Audio Engineering Society
Convention Paper 5331
Presented at the 110th Convention
2001 May 12–15 Amsterdam, The Netherlands

Restoration and Enhancement of Instrumental Recordings Based on Sound Source Modeling

Paulo Esquef, Vesa Välimäki, and Matti Karjalainen
Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing
FIN-02015 HUT, Espoo, Finland

ABSTRACT

This paper presents new propositions to audio restoration and enhancement based on Sound Source Modeling. We describe a case based on the commuted waveguide synthesis algorithm for plucked string tones. 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 which 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, e.g., wax cylinders, 78 rpm, long-plays, 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/reproducing system.

Traditional audio processing techniques have been used to improve the sound quality of old recordings. While localized disturbances, at least those of very short duration, are relatively easy to treat, dealing with global types of degradation is still a challenging task. Particularly, in the broadband 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 psychoacoustic criteria have been proposed to audio enhancement purposes [3, 4], still bounded by the lack of an observable clean reference signal.

Usually, audio restoration techniques only deal with the information available in the surface presentation of audio signals, i.e., only parametric representations of the signal samples are considered in the analysis and synthesis stages.

However, audio analysis and synthesis can also consider how the sound elements are structured in the audio signal [5]. Of course, it requires better understanding of the human auditory perception, as well as deeper representations of sound sources. In [6], a general framework for audio and musical signal processing is described. It shows the hierarchical scales and relationships among the several levels of audio representations, which, in fact, seem to be required for the actual challenges of the audio signal processing field, such as sound source recognition [7], sound source separation [8], automatic transcription and musical retrieval [9], content-based coding, and sound synthesis [10].

Some fields of sound source modeling (SSM) are still in the very beginning, and therefore, the practical usage of SSM analysis and synthesis techniques is limited to some specific cases. It is easy to see that for general cases, e.g., analysis and synthesis of polyphonic music, the SSM approach faces challenging tasks in the analysis part, such as 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.

Nevertheless, SSM can be employed in more restrained situations, for instance, the SSM-based analysis and synthesis of the acoustic guitar presented in [11]. The system, which deals with two-voice polyphony examples, is able to analyze the signal, isolate tones, and recreate the signal again using a guitar synthesizer. The analysis part involves signal modeling techniques, such as sinusoidal modeling [12], combined with auditory modeling to pitch determination and signal separation. The synthesis part employs the physical modeling approach, e.g., the digital waveguide method [13], which has been successfully used to synthesize realistic instrument sounds by taking into account physical properties associated with the instruments and their particular playing techniques [14].

The objective of the previously mentioned system is related to content-based coding. However, it can serve the audio restoration purposes if both the analysis and the synthesis parts can be made robust to the presence of noise or degradation in the signal.

In principle, a content-based analysis would help to distinguish between a noise-like signal component to be preserved, and a degrading noise to be removed. This possibility could guide further choices of the signal components to be reconstructed in the synthesis part. Additionally, the SSM approach allows taking advantage of previous knowledge of the model parameters associated with a high quality instrument sound, and 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 the high frequencies either lost or severely degraded in the recording process, since high quality synthesis models for plucked-string tones are available, providing prior knowledge of their frequency content.

Based on what was said before, a set of simplifications is needed in the audio restoration/enhancement problem within the SSM framework. In this paper, the problem is restricted to the synthesis stage, since only single acoustic guitar tones are considered. For the SSM of plucked strings, a simple *commuted waveguide synthesis* (CWS) algorithm [15, 16] is employed. This choice allows obtaining the model parameters by analyzing recorded tones [17]. The study presented here is divided basically in two parts: a proposition to extend the bandwidth of originally bandlimited guitar tones, and a de-noising scheme for guitar tones which mixes a traditional spectral-based de-hissing method and SSM.

The paper is organized as follows. In Section 1, the CWS algorithm for plucked strings 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 de-hissing 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 THE 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 [13], and its further improvements and extensions [18] 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, the string, and the 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 [15, 16] for plucked strings. For more detailed information on the development of a guitar synthesizer based on the CWS method, see [19].

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

$$S(z) = \frac{1}{1 - z^{-L_{\rm i}} F(z) H(z)}, \tag{1}$$

where $L_{\rm i}$ and $F(z)$ are, respectively, the integer and fractional parts of the delay line associated with the length of the string, L. $H(z)$ 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 lowpass filter with transfer function given by

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

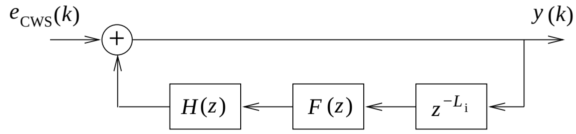

Fig. 1: Block diagram of the String Model

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. In this work, it is implemented as a fourth-order Lagrange interpolator FIR filter [20]. In this configuration, the string-model transfer function, $S(z)$, is completely defined by the length of loop delay, L, the loop filter parameters, 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 the string-model parameters

In this section, the estimation of the string parameters in the CWS model used in this work is discussed. The estimation procedure can be automatically performed by analyzing recorded tones, as shown in [17, 21].

The first step consists of 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 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 by

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

where f_s the the sampling rate of the analyzed signal.

The next step consists of 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. In this work, the estimation of the loop filter parameters is carried out in 3 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 [17].

2 BANDWIDTH EXTENSION OF GUITAR TONES

In this section, the problem of reconstruction of 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 interesting, to reconstruct the spectral information lost during a de-noising procedure. The latter case will be discussed later in Section 3.

The test signal used in this study is a single guitar tone which was lowpass filtered in order to remove the high frequency harmonics but 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.

The estimation of the fundamental frequency is not problematic, assuming that it was preserved in the lowpass 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 that, for this type of string model, variations between 25% and 40% in the time constant of the decay are not perceived [22], it is acceptable estimating the decay rate of the missing harmonics by analyzing a similar fullband 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, $e_{\mathrm{CWS}}(k)$, usually has a large number of resonances associated with all other information except that of the vibrating string, e.g., non-linearities associated to the plucking event, body resonances, coupling between strings.

In the bandwidth extension problem, the analyzed signal is already lowpass filtered, resulting in an excitation with a lowpass 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.

A possible strategy to fulfill the information which is missing in the lowpass filtered excitation would consist of analyzing other similar tones and properly mixing their excitations. However, it would require additional criteria to match, for instance, the plucking style and the dynamics of the analyzed tones.

A rather trivial alternative way to excite the string modes consists of modifying the attack part of the excitation signal in order to provide enough energy in the whole spectrum range to properly excite the string model. This procedure can be seen as the addition of an artificially generated plucking event, $e_{\mathrm{pluck}}(k)$, directly to the string model, triggered with the attack part of $e_{\mathrm{CWS}}(k)$, as illustrated in Fig. 2.

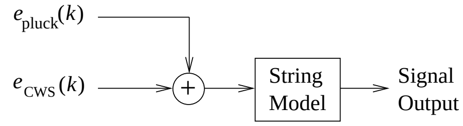

Fig. 2: Bandwidth extension scheme.

In this case, if the artificial plucking can really excite the string modes when resynthesizing the tone, it 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, estimating the spectral envelope of its attack part.

Additionally, it would be desirable to leave undistorted the harmonics originally present in the lowpass filtered tone. 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, whose magnitude response must approximate the spectral envelope of the very beginning of a full bandwidth excitation. The highpass filter $H_{\mathrm{hp}}(z)$ 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.

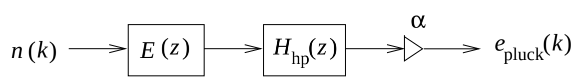

Fig. 3: Generation of the synthetic plucking event.

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 $H_{\mathrm{hp}}(z)$, and the value of the gain factor.

Further details concerning the choices of the filters $E(z)$ and $H_{\mathrm{hp}}(z)$, the length of the noise burst, and the value of gain α are given in Section 4.1.

3 SSM AND DE-NOISING OF GUITAR TONES

In this section, a processing scheme which involves SSM of guitar tones and traditional methods of audio restoration is proposed to improve the perceptual quality of de-hissed guitar tones.

Usually, spectral-based audio de-hissing methods suffer from a difficult tradeoff between the reduction of the noise effects and the distortion of the signal to be restored [2, 1]. 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 de-hissing methods is based on the 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 de-hissing acoustical musical signals, signal losses are more prominent at high frequencies, since lower SNR's are observed in this range. Naturally, auditory properties such as masking and critical bands help to explain the improved perceptual quality of de-hissed audio signals. However, the lack of a reference signal prevents the appealing use of auditory-based approaches as well as the employment of either pure objective or perceptual-based measures to evaluate the perceptual quality of the de-hissed signals.

The results of bandwidth extension, described in Section 2, provide a useful appeal to the de-hissing 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 afterwards if a sound source model is available for the signal. This possibility poses an alternative view on the de-hissing problem as discussed in the following sections.

3.1 Aggressive de-hissing and bandwidth extension

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

In this case, the estimation of the string-model parameters faces similar problems as those discussed in Section 2, when dealing with bandlimited guitar tones. Here, the high frequency harmonics are either masked by the corrupting noise or absent due to the aggressive de-hissing procedure. Thus, their decay rate estimation is prevented. Anyway, the same considerations drawn in Section 2 regarding to 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 de-hissing and bandwidth extension

Another strategy to the de-hissing problem consists of integrating the de-hissing and signal reconstruction procedures into a unique stage. This can be achieved by adapting the de-hissing 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 de-hissing method is still applicable to the excitation signal, since it has important resonances associated with the body modes. Of course, an aggressive de-hissing 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 on 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 de-hissing procedure, it will provide enough energy to properly excite the string model in order to resynthesize a non-smoothed and noise-free tone.

A simple way to protect the attack part of the excitation from the aggressive de-hissing is to control artificially the noise variance estimation used in the spectral-based de-hissing method. 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 de-hissing procedure. Firstly, a local estimate of the excitation variance at the attack part, σ^2_{attack}, should be obtained. Then, for each frame index i, the estimate of the noise variance is multiplied by a gain given by

$$\mu = \min\left(\mu_{\max}, \frac{\sigma^2_{\text{attack}}}{\sigma^2 + \epsilon}\right), \tag{4}$$

where σ^2 is a locally estimated excitation variance within a given frame i, ϵ is a very small positive value to prevent a division by zero, and $\mu_{\max}$ is a constant value which represents the maximum value of μ allowed. Since for most frames but those associated with the attack part of the excitation $\sigma^2 << \sigma^2_{\text{attack}}$, their ratio will assume higher values than $\mu_{\max}$, implying $\mu = \mu_{\max}$.

Further details on the implementation of the integrated de-hissing procedure are described in Section 4.2.

4 EXPERIMENTAL RESULTS

In this section, experimental results on the bandwidth extension and the de-hissing of single guitar tones are described. The test signal used in both cases is an F_4 tone with 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 lowpass filtered using a 101^{th} order equiripple FIR filter with cutoff frequency at 1 kHz, transition band of 1 kHz, and attenuation of 80 dB on the rejection band. In this case, regardless of the filtering procedure, the fundamental and the next 2 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 the lowpass 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 [23]. The radius of the poles was arbitrarily set to 0.8. It can be seen from Fig. 4 that, with these parameters, the frequency response of $E(z)$ approximates quite well the spectral envelope associated with the attack part of a full bandwidth excitation.

The highpass 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 window. 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.

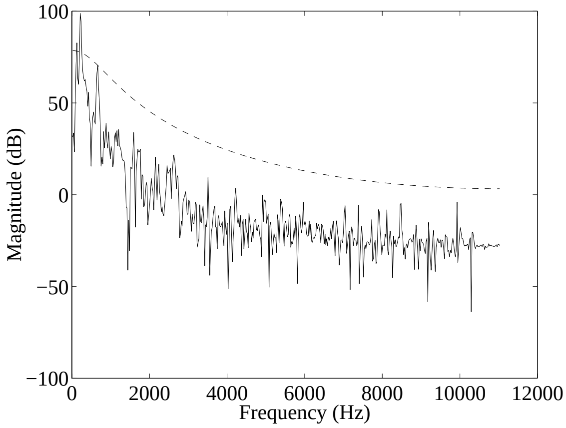

Fig. 4: Squared magnitude response of $E(z)$ (dashed line), as defined in Section 4.1, compared with the power spectrum associated with the attack part of a full bandwidth excitation.

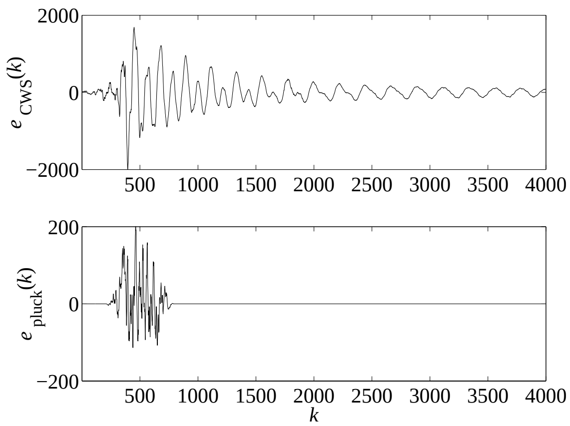

Fig. 5: Synchronization of the noise burst (bottom) with the attack part of the excitation (top).

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 lowpass 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, e.g., 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 lowpass 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 obtained results were similar to that of the previous case. However, the perceptual differences between the original and the 3000 Hz bandlimited tone are already less prominent. Therefore, the effects of the bandwidth extension procedure are more difficult to perceive. Sound examples are available at URL: `http://www.acoustics.hut.fi/publications/papers/aes110-ssm/`

4.2 De-hissing

In the de-hissing 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.

The first step of the de-hissing and post-processing approach consisted of de-hissing 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. Additionally, 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.

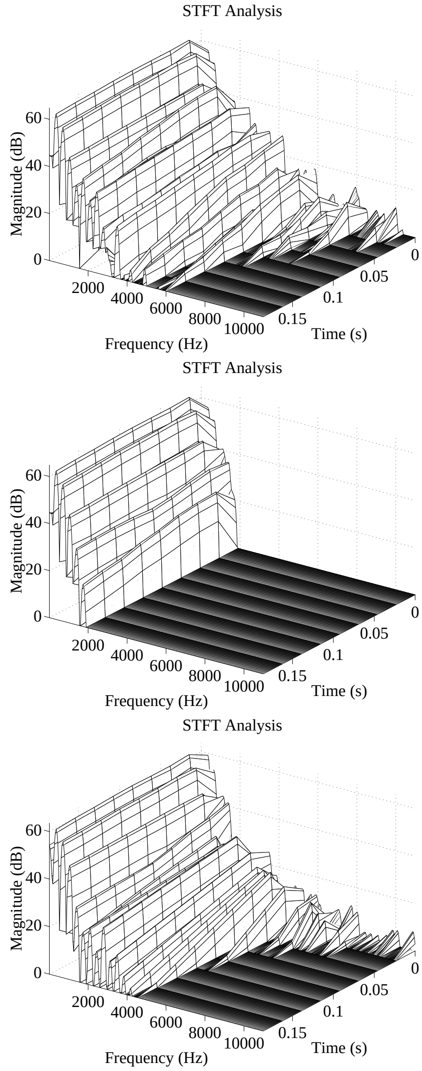

Fig. 6: Time-frequency analysis of the original tone (top), the lowpass filtered tone (middle), and the resynthesized tone (bottom).

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.

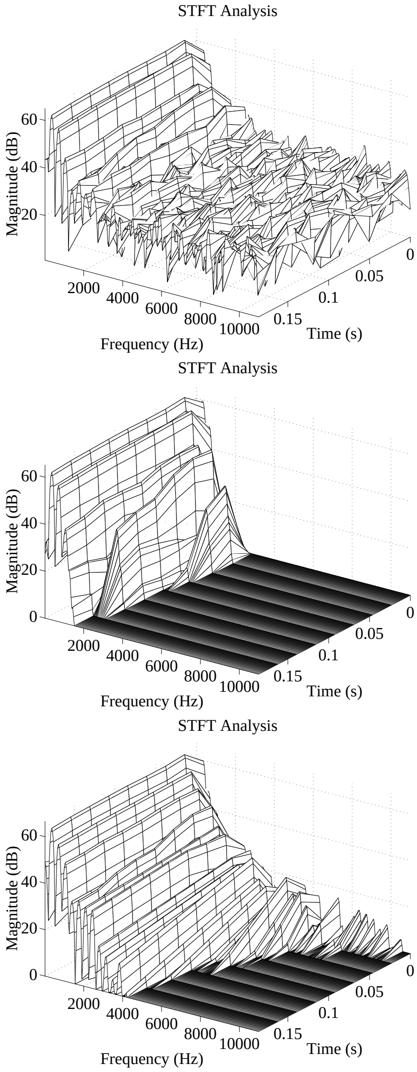

Fig. 7: Time-frequency analysis of the noisy tone (top), the de-hissed tone (middle), and the bandwidth extended tone (bottom).

The last step consisted of applying the bandwidth extension procedure to the aggressively de-hissed 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 de-hissed, and bandwidth extended signals are shown.

As can be seen, the noise masks the high-frequency harmonics, which together with the noise are also removed during the de-hissing procedure. Nevertheless, the SSM-based to bandwidth extension is able to reconstruct the missing harmonics in the resynthesized tone. 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 de-hissing 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 de-hissed 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 de-hissing 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. Additionally, the noise floor estimates were multiplied by gain factors μ, defined in Eq. (4), which were also computed for each frame.

The computation of μ involved the following choices. The value of $\sigma^2_{\mathtt{attack}}$ was set as the power of the attack part of the noisy excitation. The value of $\mu_{\max}$ 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 which 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.

The sequence of values of μ as a function of the frame index obtained after de-hissing the noisy excitation is shown in Fig. 8 (plot on the top).

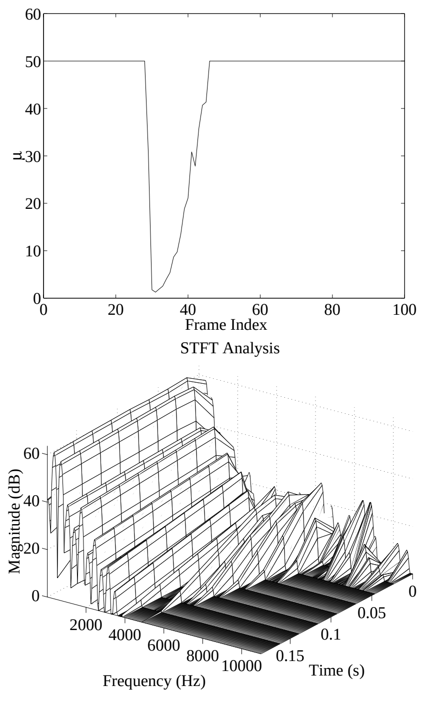

Fig. 8: Values of the noise floor gain as a function of the frame index (top), and time-frequency analysis of the restored tone (bottom).

Note that, outside the attack part of the excitation, which starts at frame 30, $\mu = \mu_{\max}$. Thus, an aggressive de-hissing 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 on the bottom of Fig. 8 shows a time-frequency analysis of the resynthesized signal obtained from the previously de-hissed 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 that of the restored tone obtained in the previous experiment. Sound examples are available at URL: `http://www.acoustics.hut.fi/publications/papers/aes110-ssm/`

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 de-hissing scheme, which mixed a traditional spectral-based method with SSM. The obtained results for both the bandwidth extension and the de-hissing 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, e.g., 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 very 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 even in the simple case of restoring solo guitar music, the use of SSM implies challenging tasks related to content-based representations of music. As an example, the separation of tones whose content overlaps both in time and frequency as well as the extraction of their musical features can be mentioned. Extensions to more general cases can be viewed as a multi-layered problem, which would include separation of more general musical elements in complex sound sources. 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.

ACKNOWLEDGMENTS

The work of P. Esquef has been supported by a scholarship from 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.

REFERENCES

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

[2] O. Cappé, “Elimination of the Musical Noise Phenomenon with the Ephraim and Malah Noise Suppressor,” *IEEE Trans. Speech and 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,” in *Proc. AES 108th Convention*, (Paris, France, 2000 Feb.). Preprint 5150.

[4] D. Tsoukalas, J. Mourjopoulos, and G. Kokkinakis, “Perceptual Filters for Audio Signal Enhancement,” *J. Audio Eng. Soc.*, vol. 45, no. 1, pp. 22–36 (1997 Jan.).

[5] B. L. Vercoe, W. G. Gardner, and E. D. Scheirer, “Structured Audio: Creation, Transmission, and Rendering of Parametric Sound Representations,” *Proceedings of IEEE*, vol. 86, no. 5, pp. 922–940 (1998).

[6] M. Karjalainen, “Immersion and Content — A Framework for Audio Research,” in *Proc. IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, (New Paltz, NY, 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 Jun.).

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

[10] 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. (A PostScript version available at URL: `http://www.acoustics.hut.fi/publications/`).

[11] T. Tolonen, “Object-Based Sound Source Modeling for Musical Signals,” in *Proc. AES 109th Convention*, (Los Angeles, California, USA, 2000 Sept.). Preprint 5174.

[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* (M. Kahrs and K. Brandenburg, eds.), pp. 314–416, Boston, USA: Kluwer Academic Publishers, 1 ed., 1998.

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

[14] M. Laurson, J. Hiipakka, C. Erkut, M. Karjalainen, and V. Välimäki, "From Expressive Notation to Model-based Sound Synthesis: a Case Study of the Acoustic Guitar," in *Proc. Int. Computer Music Conf. (ICMC'99)*, (Beijing, China, 1999), pp. 1–4.

[15] J. O. Smith, "Efficient Synthesis of Stringed Musical Instruments," in *Proc. Int. Computer Music Conf. (ICMC'93)*, (Tokyo, Japan, 1993 Sept.), pp. 64–71.

[16] M. Karjalainen, V. Välimäki, and Z. Jánosy, "Towards High-quality Sound Synthesis of the Guitar and String Instruments," in *Proc. Int. Computer Music Conf. (ICMC'93)*, (Tokyo, Japan, 1993 Sept.), pp. 56–63.

[17] 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, no. 5, pp. 331–353 (1996 May).

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

[19] V. Välimäki and T. Tolonen, "Development and Calibration of a Guitar Synthesizer," *J. Audio Eng. Soc.*, vol. 46, no. 9, pp. 766–778 (1998 Sept.).

[20] 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, no. 1, pp. 30–60 (1996).

[21] T. Tolonen, "Model-Based Analysis and Resynthesis of Acoustic Guitar Tones," Master's thesis, Helsinki Univ. of Technology, Laboratory of Acoustics and Audio Signal Processing, Espoo, Finland. Report no. 46 (1998 Jan.), (A PostScript version available at URL: `http://www.acoustics.hut.fi/publications/`).

[22] T. Tolonen and H. Järveläinen, "Perceptual Study of Decay Parameters in Plucked String Synthesis," in *Proc. AES 109th Convention*, (Los Angeles, California, USA, 2000 Sept.). Preprint 5205.

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