulate the duration of a typical finger–string interaction.

It is important to notice that the power spectrum density of the noise burst described before is flat and thus will excite almost equally all the string modes. However, it would be desired that the additional noise burst, composed with the filtered excitation, could emulate a typical spectral behavior of the attack part of an excitation corresponding to a full-bandwidth tone. A simple option to realize that is to color the noise burst in a proper way, for instance, according to known information about typical spectral characteristics of guitar bodies. Alternatively, one can obtain this information through the excitation of a full-bandwidth tone, for instance, by estimating the spectral envelope of its attack part.

In addition it would be desirable to leave the harmonics originally present in the low-pass-filtered tone undistorted. However, in real cases it is not a trivial task, since no previous information about the bandwidth limitation of the analyzed tone is available. If it can be roughly inferred, an arbitrary attenuation in the spectrum of the noise burst can be included to compensate for the unnecessary extra energy within the original bandwidth.

The generation of the noise burst, which simulates a plucking event, can be carried out as depicted in Fig. 3. The input sequence $n(k)$ is a zero-mean Gaussian white-noise sequence, the filter $E(z)$ is a coloring filter the magnitude response of which must approximate the spectral envelope of the very beginning of a full-bandwidth excitation. The high-pass filter $H_{hp}(z)$ is optional and can be included to compensate for the unnecessary addition of energy within the effective bandwidth of the analyzed tone. The gain factor α controls the local signal-to-noise ratio (SNR) at the part of the excitation to be modified.

Naturally, the noise burst is windowed before its addition to the excitation. In this context, the characteristics of the synthetic pluck depend on its length in samples, the magnitude response of the filters $E(z)$ and $H_{hp}(z)$, and the value of the gain factor. Further details concerning the choices of the filters $E(z)$ and $H_{hp}(z)$, the length of the noise burst, and the value of gain α are given in Section 4.1.

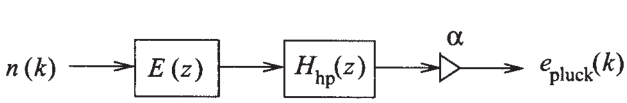

Fig. 3. Generation of synthetic plucking event.

3 SSM AND DENOISING OF GUITAR TONES

In this section a processing scheme that involves the SSM of guitar tones and traditional methods of audio restoration is proposed to improve the perceptual quality of dehissed guitar tones.

The basic principle behind spectral-based audio dehissing methods is to split the noisy signal into a certain number of frequency bands and to attenuate the signal on those bands where the SNR is below a given value [27], [28]. The spectral analysis and the corresponding signal reconstruction can be realized via either filterbanks or short-time Fourier transform. Usually dehissing methods suffer from a difficult tradeoff between the reduction of the noise effects and the distortion of the signal to be restored [2], [1], [27]. As both the noise and the signal share the same spectral range, any attempt to have the noise reduced leads to a degradation of the signal information to some extent.

One of the most common audio dehissing methods is based on digital Wiener filtering [1], in which the signal is segmented in short-time frames and the magnitude spectrum of each frame is weighted according to local estimates of the SNR at each frequency bin. The lower the SNR at a certain bin, the more attenuated is its magnitude value. When dehissing acoustical musical signals, signal losses are more prominent at high frequencies, since lower SNRs are observed in this range. Naturally, auditory properties such as masking and critical bands help to explain the improved perceptual quality of dehissed audio signals. However, the lack of a reference signal prevents the appealing use of auditory-based approaches as well as the employment of either purely objective or perceptual-based measures to evaluate the perceptual quality of the dehissed signals.

The results of bandwidth extension, described in Section 2, provide a useful appeal to the dehissing problem. The hard tradeoff between the preservation of the valuable signal information and the noise reduction can be softened on the grounds that the signal information can be reconstructed afterward if a sound source model is available for the signal. This possibility poses an alternative view on the dehissing problem as discussed in the following sections.

3.1 Aggressive Dehissing and Bandwidth Extension

The possibility of reconstructing lost frequency information of guitar tones, as described in Section 2, can serve the dehissing problem in the following way. A spectral-based dehissing method with an overestimated value for noise variance can be employed to perform an aggressive type of denoising. This choice will surely reduce the perceptual effects of the residual noise, but it will lead to an oversmoothed restored signal. A remedy to the oversmoothing problem is to apply a postprocessing stage such as the SSM-based bandwidth extension to recover the signal information that was lost due to the aggressive dehissing procedure.

In this case the estimation of the string-model parameters faces similar problems as those discussed in Section 2, when dealing with band-limited guitar tones. Here the high-frequency harmonics are either masked by the corrupting noise or absent due to the aggressive dehissing procedure. Thus their decay rate estimation is prevented. Anyway, the same considerations as in Section 2 regarding the estimation of the loop filter parameters are applicable in this case. Further details on the implementation of the previously described approach are given in Section 4.2.

3.2 Integrated Dehissing and Bandwidth Extension

Another strategy for the dehissing problem consists of integrating the dehissing and signal reconstruction procedures into a single stage. This can be achieved by adapting the dehissing method to process the excitation corresponding to the noisy signal. In fact, in the noisy excitation the energy at the harmonic frequencies is attenuated and the corrupting noise is colored. Nevertheless, a spectral-based dehissing method is still applicable to the excitation signal, since it has important resonances associated with the body modes. Of course, an aggressive dehissing procedure will lead to losses mainly in the high-frequency content of the excitation signal. As was seen in Section 2, the attack part of the excitation has an important role in the reconstruction of the frequency information of the tone. In this context, if only the attack part of the excitation is spared from the aggressive dehissing procedure, it will provide enough energy to properly excite the string model in order to resynthesize a non-smoothed and noise-free tone.

A possible way to protect the attack part of the excitation from the aggressive dehissing is to control the noise variance estimation used in the spectral-based dehissing method artificially. For instance, a gain can be assigned to the noise variance estimate in such a way that it is set to a high value elsewhere except at the attack part, where the value of the gain should be set to unity.

Considering that the highest local variance of the excitation is observed during its attack part, an automatic procedure can be devised within the frame-by-frame dehissing procedure. First a local estimate of the excitation variance at the attack part, s^2_{attack}, should be obtained. Then for each frame index i, the estimate of the noise variance is multiplied by a gain given by

$$m = \min\left(m_{\max}, \frac{s^2_{\text{attack}}}{s^2 + \epsilon}\right) \tag{4}$$

where s^2 is a locally estimated excitation variance within a given frame i, ϵ is a small positive value to prevent a division by zero, and $m_{\max}$ is a constant value which represents the maximum value of m allowed. Since for most frames but those associated with the attack part of the excitation $s^2 << s^2_{\text{attack}}$, their ratio will assume higher values than $m_{\max}$, implying $m = m_{\max}$. Further details on the implementation of the integrated dehissing procedure are described in Section 4.2.

4 EXPERIMENTAL RESULTS

In this section experimental results in the bandwidth extension and the dehissing of single guitar tones are described. The test signal used in both cases is an F_4 tone with a fundamental frequency of 347 Hz. The tone was recorded in an anechoic chamber and sampled at 22.05 kHz.

4.1 Bandwidth Extension

The test signal used in the bandwidth extension experiments was low-pass filtered using a 101th-order equiripple FIR filter with a cutoff frequency at 1 kHz, a transition band of 1 kHz, and an attenuation of 80 dB on the rejection band. In this case, regardless of the filtering procedure, the fundamental and the next two harmonic frequencies of the tone were preserved.

In the experiment described here, the string-model parameters were estimated using the original tone. This choice was taken as an attempt to isolate the problems associated with the estimation of the model parameters and the bandwidth extension procedure. In the following step, the excitation corresponding to the low-pass-filtered tone was obtained by inverse filtering.

The artificially generated plucking event $e_{\text{pluck}}(k)$ was obtained as shown in Fig. 3. In this experiment $n(k)$ was chosen as a zero mean white Gaussian noise sequence, and $E(z)$ as a second-order resonator tuned at 200 Hz. This frequency corresponds to the lowest mode of the top plate of the guitar body [25]. The radius of the poles was arbitrarily set to 0.8. It can be seen from Fig. 4 that with these parameters the magnitude response of $E(z)$ approximates quite well the spectral envelope associated with the attack part of a full-bandwidth excitation.

The high-pass filter H_{hp} was not included to keep the generality of the method, since the bandwidth limitation of the analyzed tone is usually not known beforehand. Finally the noise burst was then multiplied by a Hanning window of 600 samples, scaled, and added to the attack part of the excitation. The procedure was automated by detecting the attack part of the excitation using a magnitude criterion and then synchronizing both the attack and the window maxima, as shown in Fig. 5.

It should be noted that the noise burst can be fully characterized by the coloring filter and the length of the win-

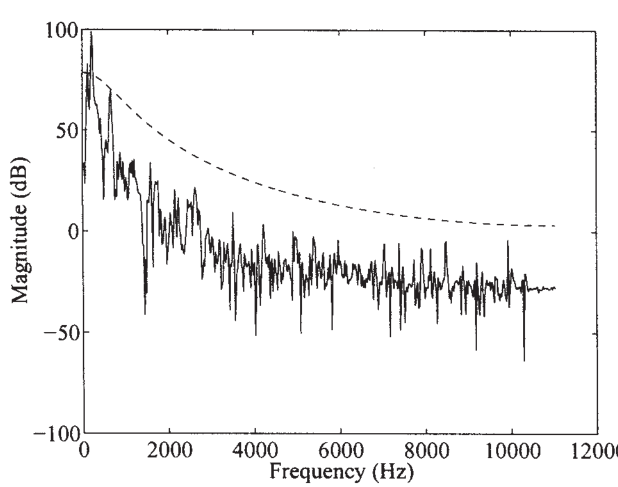

Fig. 4. Squared magnitude response of $E(z)$ (– – –) and power spectrum of attack part of a full-bandwidth excitation.

dow. While the latter was chosen according to an estimate of the duration of finger–string interaction, the former was designed by considering known features associated with the guitar body characteristics.

Based on informal listening tests, it was observed that coloring the noise burst has an important effect on the quality of the timbre of the resynthesized tone. The timbre of the resynthesized tone also varies depending on the power of the noise burst, which can be adjusted to produce a certain local SNR at the attack part of the excitation.

For instance, in this experiment, if the SNR is set to 40 dB, the resynthesized tone does not exhibit great perceptual differences compared to the low-pass-filtered tone. Reducing the value of the SNR tends to increase the perceptual differences and emphasize the plucking event. An SNR of about 20 dB was found to be a suitable value to achieve a resynthesized tone with perceptual quality close to that of the original one. On the other hand, lower values of SNR, such as 10 dB, overemphasize the plucking event, compromising the perceptual quality of the tone.

The capability of the method to extend the bandwidth of guitar tones is illustrated in Fig. 6, which shows time–frequency analysis plots of the original, the low-pass-filtered, and the resynthesized tones. In this case the noise burst was generated as described before and scaled to produce an SNR of 20 dB at the attack part of the excitation.

Additional tests were performed on the test guitar tone in which its bandwidth was limited to 500 Hz and 3000 Hz. The results obtained were similar to those of the previous case. However, the perceptual differences between the original and the 3000-Hz band-limited tone are already less prominent. Therefore the effects of the bandwidth extension procedure are more difficult to perceive.

4.2 Dehissing

In the dehissing experiments a zero mean Gaussian white noise was added to the test guitar tone signal, and its variance was adjusted to generate a global SNR of 20 dB.

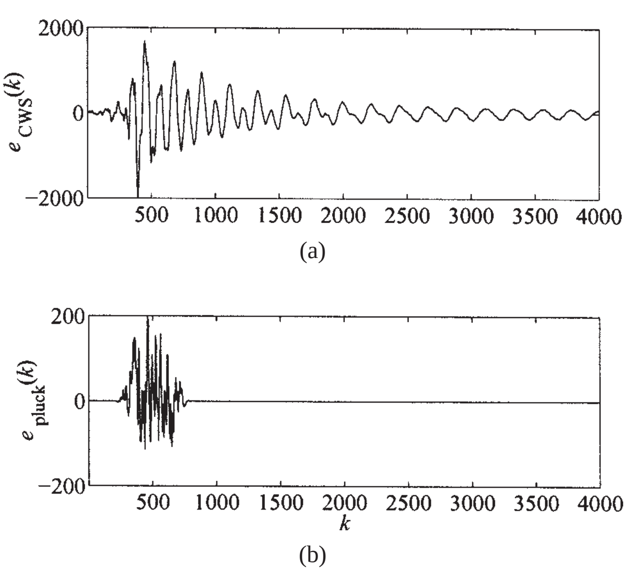

Fig. 5. (a) Attack part of excitation. (b) Synchronization of noise burst.

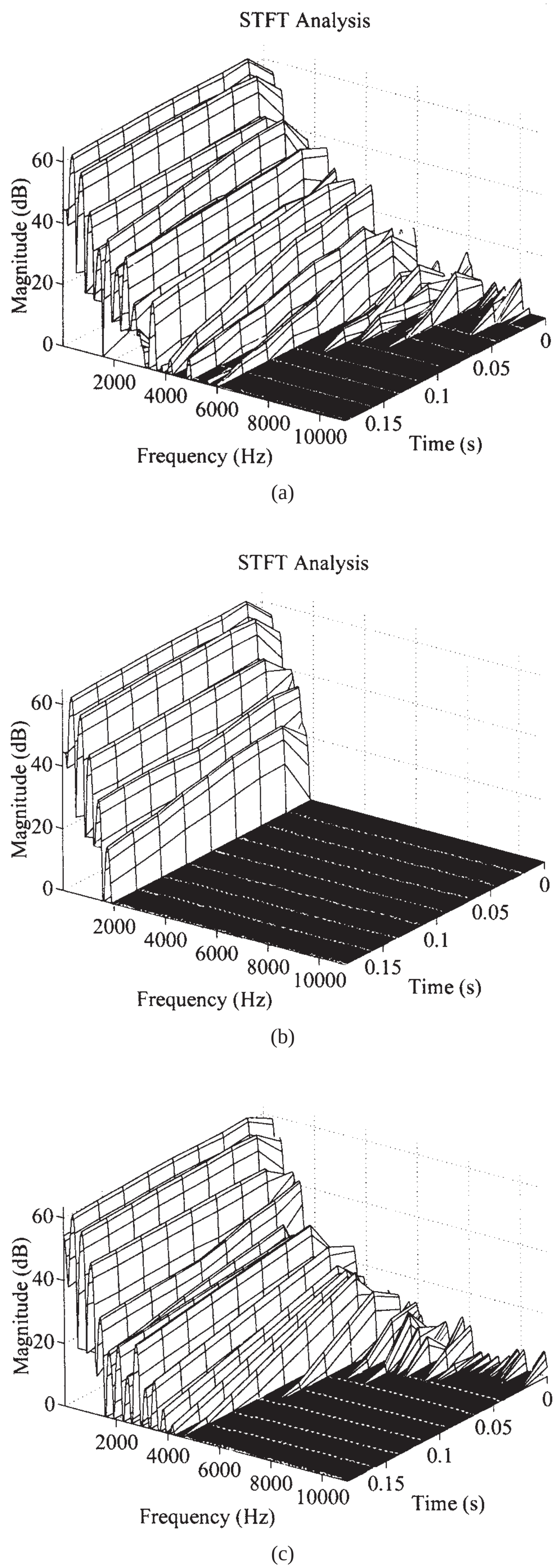

Fig. 6. Time–frequency analysis. (a) Original tone. (b) Low-pass-filtered tone. (c) Resynthesized tone.

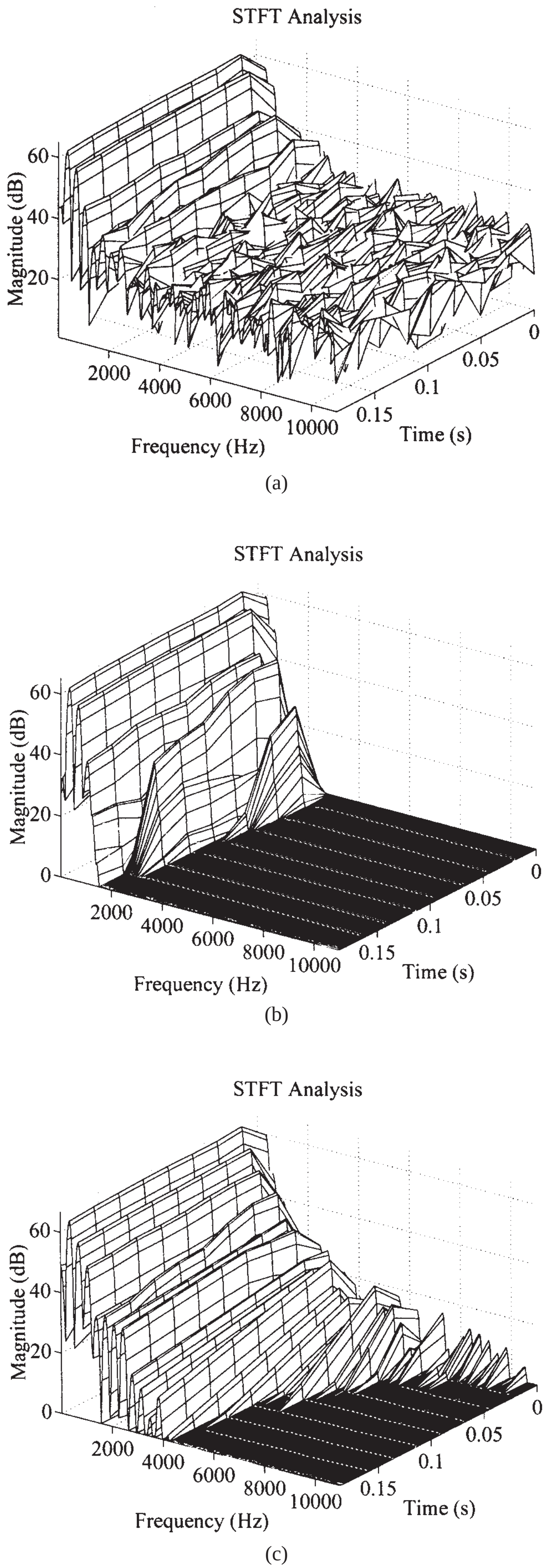

Fig. 7. Time–frequency analysis. (a) Noisy tone. (b) Dehissed tone. (c) Bandwidth-extended tone.

The first step of the dehissing and postprocessing approach consisted of dehissing the noisy signal through a Wiener filtering scheme, as described in [1]. In this experiment signal frames of 256 samples were used with an overlap of 50%. The noise variance was estimated in the frequency domain by taking the mean value of the upper quarter of the power spectrum. In addition a gain was assigned to the noise variance estimate. This gain, which hereafter will be called noise floor gain, worked as a control parameter for the amount of noise to be removed.

Since a single guitar tone is not a complex signal, it does not help in masking the residual noise effects in the restored tone, mainly after its attack part. However, they can be reduced by overestimating the noise variance within the Wiener filtering scheme. Considering the Wiener filter configuration and the test signal used in this experiment, it was found that a noise floor gain of 30 suffices to almost eliminate the residual noise effects in the restored signal despite its strongly smoothed characteristic.

The last step consisted of applying the bandwidth extension procedure to the aggressively dehissed tone in order to reconstruct the lost harmonic frequencies. In this experiment good results were attained by employing the same approach and parameters that were used in Section 4.1.

In Fig. 7 time–frequency analysis plots of the noisy, aggressively dehissed, and bandwidth-extended signals are shown. As can be seen from Fig. 7(a), the noise masks the high–frequency harmonics, which together with the noise are also removed after the aggressive dehissing procedure [Fig. 7(b)]. Nevertheless, the SSM-based bandwidth extension scheme is able to reconstruct the missing harmonics in the resynthesized tone [Fig. 7(c)]. In this case the timbre of the restored tone is similar to the original one, and the effects of the residual noise are greatly reduced.

The experiment employing the integrated dehissing and reconstruction approach was performed on the artificially corrupted tone described before. The procedure was carried out by first estimating the string-model parameters and then obtaining the noisy excitation through inverse filtering. The noisy excitation was dehissed via the Wiener filtering method, here adapted to account for the colored noise at the excitation as well as for the application of a varying gain to the noise floor estimate (see Section 3.2).

As in the previous dehissing experiment, signal frames of 256 samples were used with an overlap of 50%. The estimates for the noise floor within each signal frame were obtained in the same way. In addition the noise floor estimates were multiplied by gain factors η defined in Eq. (4), which were also computed for each frame.

The computation of η involved the following choices. The value of σ^2_{attack} was set as the power of the attack part of the noisy excitation. The value of η_{max} was chosen as 50. The value of σ^2 was set as the power of the noisy excitation within a given frame; therefore it is the only parameter that varies as a function of the frame index. Since the presence of the noise prevents σ^2 to assume a null value, the value of ϵ was chosen as zero.

In this case the sequence of values of η as a function of the frame index looks like the one shown in Fig. 8(a). Note that outside the attack part of the excitation, which starts at frame 30, $\mu = \mu_{max}$. Thus an aggressive dehissing is performed on the whole excitation except at its beginning. For the sake of simplicity, the values of μ are shown for the first 100 signal frames, which corresponds to approximately 0.6 s in time.

The plot in Fig. 8(b) shows a time–frequency analysis of the resynthesized signal obtained from the previously dehissed excitation. As can be seen, the procedure is capable of removing the noise and reconstructing the lost harmonics. However, in this case the spectral tilt associated with the attack part of the excitation is determined by that of the noisy excitation. Therefore an undesirable positive bias on the powers of the high-frequency harmonics is observed. This reduces the perceptual quality of the attack part of the resynthesized tone, which has a more synthetic quality than the restored tone obtained in the previous experiment.[1]

5 DISCUSSION AND CONCLUSIONS

In this paper the enhancement of guitar tones was presented within a sound source modeling framework. First it was shown how the reconstruction of spectral information in guitar tones can be attained by means of SSM techniques. Then that issue was taken into account in a dehissing scheme, which mixed a traditional spectral-based method with SSM. The results obtained for both the bandwidth extension and the dehissing experiments demonstrate that the proposed schemes are effective in improving the perceptual quality of the restored tones.

Although showing some potential, the use of SSM for audio enhancement purposes is still restricted to special cases. For instance, when attempting to restore severely degraded instrumental recordings addressing one single instrument, the most prominent music events, such as the melodic lines, should be detected and isolated. Further, their features could be used to calibrate a synthesizer in order to reconstruct another signal, which would sound similar to the original source but free of noise. The choice of a synthesizer based on physical modeling would provide more flexibility for adjusting the model parameters according to the extracted features of the music events as well as the possibility of taking advantage of other available information about high-quality sound sources.

It is important to note that the enhancement of audio signals using SSM is still restricted to simple cases. For instance, restoration of solo guitar within the SSM framework would in itself imply challenging tasks due to the structural representation required for the musical content. As a consequence, one needs effective and robust methods to detect and locate the occurrence of musical events, to separate notes or chords whose content overlaps both in time and frequency, and to extract musical features associated with the sound events from the observable sound waveform.

Extensions to more general cases can be viewed as a multilayered problem in which even more demanding tasks would be required, such as recognition and separation of more general musical elements in complex sound source mixtures. In addition it is plausible to expect that the signal analysis performance decreases when dealing with real recorded sounds. This is due to the possible presence of spurious noises, nonlinear distortions, and even strong reverberation in the signal to be analyzed. On the synthesis side of the chain, the requirements are related to the development of model-based music synthesizers with more realistic sounds, and capable of simulating the playing features of real performances.

SSM- and content-based audio processing is still in a youthful stage of development. However, as long as it develops into better ways to represent and recreate sound sources, performing audio enhancement within the SSM framework can lead to better results compared to those attained by traditional techniques.

[1] Sound examples are available at URL: http://www.acoustics.hut.fi/publications/papers/jaes-ssm/.

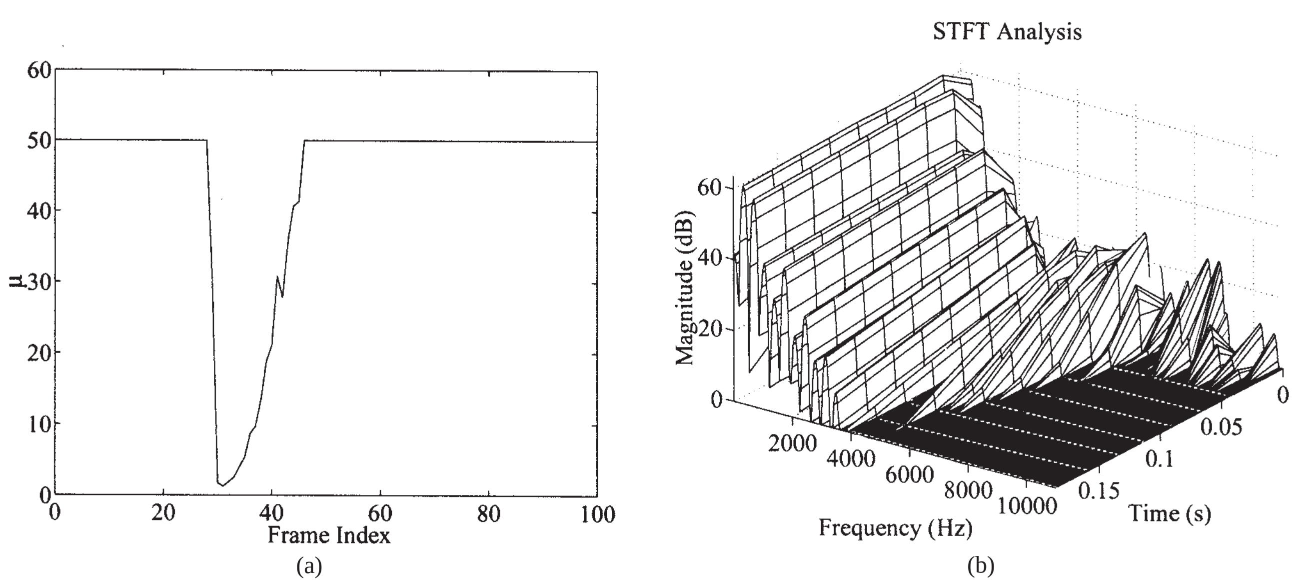

Fig. 8. (a) Values of noise floor gain as a function of frame index. (b) Time–frequency analysis of restored tone.

6 ACKNOWLEDGMENT

The work of P. Esquef has been supported by a scholarship from the Brazilian National Council for Scientific and Technological Development (CNPq-Brazil) and the Sound Source Modeling Project of the Academy of Finland. V. Välimäki has been financed by a postdoctoral research grant from the Academy of Finland.

7 REFERENCES

[1] S. J. Godsill and P. J. W. Rayner, *Digital Audio Restoration—A Statistical Model Based Approach* (Springer, London, 1998).

[2] O. Cappé, "Elimination of the Musical Noise Phenomenon with the Ephraim and Malah Noise Suppressor," *IEEE Trans. Speech Audio Process.*, vol. 2, pp. 345–349 (1994 Apr.).

[3] P. J. Wolfe and S. J. Godsill, "The Application of Psychoacoustic Criteria to the Restoration of Musical Recordings," presented at the 108th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 48, p. 363 (2000 Apr.), preprint 5150.

[4] D. Tsoukalas, J. Mourjopoulos, and G. Kokkinakis, "Perceptual Filters for Audio Signal Enhancement," *J. Audio Eng. Soc.*, vol. 45, pp. 22–36 (1997 Jan./Feb.).

[5] B. L. Vercoe, W. G. Gardner, and E. D. Scheirer, "Structured Audio: Creation, Transmission, and Rendering of Parametric Sound Representations," *Proc. IEEE*, vol. 86, pp. 922–940 (1998 May).

[6] M. Karjalainen, "Immersion and Content—A Framework for Audio Research," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* (New Paltz, NY, USA, 1999 Oct.), pp. 71–74.

[7] J. C. Brown, "Computer Identification of Musical Instruments Using Pattern Recognition with Cepstral Coefficients as Features," *J. Acoust. Soc. Am.*, vol. 105, pp. 1933–1941 (1999 Mar.).

[8] R. C. Maher, "Evaluation of a Method for Separating Digitized Duet Signals," *J. Audio Eng. Soc.*, vol. 38, pp. 956–979 (1990 Dec.).

[9] A. Klapuri, "Automatic Transcription of Music," Master's thesis, Tampere University of Technology, Tampere, Finland (1998). Electronic version available at URL: http://www.cs.tut.fi/~klap/iiro/.

[10] T. Tolonen, "Object-Based Sound Source Modeling for Musical Signals," presented at the 109th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 48, p. 1094 (2000 Nov.), preprint 5174.

[11] T. Tolonen, V. Välimäki, and M. Karjalainen, "Evaluation of Modern Sound Synthesis Methods," Tech. Rep. 48, Helsinki University of Technology, Lab. of Acoustics and Audio Signal Processing, Espoo, Finland (1998). Postscript version available at URL: http://www.acoustics.hut.fi/publications/.

[12] T. F. Quatieri and R. J. McAulay, "Audio Signal Processing Based on Sinusoidal Analysis and Synthesis," in *Applications of Digital Signal Processing to Audio and Acoustics*, 1st ed., M. Kahrs and K. Brandenburg, Eds. (Kluwer, Boston, MA, 1998), pp. 314–416.

[13] J. O. Smith and X. Serra, "PARSHL: An Analysis/Synthesis Program for Non-Harmonic Sounds Based on a Sinusoidal Representation," in *Proc. Int. Computer Music Conf.* (Champaign-Urbana, 1987).

[14] R. C. Maher, "Sinewave Additive Synthesis Revisited," presented at the 91st Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 39, p. 996 (1991 Dec.), preprint 3128.

[15] J. O. Smith, "Physical Modeling Using Digital Waveguides," *Computer Music J.*, vol. 16, no. 4, pp. 74–91 (1991).

[16] M. Laurson, C. Erkut, V. Välimäki, and M. Kuuskankare, "Methods for Modeling Realistic Playing in Acoustic Guitar Synthesis," *Computer Music J.*, vol. 25, no. 3, pp. 38–49 (2001).

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

[18] 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 Sept.), pp. 56–63.

[19] 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).

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

[21] V. Välimäki and T. Tolonen, "Development and Calibration of a Guitar Synthesizer," *J. Audio Eng. Soc.*, vol. 46, pp. 766–778 (1998 Sept.).

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

[23] T. Tolonen, "Model-Based Analysis and Resynthesis of Acoustic Guitar Tones," Master's thesis, Rep. 46, Helsinki University of Technology, Lab. of Acoustics and Audio Signal Processing, Espoo, Finland (1998 Jan.). Postscript version available at URL: http://www.acoustics.hut.fi/publications/.

[24] H. Järveläinen and T. Tolonen, "Perceptual Tolerances for Decay Parameters in Plucked String Synthesis," *J. Audio Eng. Soc.*, vol. 49, no. 11, pp. 1049–1059 (2001 Nov.).

[25] O. Christensen and B. B. Vistisen, "Simple Model for Low-Frequency Guitar Function," *J. Acoust. Soc. Am.*, vol. 68, pp. 758–766 (1980).

[26] M. Karjalainen, H. Penttinen, and V. Välimäki, "More Acoustic Sounding Timbre from Guitar Pickups," in *Proc. 2nd G-6 Workshop on Digital Audio Effects (DAFX'99)* (Trondheim, Norway, 1999 Dec.), pp. 41–44.

[27] J. A. Moorer and M. Berger, "Linear-Phase Bandsplitting: Theory and Applications," *J. Audio Eng. Soc.*, vol. 34, pp. 143–152 (1986 Mar.).

[28] S. V. Vaseghi and R. Frayling-Cork, "Restoration of Old Gramophone Recordings," *J. Audio Eng. Soc.*, vol. 40, pp. 791–801 (1992 Oct.).

THE AUTHORS

P. A. A. Esquef

V. Välimäki

M. Karjalainen

Paulo A. A. Esquef was born in Brazil in 1973. He received an engineering degree in electrical engineering from Polytechnic School of the Federal University of Rio de Janeiro (UFRJ) in 1997, and an M.Sc. degree in electrical engineering from COPPE-UFRJ in 1999. His master's thesis addressed digital restoration of old recordings. From 1999 to 2000 he worked on research and development of a DSP system for analysis classification of sonar signals as part of a cooperation project between the Signal Processing Laboratory (COPPE-UFRJ) and the Brazilian Navy Research Center (IPqM). Since 2000 he has been with the Laboratory of Acoustics and Audio Signal Processing at Helsinki University of Technology, where he is currently pursuing postgraduate studies. He is a grant holder from CNPq, a Brazilian governmental council for funding research in science and technology. His research interests include digital audio restoration, computational auditory scene analysis, sound synthesis, among others.

Mr. Esquef is an associate member of the IEEE. He likes to play the piano in his spare time.

●

Vesa Välimäki was born in Kuorevesi, Finland, in 1968. He studied acoustics, audio signal processing, and information sciences at Helsinki University of Technology (HUT), Espoo, Finland, and received the master of science in technology, the licentiate of science in technology, and the doctor of science in technology degrees in electrical engineering in 1992, 1994, and 1995, respectively. His doctoral thesis dealt with sound synthesis based on physical modeling. Dr. Välimäki worked at the HUT Laboratory of Acoustics and Audio Signal Processing from 1990 until 2001. In 1996 he spent six months as a postdoctoral research fellow with the University of Westminster, London, UK. He was appointed a docent in audio signal processing at HUT in 1999.

Since 2001 Dr. Välimäki has been professor of signal processing at Pori School of Technology and Economics, Tampere University of Technology, Pori, Finland, where he teaches courses on digital audio, signal processors, and fundamentals of signal processing. He also lectures on digital audio signal processing at HUT and at the Centre for Music and Technology, Sibelius Academy, Helsinki, Finland. His research interests are in the fields of musical signal processing, active noise control, and digital filter design.

Dr. Välimäki is a senior member of the IEEE Signal Processing Society and is a member of the AES, the International Computer Music Association, the Acoustical Society of Finland, and the Finnish Musicological Society. He is papers chair of the AES 22nd International Conference on "Virtual, Synthetic, and Entertainment Audio," 2002 June, Espoo, Finland. His Web site is http://www.pori.tut.fi/~vpvm/. His e-mail address is Vesa.Valimaki@pori.tut.fi.

●

Matti Karjalainen was born in Hankasalmi, Finland, in 1946. He received the M.Sc. and Dr. Tech. degrees in electrical engineering from the Tampere University of Technology in 1970 and 1978, respectively. His doctoral thesis dealt with speech synthesis by rule in Finnish.

From 1980 he was an associate professor and since 1986 he has been a full professor in acoustics and audio signal processing at the Helsinki University of Technology on the faculty of electrical engineering. In audio technology his interest is in audio signal processing, such as DSP for sound reproduction, perceptually based signal processing, as well as music DSP and sound synthesis. In addition to audio DSP his research activities cover speech synthesis, analysis, and recognition, perceptual auditory modeling, spatial hearing, DSP hardware, software, and programming environments, as well as various branches of acoustics, including musical acoustics and modeling of musical instruments. He has written 250 scientific and engineering articles and contributed to organizing several conferences and workshops. In 1999 he served as the papers chair of the AES 16th International Conference on Spatial Sound Reproduction.

Dr. Karjalainen is an AES fellow and a member of the Institute of Electrical and Electronics Engineers, the Acoustical Society of America, the European Acoustics Association, the International Computer Music Association, the European Speech Communication Association, and several Finnish scientific and engineering societies.