TRANSFORMATION OF INSTRUMENTAL SOUND RELATED NOISE BY MEANS OF ADAPTIVE FILTERING TECHNIQUES

Sandro Zabarella

Dipartimento di Elettronica e Informatica
Università di Padova
zaba@dei.unipd.it

Carlo Drioli

Dipartimento di Elettronica e Informatica
Università di Padova
adrian@dei.unipd.it

ABSTRACT

In this work we present an extension of the classic schema of a time-varying filter excited with white noise for the modeling of noise signals from musical instrument sounds. The framework used is that of statistical signal processing, and a structure that combines an Autoregressive (AR) model with an adaptive FIR filter is proposed. A comb-based structure for the AR filter is used when tuned noise is to be modeled. The analysis/resynthesis schema proposed is used to perform some basic sound transformations such as time stretching, tuning and energy envelop control, and spectral processing.

1. INTRODUCTION

In the field of speech and sound modeling, it is common practice to decompose the sounds in the sum of two components: the sinusoidal or harmonic one and the residual or stochastic one. In this way, it is possible to use different synthesis techniques for the two parts, e.g. it is possible to use frequency-domain techniques to represent and control the deterministic part, and time-domain techniques to represent and control the stochastic part [1].

The problem of the encoding and transformation of the deterministic part of sounds in term of time-varying sinusoids has been deeply explored. A set of robust techniques is today available to perform high quality low-level sound transformations, such as time-stretching and pitch shifting, as well as high-level sound transformations, such as control of expressiveness in digitally recorded musical performances. On the other hand, there is a lack of models for noisy sound transformations, such as the ones mentioned, in a time-frequency analysis/synthesis framework. A common approach is to use a noise- or pulse-driven source filter model for the analysis and synthesis of stochastic components [1],[2]. However, this model generally does not work well due to the loss of synchronization between the sinusoidal and stochastic part. Another recent approach proposes to represent a single component of the sinusoidal decomposition by partials with noise-enhanced bandwidth [3]. This representation allows
for effective and compact sound manipulations, but still has the limitations of a frame-based analysis-synthesis approach for the modeling of fast transients.
Our research aims at extending the classic schema of a time-varying filter excited with white noise, by organizing an ARMA filter in two parts, having different functionalities: the MA part is responsible for the modeling of rapidly time-varying transients, while the AR part is responsible for the stationary or slowly changing spectral characteristics.
Having these two components separated and modeled by two different filters gives better control in terms of the sound parameters, so that time-scale transformations can be done by changing the rate in the temporal sequence of the MA coefficient, while pitch related or spectral related transformations can be done by changing the coefficients of the AR part. When opportune, a comb AR filter is used to model the emphasis, in the noise spectrum, of frequency regions related to the harmonics of the deterministic part.

2. ANALYSIS AND SYNTHESIS FRAMEWORK

The theoretic framework we used is that of statistical signal processing [4]. In detail, the recursive least mean squares (RLMS) algorithm is used to identify the time-varying MA filter. The RLMS algorithm is is an adaptive algorithm based on the computation of the estimate error, e(k)=d(k)-y(k), where d(k) and y(k) are respectively the desired and the actual output signal, and on the adaptation of the coefficients of the filter **c**(k) according to the formula:

$$\mathbf{c}(k+1) = \mathbf{c}(k) + \mu e(k)\mathbf{x}(k)$$

where

$$\mathbf{c}(k) = [c_0(k),\ c_1(k),\ \dots\ ,\ c_{N-1}(k)]^T$$
$$\mathbf{x}(k) = [x(k),\ x(k-1),\ \dots\ ,\ x(k-N+1)]^T$$

x(k) is the input signal, and μ is an opportune constant.
It has been observed that RLMS algorithm offers good performance for the transients and rapidly time-varying parts of the desired signal.
Conversely, to model the slowly-varying part of the signal, a linear prediction (LP) analysis based on the correlation method, is used to identify the AR [4]. The combination of these two techniques gives good synthesis results and a meaningful control interface for transformations. The overall model is a feed forward schema composed by a time-varying MA filter excited with white noise followed by an AR filter, as in Fig. 1

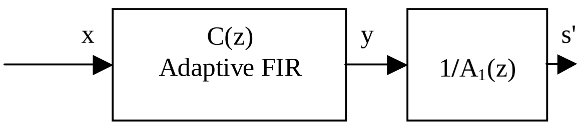

Fig. 1: Schema of the noise model.

Referring to Fig. 1, the linear prediction analysis permits to estimate the quasi-stationary spectral envelope of the desired signal *s*. This spectral envelope can be removed by deconvolving *s* with the inverse filter, i.e. the all-zeros filter $A_1(z)$, and an excitation signal *y* is obtained as prediction error. If *s* does not have rapidly changing transients, the prediction error is 'nearly' white. Otherwise, a further modeling of the rapidly changing characteristics is required. This task is accomplished by the adaptive filter *C(z)*, that is realized using an RLMS algorithm. This process permits to reproduce the signal *y* by a time-varying FIR filter excited with a real white gaussian noise input, with variance equal to the noise energy.
It has to be noted that, in this preliminary form the filter coefficients are computed and stored at sample rate. This makes the algorithm quite inefficient from an encoding point of view, and suitable strategies are under investigation.

In some cases, sounds from acoustic instruments are characterized by a noise energy that is weak relative to the sinusoidal energy, so that the stochastic component retains some of the tuning characteristics of the deterministic component. Even though a single AR filter with sufficiently high order will fit the desired spectral envelope, it is common practice to separate the linear prediction schema in a long-term predictor, responsible for the periodicity of the signal (i.e., the fine structure of the spectrum) and a short-term predictor, responsible for the gross structure of the spectrum. The schema is shown in Fig. 2.

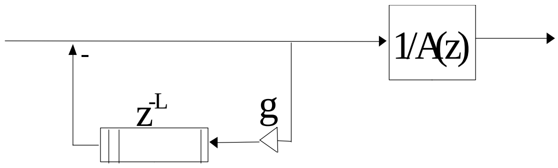

Fig. 2: Schema of the comb-based AR filter.

This representation provides a mean of manipulating the noise independently form the deterministic part, as the comb filter can be easily tuned, and the all-poles filter can be interpreted as a formant modeler. The AR filter corresponding to this configuration can be written as

$$\frac{1}{A_1(z)} = \frac{1}{1+gz^{-L}} \cdot \frac{1}{A(z)} = \frac{1}{A(z)+gA(z)z^{-L}}$$

A two-step approach is usually adopted for the identification of the two cascaded filters: first, a low-order LPC filter is used to remove the short-term correlation in the signal; then a comb filter is identified to remove the long term correlation [6]. Here, we use a different approach. We observe that the denominator polynomial provided by a conventional LP analysis of tuned noise presents most of the times a structure that is coherent with the representation of Fig. 2 (see, for example, the dashed line of Fig. 3, representing the coefficient of the polynomial): i.e. it has the right-most part that resembles the left-most part, scaled by a constant less than one in modulus, and it presents low-energy coefficients in the middle. Based on this observations, the algorithm is as follows: first, a conventional LP analysis of sufficiently high order is performed, resulting in a high-order polynomial $A_1(z)$ with coefficients $[1\ a_{1,1} \dots a_{1,P}]$. Then, an autocorrelation analysis on the coefficients is used to estimate the parameter *L* (and, as a consequence, the order N of the filter *A(z)*). If the similarity of the first and the last *N+1* coefficient is pronounced, then the easiest way to have the desired approximation is to simply take $A(z)=[\ 1\ a_{1,1} \dots a_{1,N}]$ and $g= a_{1,L}$ (Fig. 4 show the corresponding results). This approach has the advantage that the final representation is stable for pitch transformations.

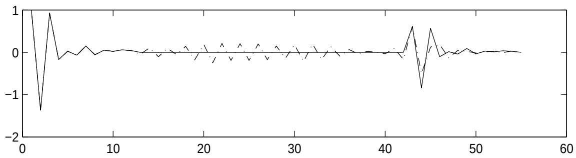

Fig 3: Coefficients of the polynomial $A_1(z)$. Dashed line: original LP analysis with filter order $P=55$; continuos line: adaptation of $A_1(z)$ to the comb-based schema with $N=13$ and $L=42$.

However, since this approximation is not always accurate, especially for low pitch tuned noises, a valid alternative is to perform a second optimization step (once L and N are known), solving a constrained LS problem. In this case, the Youle-Walker equations are solved with an iterative method, and the coefficients $a_{1,(N+1)} \dots a_{1,(L-1)}$ are forced to assume low-energy values.

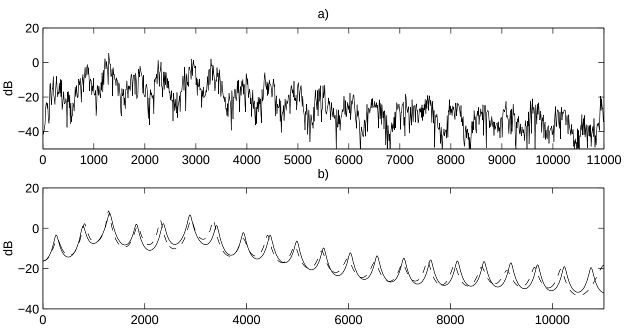

Fig 4: Identification of the AR filter. Upper plot: desired spectrum. Lower plot: AR model identification, traditional LP analysis (dashed line) vs adapted comb-based AR model (continuos line).

3. RESULTS AND DISCUSSION

The model just presented above is suitable to transformations of signals, since every part of the model controls a determined characteristic of the process and a sound transformation schema is organized as follows.

Two different realizations of time stretching are used for transients and quasi-stationary regions: transients are stretched by an opportune interpolation of both the MA adaptive filter coefficients **c**, and the time-varying gain of the input white noise, which has been estimated in the analysis step as the energy envelop of the signal. Slowly-varying regions of the signal are stretched, in the simplest case, by acting only on the AR filter (in this case the MA filter is reduced to a constant). Here, the gain of the input white noise, estimated as before, is interpolated. However, when time stretching is to be performed on a region of the signal modeled by both the AR and the MA filters, the two actions described above are synchronized to obtain the desired effect.

When a comb-based extended structure of the AR filter is used for tuned noise, the tuning characteristics of the signal are controlled by changing the length L of the comb delay line. The parameter g is used to determine the bandwidth of the partial-related energy bands, in order to control the tuning degree of the noise.

In any case, the AR filter is organized in second-order cells to have independent access to the single energy formants. Transformations of the short-term spectral envelope can be done by changing the central frequencies and bandwidths of the second-order AR cells.

The analysis/resynthesis schema has been tested on some sounds from monophonic instruments: in particular, two examples of a flute and an oboe residuals (sampled at 22050 Hz, 16 bit/sample) are shown in Fig. 5, 6 and 7. Informal listening tests demonstrate that the signals generated with the proposed model are perceptually indistinguishable from the target signals, and that the time scale and tuning transformations are quite realistic.

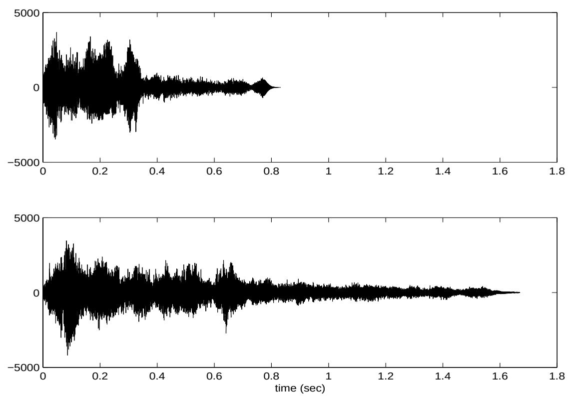

Fig. 5: time stretching of a flute residual noise (attack transient). It has been stretched by interpolation of the

adaptive MA filter coefficients and input gain, from start to *0.4* sec. Then, a simple AR model is used for the remaining release part, and the time stretching is done by interpolating the analysis gain for the excitation white noise.

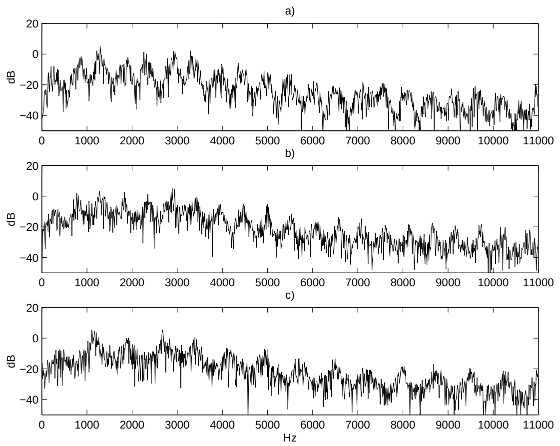

Fig. 6: Example of a tuning control, by means of a change of the parameter *L* in the comb-filter. Upper figure: target spectrum. Middle figure: spectrum of the signal resynthesized by means of the adapted AR model. Lower figure: signal resynthesized with a shorter delay line of the comb filter.

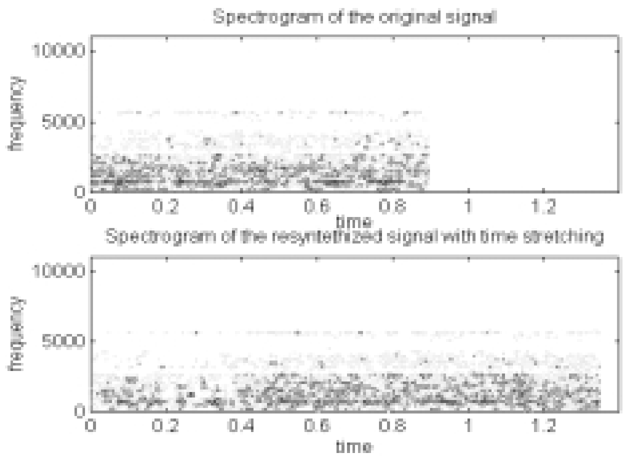

Fig. 7: Example of a time stretching by means of a AR model of the stationary region of the residual noise.

4. CONCLUSIONS

A model of the noisy part of sounds has been proposed. It consists of two parts: the first is represented by a FIR filter which uses the adaptive technique of RLMS algorithm; the second is instead represented by an all-pole filter obtained by an algorithm of linear prediction coding. The first filter is responsible for the modeling of transients, while the latter filter is responsible for the stationary spectral structure of the noise. A comb-filter structure has been used for tuned noise, where the AR original filter is splitted in two AR filters taking into account respectively the tuning and the formantic structure of the spectrum.

This model has given very good performances, both in resynthesis of sounds and in their transformations, such as time stretching. It has been proved for the noisy part of sounds from monophonic instruments, in particular flute and oboe.

5. REFERENCES

[1] X. Serra. "Musical sound modeling with sinusoids plus noise", in C. Roads and others (eds.), Musical Signal Processing, Swets and Zeitlinger Publishers, 1997.

[2] Y. Ding, X.Qian. “Sinusoidal and residual decomposition and modeling of musical tones using the QUASAR signal model”, Proc. ICMC97, pp. 35-42, Thessaloniki, 1997.

[3] K. Fitz, L. Haken, P. Christensen. “A new algorithm for bandwidth association in bandwidth-enhanced additive sound modeling”, Proc. ICMC 2000, pp.384-387, Berlin, 2000

[4] L. Ljung. “System identification - Theory For the User, 2nd ed”, PTR Prentice Hall, Upper Saddle River, N.J., 1999.

[5] J. Makhoul. “Linear prediction: a tutorial review”, Proc. IEEE, vol. 63, no. 4, pp. 561-580, Apr. 1975.

[6] J. R. Deller, J. G. Proakis, J. H. Hansen. “Discrete-Time Processing of Speech Signals”, Macmillan Publishing Company, New York, 1993.