ENVELOPE MODEL OF ISOLATED MUSICAL SOUNDS

Kristoffer Jensen

University of Copenhagen
Department of Computer Science
Universitetsparken 1, DK-2100 Copenhagen Ø, Denmark
krist@diku.dk

ABSTRACT

This paper presents a model of the envelope of the additive parameters of isolated musical sounds, along with a new method for the estimation of the important envelope split-point times.
The model consists of start, attack, sustain, release, and end segments with variable split-point amplitude and time. The estimation of the times is done using smoothed derivatives of the envelopes.
The estimated split-point values can be used together with a curve-form model introduced in this paper in the analysis/synthesis of musical sounds. The envelope model can recreate noise-less musical sounds with good fidelity, and the method for the estimation of the envelope times performs significantly better than the classical percentage-based method.

1. INTRODUCTION

The envelope is the evolution over time of the amplitude of a sound. It is one of the important timbre attributes. The envelope model presented in this paper is relatively simple, as compared with the piece-wise linear model [1], having only 4 split-points. The five segments of the model are the start, attack, sustain (or decay), release and end segment.
The model introduced in this work combines the intuitive simplicity of the envelope model with the flexibility of the additive model. The idea is to model the amplitude of each partial as four time/value pairs. Furthermore, the interval between each split-point is modeled by a curve the quality of which (exponential/logarithmic) can be varied with one parameter. All split-point are variable and in particular, by using variable start of release split-point amplitude, both sustained and percussive sounds can be modeled. The envelope model can be seen in figure 1.
A new method introduced here finds the envelope split-point times by analyzing the derivatives of the amplitude. The derivative of the envelope has been shown to be correlated with the perceptual attack time of musical sounds [2].

This method performs significantly better than the classical percentage-based method, in which the times are found by looking at the first (and last) time the amplitude is above a threshold. The method presented here permits, for instance, to estimate correctly the start of release times of the piano partials.
The envelope model can be used in synthesis of musical sounds, and the envelope model parameters can be used in research involving the perception of sounds [3], in the analysis of the evolution of the envelope parameters [4] and in the classification of musical sounds [5]. The envelope model is also suitable for use in timbre morphing [4].

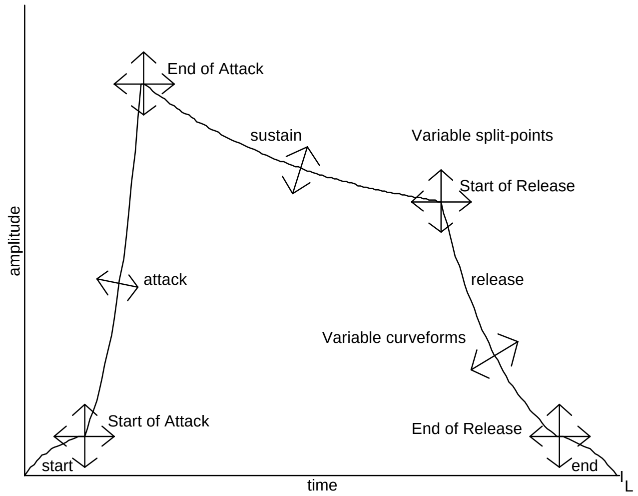

Figure 1. The envelope model.

This paper is organized as follows: first the additive parameters are defined, then the estimation of the attack and release times is presented, then the curve form model of the segments is introduced, the envelope reconstruction is outlined, and finally a conclusion is given.

2. ADDITIVE PARAMETERS

The additive parameters are analyzed with the linear time frequency method introduced in [6]. Only quasi-harmonic components are analyzed, and the amplitudes and frequencies are estimated once each period. The model of the sounds is

$$snd(t) = \sum_{k=1}^{N_p} amp_k(t) \cdot \sin(2\pi freq_k t), \qquad 0 \le t \le L. \qquad (1)$$

The sound is thus the sum of a number of sinusoidals, with static frequencies and time-varying amplitudes. The sounds are actually analyzed with time-varying frequencies, but these are not used in this paper. The resynthesis of the sounds using formula (1) with time-varying frequencies has been shown to have an impairment between *imperceptible* and *perceptible, but not annoying* [4].

3. SPLIT-POINT TIME ESTIMATION

The split-point time estimation is done individually on each of the partials of the sound. Assuming that there exists N_p partials, each defined for *0• t• L*, and that furthermore,

$$\begin{aligned} &amp_k(t) \rightarrow 0, t \rightarrow \pm\infty \\ &amp_k(t) \ge 0, 0 \le t \le L. \end{aligned} \qquad (2)$$

Each non-zero envelope thus has a positive slope (attack) and a negative slope (release). This section presents a estimation method of the start and end of the attack and release.

The split-point time estimation is done on smoothed envelopes. The smoothing is performed by convoluting the envelope with a gaussian,

$$env_\sigma(t) = amp_k * g_\sigma(t), \qquad g_\sigma(t) = \frac{1}{2\pi\sigma} e^{-\frac{t^2}{2\sigma^2}}. \qquad (3)$$

The middle of the attack and release are now found by finding the maximum and minimum of the time derivative of the smoothed envelope,

$$\frac{\max}{\min} L_{t,\sigma}(t), \qquad L_{t,\sigma}(t) = \frac{\partial}{\partial t} env_\sigma(t) \qquad (4)$$

The start and end of the attack and release is found by following $L_{t,\sigma}$ forwards and backwards (in time) until it is close to zero (about one tenth of the maximum derivative). The start of release threshold is greater than the other three thresholds to estimate correctly the start of release in the common decay-release envelope type, as seen in the piano for instance.

The resulting envelope times for the fundamental of four test sounds, viola (left), piano, trumpet and flute (right), can be seen in figure 2. The smoothed envelope is shown top, and the derivative is shown bottom.

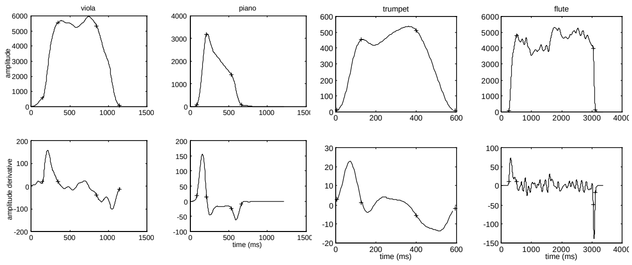

Figure 2. Smoothed fundamental envelopes (top) and time derivative (bottom) for the viola, piano, trumpet and flute.

The envelope is then followed from the smoothed to the unsmoothed case in several steps by a method borrowed from the scale-space theory used in image processing [7]. This procedure is outlined in the following paragraphs.

The local maximums and minimums of the second derivative of the envelope are found, typically by searching the zero crossing of the third derivative,

$$L_{ttt,\sigma}(t) = 0, \qquad L_{ttt,\sigma}(t) = \frac{\partial^3}{\partial t^3} env_\sigma(t) \qquad (5)$$

This corresponds to the end points of a slope, as can be seen in figure 3.

When the envelope is less smoothed, the slope is steeper, and the slope points correspond more to the unsmoothed case. In the unsmoothed case, there are typically many points. It is thus necessary to use enough smoothing steps so the slope points can be followed.

The times are adjusted after each smoothing step so it doesn't occur in the middle of the slope, or in a local minimum. Furthermore, if the slope point is chosen from many candidates, the closest to the middle of the attack (or release) is selected. This ensures that the attack and release get shorter in the unsmoothed case, as they should.

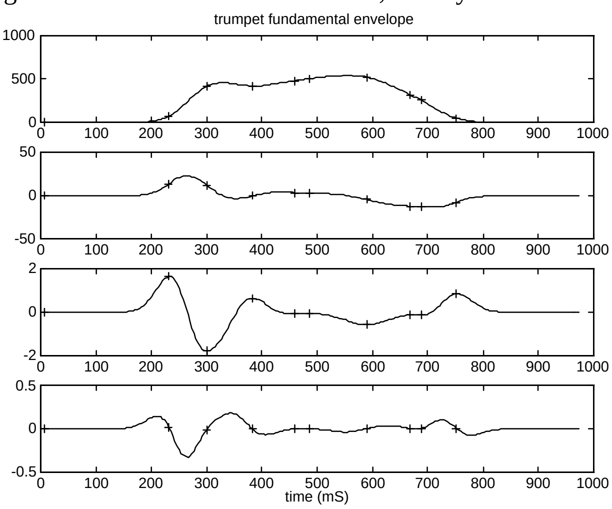

Figure 3. Envelope of smoothed trumpet fundamental (top) and first three derivatives. The zero-crossing of the

third derivative (bottom) correspond to the end points of the slopes, denoted with '+'.

The resulting times of the four test sound can be seen in figure 4. The start of attack is the lowest curve in each subplot, then follows the end of attack, the start of release, and on top the end of release.

The envelope times are rather stable across the partial index. It is clear that the difficult start of release has been correctly estimated for most of the partials of the piano sound. Furthermore, the flute sounds sustain is approximately correct for all partials, and generally the split-point times found correspond well to the theoretical times.

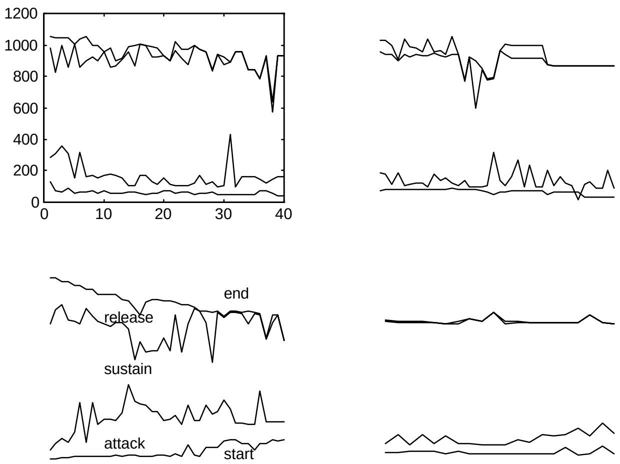

Figure 4. Envelope split-point times of four test sounds. The y-axis is times, from start of attack (bottom line), end of attack, start of release to end of release (top line), and the x-axis is partial index. The segment names are shown in the trumpet subplot.

In addition to the envelope times, the amplitude value is retained in each split-point. The amplitudes at time $t=0$ and $t=L$ are assumed to be zero and not measured. The relative amplitudes (amplitude of split-point divided by the maximum amplitude of the partial) can be seen in figure 5.

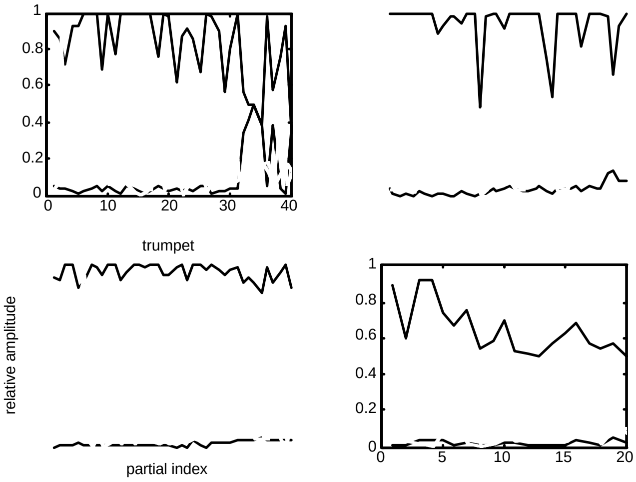

Figure 5. Relative amplitudes of the split-points of the four test sounds for the attack (solid) and the release (dotted). The sustain split-points are generally close to 1.

The envelopes found with the percentage-based method are shown in figure 6. The thresholds are 0.1 for the start of attack and end of release, 0.9 for the end of attack and 0.7 for the start of release.

It is clear that the times are more noisy than for the slope-based method, and the start of release estimation has failed for the piano sound. The sensibility to noise can clearly be seen in the flute sound, where most of the partials end of attack are found in the middle of the sound.

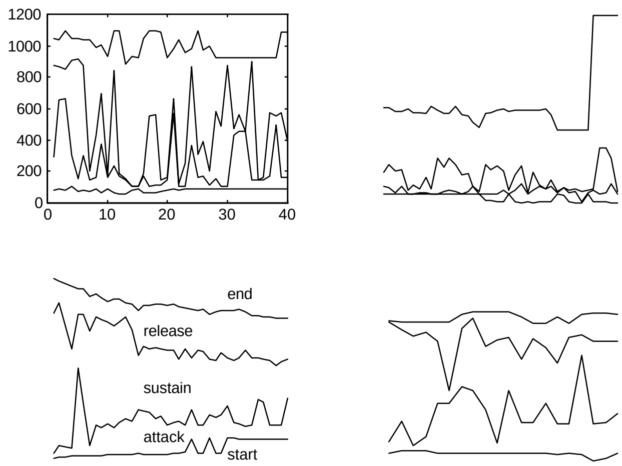

Figure 6. Envelope split-point times found with the percentage-based method. The segment names are shown in the trumpet subplot.

4. CURVE FORM

The slope of the envelope curve between split points is supposed to rise (or fall) continuously. Therefore, the slope

is modeled by a curve whose curve form (exponential/logarithmic) can be set with one parameter.
The curve used for the modeling of the envelope for one segment is the exponential curve,

$$\hat{env}(t) = v_0 + (v_1 - v_0)\frac{n^{t/L_s} - 1}{n - 1}, \qquad 0 \le t \le L_s \tag{6}$$

with v_0 and v_1 the start and end values, L_s the segment length, t the time and n the curve form coefficient, which is found by minimizing the least-square error for the segment,

$$Error = \sum \left(\hat{env}(t) - env_0(t) \right)^2 . \tag{7}$$

The curve-fitting problem is nonlinear, and the Levenberg-Marquardt method [8] is used to solve it. The initial values for the curve fit are found in the log domain of equation (6) using LMS [9].

5. ENVELOPE RECONSTRUCTION

The envelope can now be recreated, using the start, attack, sustain, release and end split-point times, amplitudes and curve forms and concatenating the five segments created using formula (6)**.**
The recreated envelopes for the fundamental of four sounds, viola, piano, trumpet and flute, can be seen in figure 7.
The resulting envelopes have the same shape as the originals, but most irregularities are no longer present. Especially the noise of the flute fundamental is missing, and this is also the perceptually most important difference.

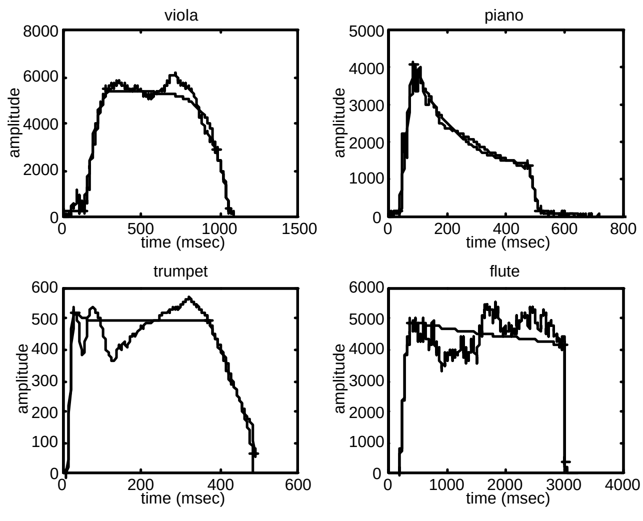

Figure 7. Fundamental envelope, parameters and recreated envelope for four sounds. Split-points are denoted with a '+'.

The original additive partials of the four test sounds, viola, piano, trumpet, and flute are shown in the upper plots of figure 8 and **Error! Unknown switch argument.**, and the recreated additive parameters are shown in the lower plot.

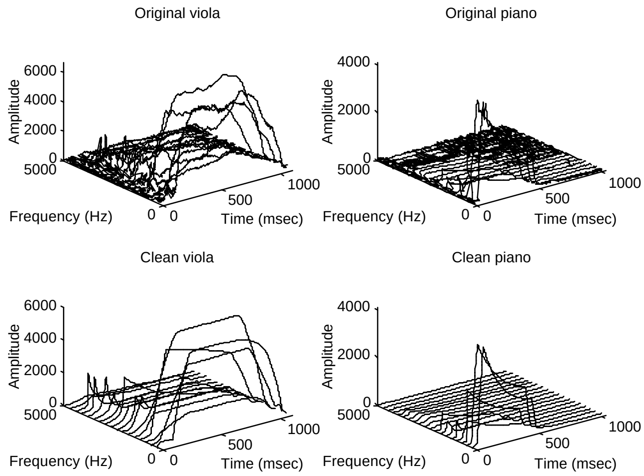

Figure 8. Original (top) and recreated additive parameters (bottom) for the viola and the piano.

Informal listening tests reveal that the sounds recreated with the clean envelopes, using static frequencies, inherit most of the timbre quality of the original sounds, except for the noise of the flute sound, which sounds more dull with the clean additive parameters. Some of the liveliness is of course lost in the simplification of the sounds, but most of the quality is retained nonetheless.

6. CONCLUSIONS

This paper has presented a simple envelope model of the additive parameters of musical sounds. A novel method of estimating the envelope split-point times, using the derivative of the envelope is introduced.
The envelope time estimation presented here is more accurate and less noise sensitive than the percentage-based method commonly used. In particular, it permits the proper estimation of the start of release time in a decay-release envelope from, for instance, the piano. Resynthesis of sounds using the envelope model presented here yields results close to the original sounds, as long as the sounds contain little noise.
The envelope model forms the basis of a timbre model, with impairment *perceptible, but not annoying* as compared with the original sounds [4]. The envelope parameters are also instrumental in the classification of musical sounds [5].

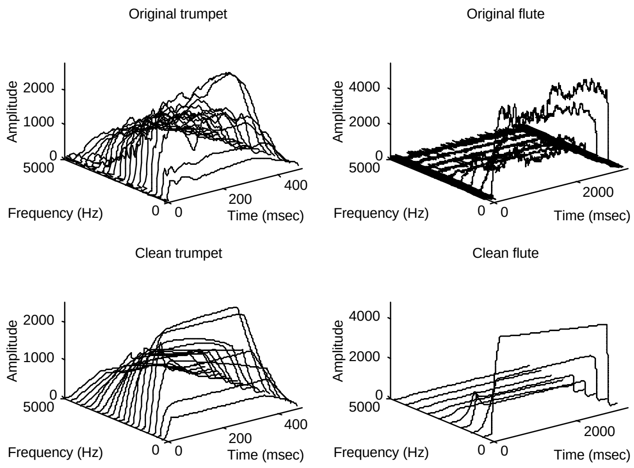

Figure 9. Original (top) and recreated additive parameters (bottom) for the trumpet and the flute.

7. REFERENCES

[1] A. D. Bernstein, E. D. Cooper, "The piecewise-linear technique of electronic music synthesis". *J. Audio Eng. Soc*. Vol. 24, No. 6, July/August 1976.

[2] J. W. Gordon, "The perceptual attack time of musical tones". *J. Acoust. Soc. Am*. 82(2), July 1987.

[3] S. McAdams, S. Winsberg, S. Donnadieu, G. de Soete, J. Krimphoff, "Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes". *Psychological Research*, 58, 1992.

[4] K. Jensen, *Timbre Models of Musical Sounds*. Ph.D. dissertation, Department of Computer Science, University of Copenhagen, DIKU Report 99/7, 1999.

[5] K Jensen, "Binary tree classification of musical sounds". *Proceedings of the ICMC*, 1999.

[6] P. Guillemain, R. Kronland-Martinet, "Characterization of acoustic signals through continuous linear time-frequency representations". *Proc. of the IEEE*, 84(4), April, 1996.

[7] T. Lindeberg, "Edge detection and ridge detection with automatic scale selection", *CVAP Report*, KTH, Stockholm, 1996.

[8] J. J. Moré, "The Levenberg-Marquardt algorithm: Implementation and theory". Lecture notes in mathematics, Edited by G. A. Watson, Springer-Verlag, 1977.

[9] H. R Schwarz, *Numerical Analysis*, Wiley, 1989.