Audio Restoration Using Sound Source Modeling

Paulo Esquef, Vesa Välimäki, and Matti Karjalainen

Helsinki University of Technology
Laboratory of Acoustics and Audio Signal Processing
P.O.Box 3000, FIN-02015 HUT, Espoo, Finland
esquef@acoustics.hut.fi

ABSTRACT

This paper presents new propositions to audio restoration and enhancement based on Sound Source Modeling (SSM). The main motivation is to take advantage of prior information of generative models of sound sources when restoring or enhancing musical signals. We describe a case based on the commuted waveguide synthesis algorithm for plucked string tones and devise a scheme to extend the bandwidth of guitar tones. Then, we study the de-hissing of guitar tones and propose a scheme in which the bandwidth extension method is applied as a post-processing stage to a spectral-based de-hissing procedure. According to our experiments, this is effective for the reduction of common side-effects associated with spectral-based de-hissing methods, such as musical noise and signal distortion.

1. INTRODUCTION

Signal modeling techniques have been widely used in audio restoration purposes. In these techniques, the analysis and synthesis parts of the processing only deal with the information available in the surface presentation of audio signals. However, audio analysis and synthesis can also consider how the sound elements are structured in the audio signal [1]. This kind of approach asks for better understanding of the human auditory perception, as well as deeper representations of sound sources, which, in fact, are important requirements for the actual challenges of the audio signal processing field, such as sound source recognition, sound source separation, automatic transcription and musical retrieval [2], content-based coding, and sound synthesis [3].

In principle, a content-based audio 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, SSM allows taking advantage of previous knowledge of the model parameters associated with a high quality instrument sound to enhance the sound quality of a poorly recorded instrument. However, the practical usage of SSM is still limited to some specific cases, e.g., analysis and synthesis of monophonic instrument sounds.

In this paper, we show 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. Our 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 is employed [4, 5]. This choice allows obtaining the model parameters by analyzing recorded tones [6]. 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.

2. 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

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

where L_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 low-pass filter with transfer function given by

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

The magnitude response of the filter $H(z)$ 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

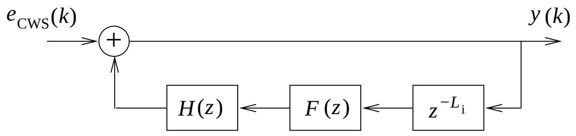

Fig. 1. Block diagram of the string model.

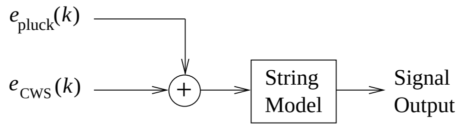

Fig. 2. Bandwidth extension scheme.

interpolator FIR filter [7]. 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.

The value of L is obtained by

$$L = \frac{f_{\mathrm{s}}}{\hat{f}_0}, \tag{3}$$

where f_{s} the the sampling rate of the analyzed signal, and $\hat{f}_0$ is an estimate of the fundamental frequency of the tone.

The parameters of the loop filter are obtained by first estimating the decay rate of the harmonics. Then, the resulting loop gains are used as a target magnitude response for the loopfilter. A detailed description of the procedures used to estimate the string model parameters can be found in [6].

The excitation $e_{\mathrm{CWS}}(k)$, shown in Fig. 1, is obtained by inverse filtering the guitar tone through the previously estimated string-model.

3. 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 to reconstruct the spectral information lost during a de-noising procedure.

Let us consider a single guitar tone which was lowpass 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 [6]. Due to the simplicity of the string-model we are employing, and perceptual aspects [8], it is acceptable to analyze a similar fullband guitar tone to overcome the impossibility of estimating the decay rate of the missing harmonics.

Based on the CWS properties, the tone can be inverse filtered resulting in an excitation, $e_{\mathrm{CWS}}(k)$. If the analyzed tone is already lowpass filtered, the corresponding $e_{\mathrm{CWS}}(k)$ will have a lowpass characteristic as well. This means that we need to provide extra energy to the excitation in order to fully excite the string-model modes.

A possible way to achieve that consists of adding to the attack part of the excitation an artificially generated plucking event, $e_{\mathrm{pluck}}(k)$, as illustrated in Fig. 2.

A suitable option for $e_{\mathrm{pluck}}(k)$ is to generate an impulsive noise burst, for instance, by windowing a zero-mean Gaussian white noise sequence. 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. This can be achieved by coloring the noise burst sequence according to known information about typical spectral characteristics of guitar bodies.

The generation of the noise burst, which simulates a plucking event, is 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.

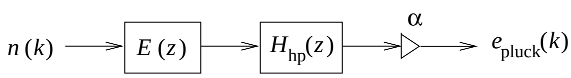

Fig. 3. Generation of the synthetic plucking event.

The capability of the previously described method to extend the bandwidth of guitar tones is illustrated in Fig. 4. In this example, a test signal consisting of an F_4 tone with fundamental frequency of 347 Hz, sampled at 22.05 kHz was used. The tone 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. Filter $E(z)$ was chosen as a second-order resonator tuned at 200 Hz. This frequency corresponds to the lowest mode of the top plate of the guitar body [9]. The radius of the poles was arbitrarily set to 0.8. With these parameters, the frequency response of $E(z)$ approximates the spectral envelope associated with the attack part of a full bandwidth excitation. The highpass filter $H_{\mathrm{hp}}(z)$ was not included. Finally, the noise burst was multiplied by a Hanning window of 600 samples, scaled, and added to the attack part of the excitation.

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. Additional tests were performed on the same guitar tone but with bandwidth limited to 500 Hz and 3000 Hz. The obtained results were similar to that of the previous case.

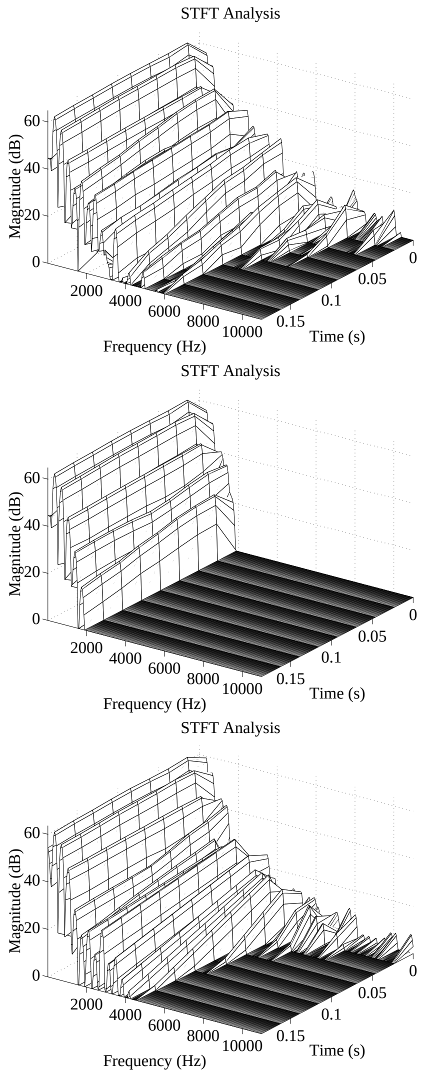

Fig. 4. Time-frequency analysis of the original tone (top), the lowpass filtered tone (middle), and the resynthesized tone (bottom).

4. SSM AND DE-NOISING OF GUITAR TONES

Usually, spectral-based de-hissing methods suffer from a difficult tradeoff between the reduction of the noise effects and the introduction of distortion in the restored signal [10, 11]. The results of the SSM-based bandwidth extension of guitar tones, described in Section 3, can be useful in the de-hissing problem. The hard tradeoff between noise reduction and preservation of the signal information can be softened on the grounds that the spectral content can be reconstructed afterwards if a sound source model and a synthesis algorithm are available for the analyzed signal.

A possible option to remove the noise effects from a guitar tone corrupted by zero mean white Gaussian noise is to de-hiss it through a spectral-based method using an over-estimated value for the variance of the corrupting noise. However, the side-effect of this approach is to end up with an oversmoothed restored tone which lacks high frequencies. A remedy to the oversmoothing effect is to apply the SSM-based bandwidth extension to recover the signal information that was lost due to the aggressive de-hissing procedure.

The previous two-steps strategy was found to be effective to de-hiss guitar tones, as can be seen in Fig. 5. In this experiment, a zero mean white Gaussian noise sequence was added to the test guitar tone and its variance was adjusted to produce a global SNR of 20 dB. As can be seen in the top plot of Fig. 5, the noise masks the high-frequency harmonics of the tone.

The first step of the restoration procedure consisted of de-hissing the noisy signal through a Wiener filtering scheme, as described in [11]. Here, 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.

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. This can be verified in the middle plot of Fig. 5, as well as the strongly smoothed, i.e. lowpass filtered, characteristic.

The last step consisted of extending the bandwidth of the previously de-hissed guitar tone using the SSM-based scheme described in Section 3. The same approaches to estimate the string-model parameters and to generate the additional excitation signal were employed in this experiment. These choices were found to generate a restored tone whose timbre is similar to that of uncorrupted tone, without the annoying effects of the residual noise as can be seen in Fig. 5 (bottom). Sound examples are available at URL: http://www.acoustics.hut.fi/publications/papers/fs01-ssm/

5. 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, the SSM-based bandwidth extension scheme was applied as a post-processing stage after a traditional spectral-based de-hissing method. 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. Even if the attempt is to restore solo guitar music, SSM-based techniques face 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

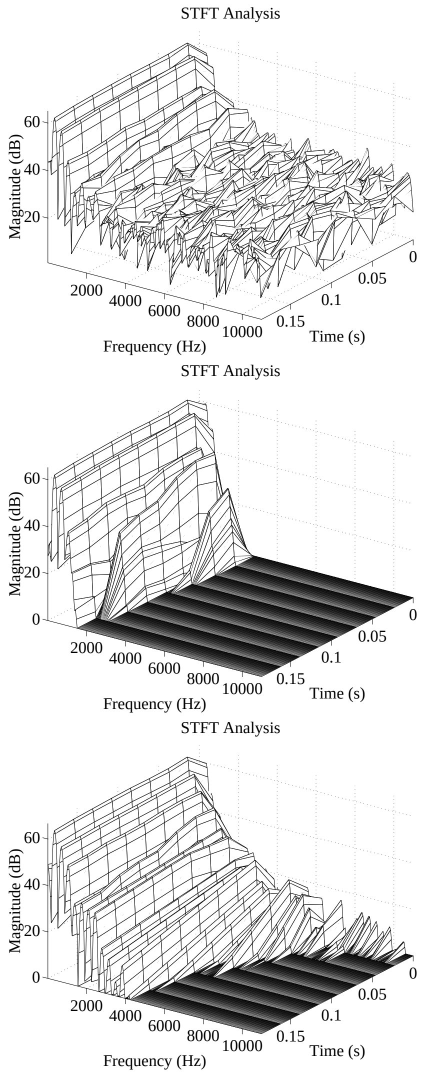

Fig. 5. Time-frequency analysis of the noisy tone (top), the de-hissed tone (middle), and the bandwidth extended tone (bottom).

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.

Finally, it is worth mentioning that 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 using traditional techniques.

ACKNOWLEDGMENT

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] 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.

[2] A. Klapuri, "Automatic transcription of music," M.Sc. thesis, Tampere Univ. of Technology, Tampere, Finland, 1998, (Electronic version available at URL: http://www.cs.tut.fi/˜klap/iiro/).

[3] T. Tolonen, V. Välimäki, and M. Karjalainen, "Evaluation of modern sound synthesis methods," Tech. Rep. 48, Helsinki Univ. of Technology, Lab. of Acoustics and Audio Signal Processing, Espoo, Finland, 1998, (Electronic version available at URL: http://www.acoustics.hut.fi/publications/).

[4] J. O. Smith, "Efficient synthesis of stringed musical instruments," in *Proc. Int. Computer Music Conf., ICMC'93*, Tokyo, Japan, 1993, pp. 64–71.

[5] 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, pp. 56–63.

[6] 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.

[7] T. I. Laakso, V. Välimäki, M. Karjalainen, and U. K. Laine, "Splitting the unit delay — tools for fractional delay filter design," *IEEE Signal Proc. Mag.*, vol. 13, no. 1, pp. 30–60, 1996.

[8] T. Tolonen and H. Järveläinen, "Perceptual study of decay parameters in plucked string synthesis," in *Proc. AES 109th Convention*, Los Angeles, California, USA, 2000, Preprint 5205.

[9] O. Christensen and B. B. Vistisen, "Simple model for low-Frequency guitar function," *J. Acoust. Soc. Am.*, vol. 68, pp. 758–766, 1980.

[10] O. Cappé, "Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor," *IEEE Trans. Speech and Audio Process.*, vol. 2, no. 2, pp. 345–349, 1994.

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