Katholieke Universiteit Leuven

Departement Elektrotechniek ESAT-SISTA/TR 2000-49

Dynamic timbre matching for sound synthesis using time-varying parametric filters.[1]

Axel Nackaerts[2], Bart De Moor[3] and Rudy Lauwereins.[4]

August 2000
Published in the Proc. of the 43rd IEEE Midwest Symposium on Circuits and Systems

[1] This report is available by anonymous ftp from *ftp.esat.kuleuven.ac.be* in the directory *pub/sista/nackaerts/reports/report_timbrematch.ps.gz*

[2] E-mail: *axel.nackaerts@esat.kuleuven.ac.be* This work is supported by several institutions: the Flemish Government (Mefisto-666), the FWO, and Research Communities (ICCoS and ANMMM), IWT and IWT projects (EUREKA 1562-SINOPSYS, EUREKA 2063-IMPACT, STWW), the Belgian State, Prime Minister's Office - OSTC - (IUAP P4-02 1997-2001 and IUAP P4-24 1997-2001, Sustainable Mobility Programme - Project MD/01/24 1997-2000), the European Commission (TMR Networks: ALAPEDES and System Identification, Brite/Euram Thematic Network : NICONET) and by Texas Instruments (TI-Elite program). ESAT-K.U.Leuven is member of DSP Valley.

[3] K.U.Leuven, Dept. of Electrical Engineering (ESAT), Research group SISTA, Kardinaal Mercierlaan 94, 3001 Leuven, Belgium, Tel. 32/16/32 17 09, Fax 32/16/32 19 70, WWW: *http://www.esat.kuleuven.ac.be/sista*. E-mail: *bart.demoor@esat.kuleuven.ac.be*

[4] K.U.Leuven, Dept. of Electrical Engineering (ESAT), Research group ACCA, Kardinaal Mercierlaan 94, 3001 Leuven, Belgium, Tel. 32/16/32 10 50, Fax 32/16/32 19 86, WWW: *http://www.esat.kuleuven.ac.be/acca*. E-mail: *rudy.lauwereins@esat.kuleuven.ac.be*

Abstract

A method based on parametric filters for dynamic amplitude and timbre-change in musical instrument sound synthesis is presented. The filter coefficients are determined by analysis of an instrument recording. Results show better qualitative performance than single-tone variable-amplification and two-tone crossfading wavetable synthesis.

Dynamic timbre matching for sound synthesis using time-varying parametric filters

Axel Nackaerts, Bart De Moor, Rudy Lauwereins
Department Elektrotechniek-ESAT
Katholieke Universiteit Leuven
3000 Leuven, Belgium

***Abstract*—A method based on parametric filters for dynamic amplitude and timbre-change in musical instrument sound synthesis is presented. The filter coefficients are determined by analysis of an instrument recording. Results show better qualitative performance than single-tone variable-amplification and two-tone crossfading wavetable synthesis.**

I. INTRODUCTION

Most sound synthesis algorithms have two basic properties in common: the synthesis engine has to generate a tone with a specified pitch and with a specified loudness. In this work, the second property is addressed, specifically for wind instrument sound synthesis where dynamic input changes are common.

Several methods exist to obtain output with the desired loudness [1]. The one commonly used is simply amplifying the output. In reality, when an instrument is played at a different loudness, not only the total, time-domain amplitude changes, but also the relative amplitude of partials in the frequency domain. This can easily be understood by looking at the acoustical properties of musical instruments [2]. When an excitation signal is applied, the output can be regarded as an energy distribution over a set of allowed vibrational modes. It is possible that below a certain input level, some modes will only scarcely be excited and thus have a small, if not zero, amplitude. In subjective terms, an instrument played loudly tends to sound "brighter" than a softly played one. This input-dependent timbre-change cannot be realised using a global amplitude change only, as this leaves the relative amplitudes of the partials unchanged.

The most widespread algorithm for sound synthesis in use today is wavetable synthesis. Here, a prerecorded waveform is played back at the desired pitch and amplitude [3]. To obtain the timbre-change, one uses a crossfading technique between several waveforms of the same tone, recorded at different playing levels. Careful choice of the waveforms yields high-quality results [4]. On the other hand, this technique requires more memory and is more difficult to apply when continuous control of the amplitude is needed (e.g. for wind instruments). Analysis of the timbre-change is another possible approach.

II. ANALYSIS

The method used in this paper is to subtract undesired partials from a signal that contains all the partials. This is similar to source-filter synthesis [3]. A good candidate for such a signal is a recording of the instrument played at high input level.

Axel Nackaerts is a Research Assistant with the I.W.T. at the Katholieke Universiteit Leuven. Bart De Moor is a senior Research Associate with the F.W.O. and a Professor Extra-Ordinary at the K.U. Leuven. Rudy Lauwereins is Professor at the K.U.Leuven.

Analysis of a soft-to-loud input sequence with the Short-Time Fourier Transform (STFT) shows that the output of a softly played instrument only contains the fundamental frequency. At higher levels, harmonics appear and at maximum input, the amplitudes of the harmonics can be as high as that of the fundamental. We would like to use a simple, low-order linear filter (the "expression filter") to remove or at least attenuate the undesired partials at only a small calculational cost.

The spectral envelope of a wind instrument tone (for one time frame) can be approximated with a second order IIR resonant lowpass spectrum described by

$$\mathcal{H}(s) = \frac{A\omega_n^2}{s^2 + 2\xi\omega_n s + \omega_n^2}, \tag{1}$$

in which A denotes the amplitude, ξ the damping coefficient ($0 < \xi \leq 1$) and ω_n the resonant frequency in the absence of damping. The lower limit for the parameter ξ ($\xi > 0$) ensures that a stable filter is obtained. Without the upper limit $\xi \leq 1$, the poles of system would split to two different, real poles. The coefficients of this filter can be found by minimising the weighted least squares problem (WLS) with cost function

$$F = \sum W|S|(|S| - |\mathcal{H}|)^2. \tag{2}$$

Here, S is the spectrum of the signal as determined by the STFT. W is a scaling factor, chosen large when $|S| > |\mathcal{H}|$ and small otherwise. This choice ensures that only the dominant peaks in the amplitude spectrum are important to the WLS and that the spectral envelope is "dressed" over those peaks. Figure 1 shows two such spectral envelopes. Plotting the extracted coefficients (see figure 2) as a function of time for a soft-to-loud sequence reveals that the spectral envelopes go from highly-resonant (with resonance peak around the fundamental frequency), to fully damped (with -3dB frequency around three times the fundamental frequency).

The spectral envelope as a function of input amplitude is now known. To use this extracted envelope for timbre-matching, one simply calculates the ratio between the envelope at the desired amplitude and the envelope of the waveform used for synthesis. This ratio is then the transfer function of the expression filter. We now have two second-order spectra of the form of equation (1), $\mathcal{H}_{\text{desired}}$ and $\mathcal{H}_{\text{waveform}}$. The required filter is simply

$$\begin{aligned}\mathcal{H}_{\text{filter}}(s) &= \frac{\mathcal{H}_{\text{desired envelope}}}{\mathcal{H}_{\text{waveform envelope}}} & (3)\\ &= \frac{A_d\omega_d^2}{A_w\omega_w^2}\,\frac{s^2 + 2\xi_w\omega_w s + \omega_w^2}{s^2 + 2\xi_d\omega_d s + \omega_d^2} & (4)\end{aligned}$$

We know that $\mathcal{H}_{\text{waveform}}$ is fixed, so we can pre-apply the inverse of this spectrum to the stored waveform, such that only

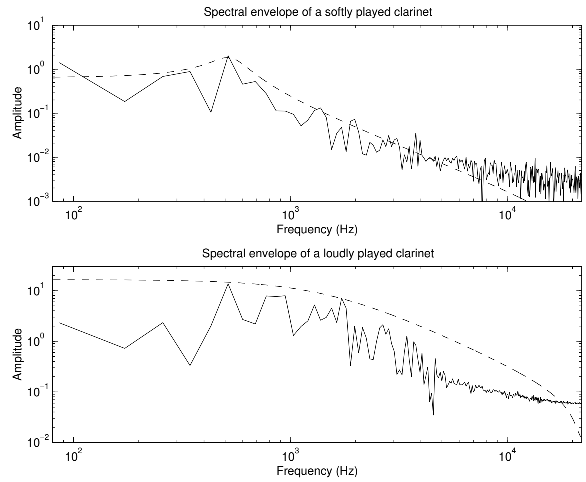

Fig. 1. The second-order envelope spectrum of equation (1) (dashed line) fitted to the spectrum of a softly and loudly played clarinet (full line).

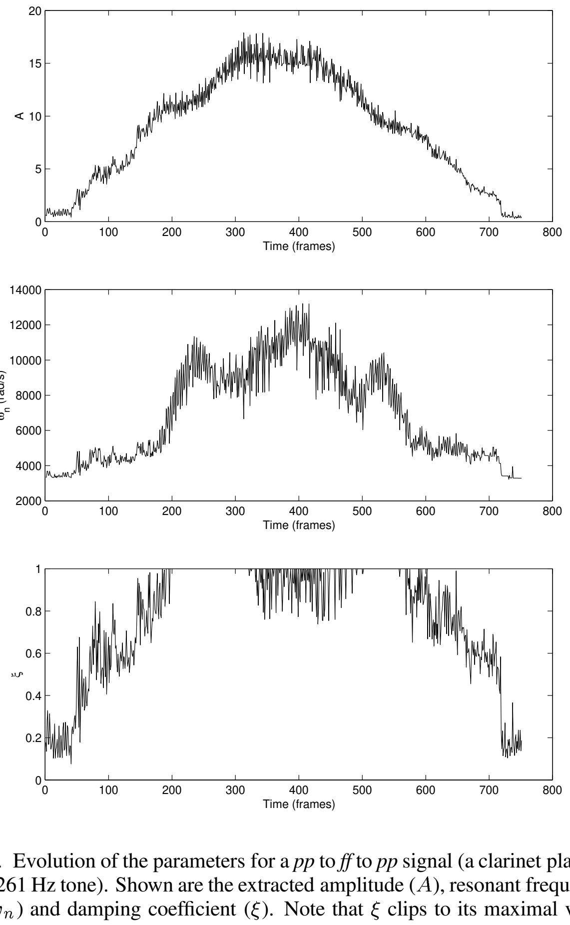

Fig. 2. Evolution of the parameters for a *pp* to *ff* to *pp* signal (a clarinet playing a 261 Hz tone). Shown are the extracted amplitude (A), resonant frequency (ω_n) and damping coefficient (ξ). Note that ξ clips to its maximal value (1) for the highest amplitudes. The frames are 1024 samples or 23.2 ms long (44.1kHz sampling rate), the complete signal has 17.4 s duration.

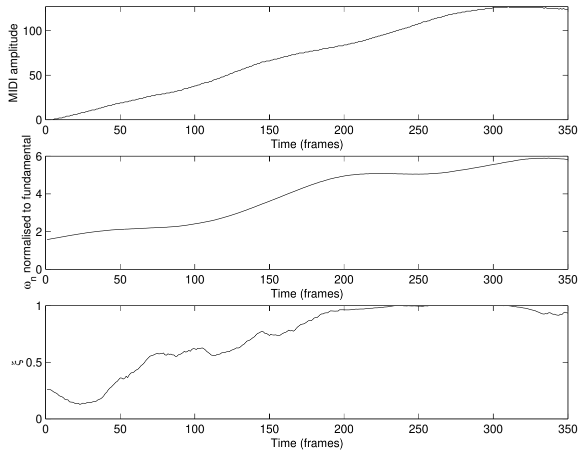

Fig. 3. The extracted parameters of figure 2 are now smoothed and normalised. The amplitude was converted to the MIDI amplitude parameter (a 7-bit value). ω_n was normalised with respect to the fundamental frequency of the note being analysed (in this case 261 Hz)

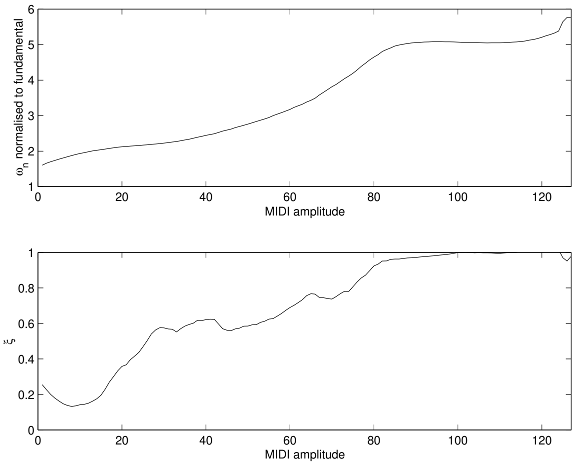

Fig. 4. The expression parameters ξ and ω_n as a function of MIDI amplitude

the effect of $\mathcal{H}_{\text{desired}}$ has to be calculated during synthesis. Note that pre-filtering the waveform enhances the high frequencies, which may deteriorate the signal-to-noise ratio so some precautions have to be taken.

The extracted filter parameters can be stored in a table or a parametric description of the coefficients can be found. The first step is to smooth the parameters. In this work, we used a Singular Value Decomposition-based method to eliminate the noise on the extracted values. The result of this step is shown in figure 3. When used for realtime performance on a synthesiser, it is useful to determine the enveloppe parameters ξ and ω_n as a function of the amplitude. This amplitude is expressed as a 7-bit value (as used within the MIDI specification). Figure 4 shows the parameters.

A digital implementation of the filter can then be sought, taking care of transient distortion [5], [6]. One can now determine the lookup tables for the coefficients of the digital filter as a function of input amplitude and the chosen filter structure.

The synthesis engine now selects the appropriate expression filter depending on the input. Such a filter change is only

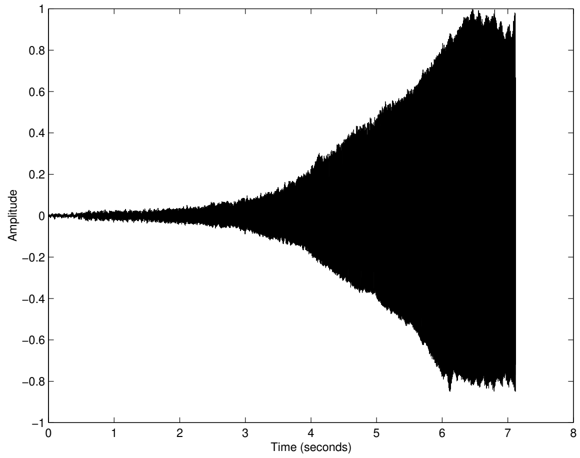

Fig. 5. The recorded refence signal to which the synthetic signals are compared. The amplitude of this signal is used as the desired amplitude during synthesis.

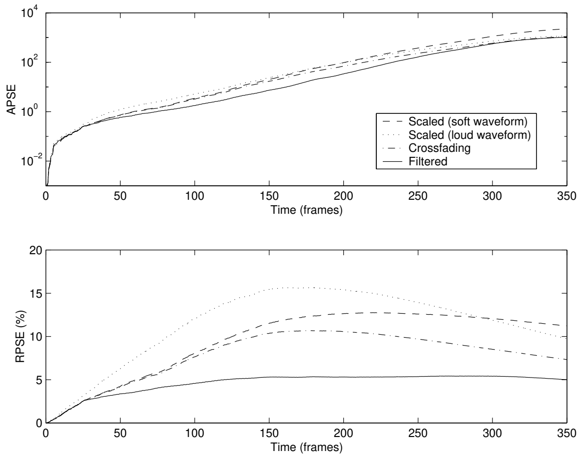

Fig. 6. The APSE (top) and RPSE (bottom) for the four test signals *ScaledSoft* (dashed line), *ScaledLoud* (dotted line), *Crossfading* (dash-dot line) and *Filtered* (full line). One time frame is 23.3 ms.

needed when an input change is detected, typically once per several hundred calculated audio samples.

III. Comparison

To assess the subjective and objective quality of this method, it was compared to the simple amplitude change and the crossfading method, as well as to a live played, soft-to-loud reference recording.

A. Generation of the signals

For the calculation of the simple amplitude change, we started from a recording of a softly played clarinet (the soft generating waveform). This waveform was looped and progressively amplified such that it followed the loudness change of the reference recording (fig. 5). This is the *ScaledSoft* signal. The procedure was repeated but starting from a recording of a loudly played clarinet (*ScaledLoud* signal).

The crossfading method uses more than one waveform to generate the desired tone. Wavetable crossfading between two or more recorded tones requires careful adjustment of pitch and phase relations between the tones; if not a chorus effect will be audible. More advanced techniques, such as Principal Component Analysis [4], have been elaborated for the selection of the waveforms used. Generally, the more tones used, the more precise the output, but at the cost of more memory and more computation time. We generated a signal (*Crossfading*) based on two recordings (soft and loud) and interpollated such that the spectrum of this signal matches the spectrum of the reference as closely as possible.

The signal generated with the method described in this paper (*Filtered*) is based on the parameters shown in fig. 4 and the same generating waveform as for the *ScaledLoud* signal.

B. Objective comparison

To measure the quality of the synthetic signals, we calculated the absolute and relative difference of the power spectra of the reference and synthetic signals. We define the absolute power spectral error (APSE) as

$$\mathrm{APSE} = \sum_{\mathrm{bins}} |\mathcal{S}^2_{\mathrm{reference}} - \mathcal{S}^2_{\mathrm{synthetic}}| \tag{5}$$

where $\mathcal{S}$ is the amplitude of the spectrum. The relative power spectral error (RPSE) is

$$\mathrm{RPSE} = \frac{\sum_{\mathrm{bins}} |\mathcal{S}^2_{\mathrm{reference}} - \mathcal{S}^2_{\mathrm{synthetic}}|}{\sum_{\mathrm{bins}} \mathcal{S}^2_{\mathrm{reference}}} \tag{6}$$

These can be seen as the absolute and relative power of the error signal.

Figure 6 show the APSE and RPSE for the four signals. We start by comparing the two scaled signals. As was expected, the APSE and RPSE are better in the amplitude range of the waveform used for synthesis. The *ScaledLoud* signal has a lower APSE and RPSE than the *ScaledSoft* signal for the high amplitudes. The inverse is true for the low amplitudes. With *ScaledSoft*, the error grows with increasing amplitude as the original signal lacks the upper harmonics. On the other hand, *ScaledLoud* has too much harmonics when scaled to a low amplitude. Note that both have an RPSE of 10-15% over most of the amplitude range. Both signals would have zero error when the generating waveforms are chosen out of the reference recording and no scaling is applied (i.e. we want a synthetic signal that has the same amplitude as the generating waveform).

The *Crossfading* signal is the best possible interpollation of the soft and loud recordings and has lower error measures (7-10% RPSE). The error is low when the amplitude matches the amplitude of one of the recorded tones and is maximal between these two tones.

The filtering method described in this paper (*Filtered*) has consistently lower errors than the other methods (RPSE below 6% for the complete amplitude range).

C. Subjective comparison

Subjective evaluation of the *ScaledSoft* signal showed that the quality is high at the "natural" loudness of the signal (soft to medium-loud), but low at loud levels. The transition from soft to loud sounds synthetic. The same holds for the *Scaled-Loud* signal. When using the single waveform, simple scaling method, one should select a medium-loud recorind as generating waveform. This ensures that the error is spread out over both loud and soft amplitudes. None of the generated signals with this method sound convincing.

With two-tone interpolation (*Crossfading*), the transition sounds unnatural as the amplitude of the higher partials does not have a linear dependence on input amplitude, but more a saturation-like, non-linear characteristic. This could be improved by using more generating waveforms but this increases the complexity of the interpollation.

The *Filtered* signal has the most convincing transition. It is not an exact match to the reference recording but it comes close. The dynamic timbre-change makes it much less synthetic-sounding.

IV. Conclusion

We presented a method based on parametric filters for dynamic amplitude and timbre-change in musical instrument sound synthesis. The filter coefficients were determined by analysis of an instrument recording. The filtering method subjectively sounds better than the simple amplification method, and the absolute and relative power spectral errors are lower. The soft-to-loud transition sounds more natural than in two-tone crossfading. The computational cost is low.

The comparison shows that the dynamic-input sound quality of single-wavetable sound synthesis can be improved at a very low global cost.

V. Acknowledgements

Axel Nackaerts is a Research Assistant with the I.W.T. (Flemish Institute for Scientific and Technological Research in Industry). Bart De Moor is a senior Research Associate with the F.W.O. (Fund for Scientific Research-Flanders) and a Professor Extra-Ordinary at the Katholieke Universiteit Leuven. Rudy Lauwereins is Professor at the K.U.Leuven. This research work was carried out at the ESAT laboratory of the Katholieke Universiteit Leuven. This work is supported by several institutions: the Flemish Government (GOA-MIPS and Mefisto-666), the FWO, and Research Communities (ICCoS and ANMMM), IWT projects (EUREKA 1562-SINOPSYS, EUREKA 2063-IMPACT, STWW), the Belgian State, Prime Minister's Office - OSTC - (IUAP P4-02 1997-2001 and IUAP P4-24 1997-2001, Sustainable Mobility Programme - Project MD/01/24 1997-2000), the European Commission (TMR Networks: ALAPEDES and System Identification, Brite/Euram Thematic Network : NICONET) and by Texas Instruments (TI-Elite program). ESAT-K.U.Leuven is member of DSP Valley. The scientific responsibility is assumed by the authors.

References

[1] C. Roads, *The Computer Music Tutorial*. The MIT Press, 1996.

[2] N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments*. New York: Springer-Verlag Inc., 1991.

[3] T. Tolonen, V. Välimäki, and M. Karjalainen, "Evaluation of modern sound synthesis methods," Tech. Rep. 48, Helsinki University of Technology, Espoo, March 1998. ISBN 951-22-4012-2, ISSN 1239-1867.

[4] A. Horner, J. Beauchamp, and L. Haken, "Methods for multiple wavetable synthesis of musical instrument tones," *Journal of the Audio Engineering Society*, vol. 41, pp. 336–355, May 1993.

[5] J. N. Mourjopoulos, E. D. Kyriakis-Bitzaros, and C. E. Goutis, "Theory and real-time implementation of time-varying digital audio filters.," *Journal of the Audio Engineering Society*, vol. 38, pp. 523–535, July/August 1990.

[6] Y. Ding and D. Rossum, "Filter morphing of parametric equalizers and shelving filters for audio signal processing," *Journal of the Audio Engineering Society*, vol. 43, pp. 821–825, October 1995.