

RESEARCH PROPOSAL

Expressive Musical Descriptors Extraction

Esteban Maestre

May 2004

Music Technology Group
Audiovisual Institute - Pompeu Fabra University
Ocata, 1. 08003 Barcelona, SPAIN
<http://www.iaa.upf.es/mtg>
emaestre@iaa.upf.es

Abstract

This document summarizes the research purposes and objectives to be worked as a part of DEA studies in Computer Music intended to be followed within the doctoral program in Computer Science and Digital Communication offered by the Audiovisual Institute of the Pompeu Fabra University, dealing with expressive audio transformations. First section introduces the most relevant descriptors suitable for defining the expressivity that a musician introduces when performing a musical piece, and proposes different approaches for extracting them. Section 2 goes into more detail for the signal energy approach when extracting some note intensity descriptors and presents some results to be obtained. The third section describes the further work on the approach presented in section 2 as a research proposal. Finally, the fourth section concludes with a statement of the research intentions following the line described along this document.

1. Introduction

In the study of the expressivity of music performance, many different parameters related to the audio signal may become important descriptors when trying to analyze, transform and synthesize audio signals following concrete expressive patterns [1], [6]. With the aim of extracting high level musical descriptors that help to perform the expressive analysis/synthesis closer to the context of music rather than to the signal processing context, low-level musical descriptors must be extracted in order to combine them into a high level performance model.

Considering the low-level musical descriptors that can be related to the musical performance, many of them might be classified into two main groups considering their context: the *intra-note* descriptors and the *inter-note* descriptors. The *intra-note* descriptors concern the characteristic features that a musician performs within a note, whilst the *inter-note* descriptors define the relation between two or more notes in the context of the performance and musical phrase.

- **Intra-note descriptors:** Concerning a signal intensity approach, descriptors like note *onset* and *offset*, time and level of *attack*, *decay*, *sustain* and *release* parts, *tremolo*, etc. should be taken into account. When analyzing frequency modulation over time, descriptors like *vibrato*, *glissando* and *portamento*¹ should be considered.

¹ *Glissando* and *portamento* might be considered as both *intra-note* and *inter-note* descriptors.

- **Inter-note descriptors:** Considering intensity evolution of the notes within a musical phrase, *crescendo* and *decrescendo* might be extracted. Descriptors like *legato* and *stacatto* shall be considered when studying the smoothness of the connection of successive notes, while descriptors like *rubato* and *marcato* are more related to the duration of notes and the sense of meter.

The study of all these descriptors from a signal processing point of view will include both the temporal analysis of the *signal energy* and the *fundamental frequency*, and, in many cases, a combination of these two main approaches will be needed in order to determine some of the descriptors already presented. The following section introduces a first approach for extracting some intra-note descriptors dealing with the signal energy over time.

Of course, the interrelation between the intra-note and inter-note descriptors for a concrete musical and performance context will be a very important issue when modelling the expressive transformation pattern to be followed.

2. Intra-note intensity descriptors: the signal energy approach

Going into more detail on the signal intensity approach when extracting intra-note descriptors, this section introduces a method for extracting the segment-split points of a musical note based on the study of the signal energy envelope and its curvature characteristic points. Some related work dealing partial amplitudes is related in [3].

2.1. Note segments

Perceptually, the limits defining the segments that an ideal ‘generic’ note is composed of are depicted in figure 1. Depending on many factors, such as the instrument type or the performance, other types of envelopes with slight variations in the number and localization of the points defining them might be considered. This first approach will deal with the standard ADSR definition of the note characteristic points fitting the ideal energy envelope of figure 1.

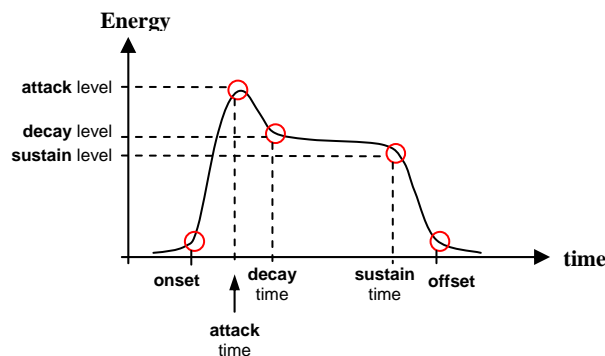


Figure 1. Characteristic points of a note's energy envelope

The importance of the extraction of these points is very clear for any expressive transformation of the note being considered. For instance, when performing a note duration transformation using SMS time-stretch techniques [2], it will be important to know accurately the time points defining the sustain part in order to lengthen or shorten

the note in between these limits, since it is in many cases the note segment that defines well the note duration.

2.2. Onset, offset and ADSR times extraction

Considering the energy as a differentiable function over time, these points of maximum curvature can be considered as the local maximum variation of the first derivative of the signal energy (first derivative inflexion points), i.e. the local maxima and minima of the second derivative [3], [5]. In this way, finding the zero-crossings of the third derivative will allow to localize these characteristic points. Figure 2 shows the ideal energy shape presented in figure 1 and its third time-derivative.

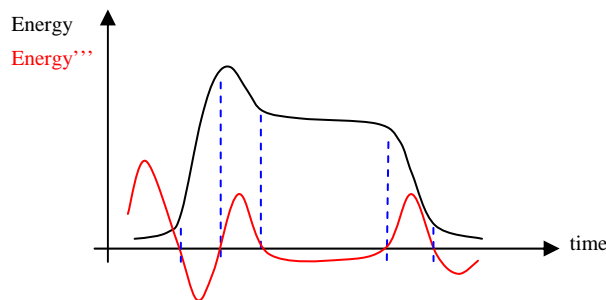


Figure 2. Zero-crossings of the third derivative of the energy envelope

By observing the zero crossings of the third derivative, can be seen how they indicate the characteristic points defining the segments described in the previous section.

2.3. Multi-resolution analysis

Considering the energy evolution of a real audio signal over time as the frame-by-frame energy computation using Short-Time Fourier Transform, it results in a noisy signal that would not let the characteristic points being clearly distinguished within the whole conjunct of the zero-crossings of the third derivative.

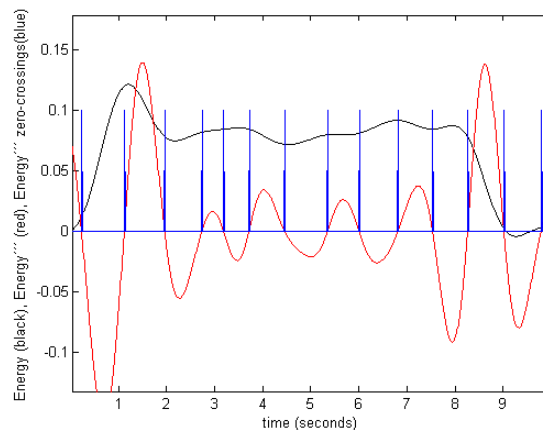


Figure 3. Zero-crossings of the third derivative of a real sax note energy envelope

This can be observed in figure 3, where the energy of a sax note recording and its third derivative are depicted. For a clear representation, the energy has been previously low-pass filtered at a cutoff frequency $f_0 = 400\text{Hz}$.

If only the most relevant zeros of the third derivative of the energy envelope must be found, the signal energy should be low-pass filtered in order to overcome the problem of having too many zero-crossings. The cutoff frequency must be selected as low as it is needed for having just the five relevant points defining the note segments described in the previous sections². This has been carried out for another real sax note recording, as it is shown in figure 4.

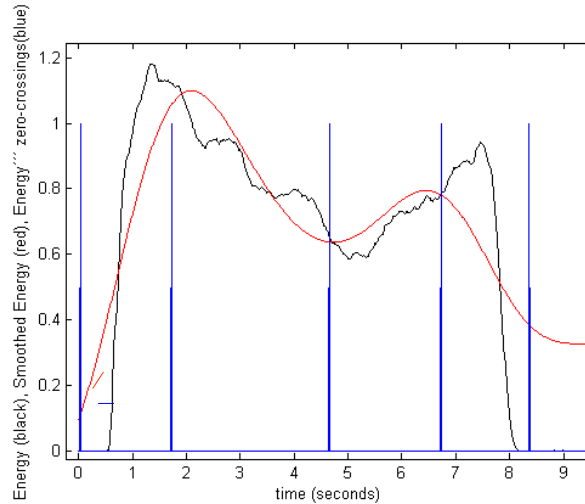


Figure 4. Smoothed energy and its third derivative zero-crossings

By projecting the zeros of the third derivative of the smoothed energy onto the real energy shape, it is clearly seen how these points do not correspond with the interesting points of the energy signal due to the filtering. The number of inflexion points of the energy curve is increasing as much as the cutoff frequency of the smoothing filter gets higher. These zeros may change their time location between two successive resolutions.

Because of this, a multi-resolution analysis is proposed in order to follow the relevant points from the lowest resolution until the one for which they sufficiently coincide with the ones that can be marked perceptually in the real signal energy, as it can also be concluded from the observation of figure 5, where frame positions of the zero-crossings of the third derivative are illustrated as a black square (bottom graph) at 120 different resolutions, for cutoff frequencies in the interval $100\text{Hz} < f_0 < 500\text{Hz}$. The paths followed by these zero-crossings through resolutions can be easily seen as the curves going from the lowest resolutions up to the highest. Their projection onto the energy curves has been depicted using circles in the top graph for two different resolutions, as define the dashed horizontal lines in the bottom graph. The red curve represents the real energy without smoothing filter and the green and blue curves are the smoothed energy at cutoff frequencies of, respectively, $f_0 = 190\text{Hz}$ and $f_0 = 500\text{Hz}$. The black circles in the bottom graph represent the starting points, for a cutoff frequency of $f_0 = 100\text{Hz}$.

The start (lowest) resolution is chosen such that just the five most relevant points defining the note segments appear, while the finish (highest) resolution is selected such that the projections onto their smoothed energy curve sufficiently coincide with the ones that would be marked manually in the real signal energy.

² Other criteria for the selection of the number of points to be followed would be considered depending on the energy shape intended to be modelled.

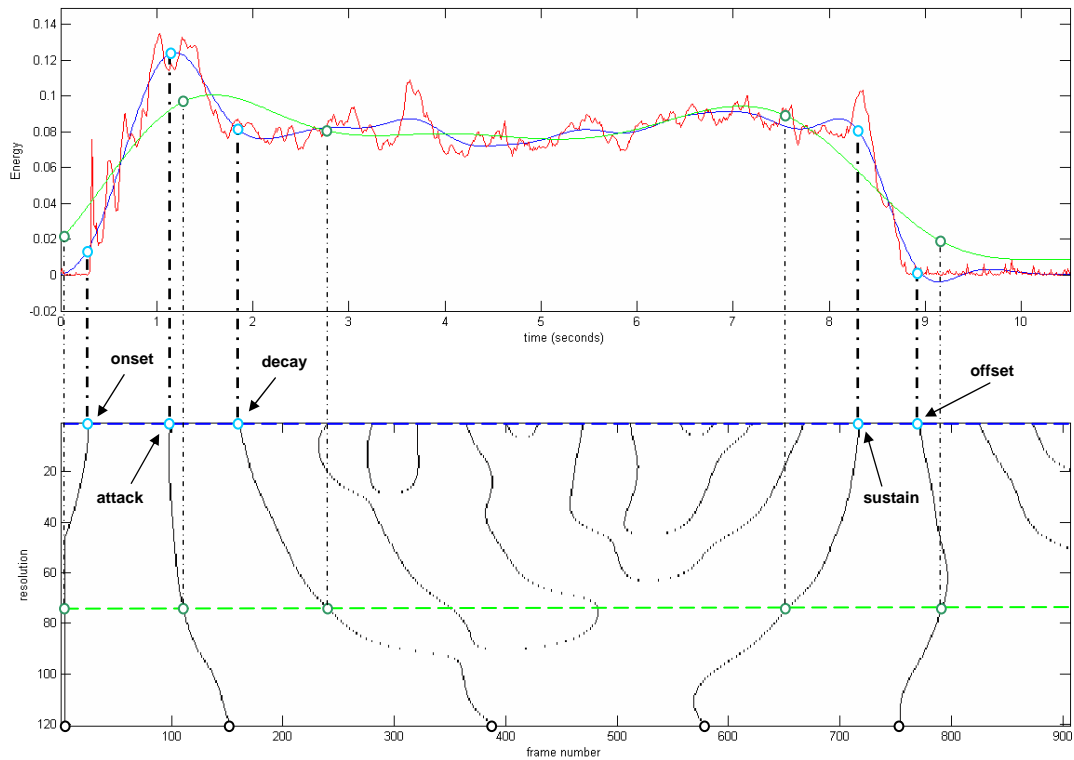


Figure 5. Time-resolution tracking of the zero-crossings of the third derivative of the energy signal

In this way, by means of a good and efficient tracking algorithm, very useful and accurate information about the intra-note descriptors related to the intensity of the sound can be obtained. An important issue when performing this multi-resolution analysis is the computational cost, due to the complexity of the operations involved. A good selection of the initial and final resolutions, as well as the resolution step would lead to a more efficient tracking.

3. Further work

Following the line presented in the previous section, several research directions remain open towards the extraction of note expressive descriptors:

- **Tracking criteria:** when performing the tracking of the zero-crossings of the third derivative of the Energy signal through the different resolutions, several criteria -not only based on the frame distance between contiguous resolutions³- are to be studied with the intention of selecting always the correct path up to the characteristic points location in the higher frequencies definition.

³ For low amount of different smoothing filters, i.e. big resolution steps, tracking based only in the frame distance fails to select the correct path.

- **Limit frequencies selection:** in order to select the most suitable initial and final resolution accomplishing with the requirements stated in the previous section, a pre-processing Fourier analysis of the frame-by-frame energy curve might be performed. This could lead to a determination of the limit cutoff frequencies and resolution step for the filter bank in the resolution analysis implying computationally faster and cheaper tracking.
- **Individual partial contribution:** Other approaches analyzing the contribution of each partial instead of treating the whole signal might be studied.
- **Wavelet Transform approach:** Another approach giving the same information, but permitting the reconstruction of the energy signal could be studied by means of Wavelet Transform. Filtering can be seen as the convolution of the energy signal with a smoothing function $g(t)$. The derivative of this convolution [3] is equal to the convolution of the signal energy with the derivative of the smoothing function.

$$\frac{d}{dt}(e(t) * g(t)) = e(t) * \frac{d}{dt}(g(t))$$

The smoothing function will be dilated or contracted for each resolution intended to be analyzed. The Wavelet Transform [4] can be considered as the convolution with a finite smoothing function⁴ at different scales (or contractions and dilatations). If the mother wavelet is chosen as the third derivative of the smoothing function $g(t)$, the zero-crossings of the Wavelet Transform [5] would give the same information as the multi-resolution analysis presented before, but with the possibility of reconstructing the energy curve shape after, for instance, transforming some of its characteristic points.

- **High level expressive descriptors extraction:** the extension of the signal energy approach to a conjunct of notes would let to extract some useful inter-note descriptors. Combination of these descriptors with those that can be obtained from a fundamental frequency approach will lead to the extraction of high level expressive descriptors.
- **Synthesis:** with the aim of actually performing the expressive transformation, it should be studied how to re-synthesize the sound from the new descriptors extracted and transformed.

4. Conclusion

A summary of the research purposes and objectives to be worked as a part of DEA studies in Computer Music dealing with expressive audio transformations has been outlined in this document. First, the problem has been presented, and the most relevant descriptors suitable for defining the expressivity that a musician introduces when performing a musical piece have been introduced. Different approaches for extracting some of them have been proposed. The signal energy approach when extracting some note intensity descriptors is described in more detail while giving some results on the

⁴ Accomplishing some mathematical requirements outlined in [4].

hypothesis explained. Remaining open research directions, as extensions of the current approach have been introduced as a research proposal in the context of expressive audio transformation dealt within the project ProMUSIC, currently developed in the Music Technology Group of the Pompeu Fabra University.

5. References

- [1] E. Gómez, M. Grachten, X. Amatriain, J. L. Arcos: "Melodic Characterization of Monophonic Recordings for Expressive Tempo Transformations" Proceedings of the Stockholm Music Acoustics Conference SMAC03, Stockholm 2003
- [2] Serra X., Bonada J.: "Sound Transformations based on the SMS High Level Attributes", Proceedings of the DAFx98, Barcelona 1998
- [3] K. Jensen: "Envelope model of isolated musical sounds" Proceedings of the 2nd COST G-6 Workshop on Digital Audio Effects DAFx99, Trondheim 1999
- [4] A. Graps., "An Introduction to Wavelets" IEEE Computational Science and Engineering vol. 2, num. 2, 1995
- [5] Stephane G. Mallat, "Zero-crossings of a wavelet transform" IEEE Trans. Information Theory, vol.37, no.4. pp.1019-1033, 1991
- [6] Josep-Lluís Arcos, Ramon López de Mántaras, X. Serra, "Saxex: A case-based reasoning system for generating expressive musical performances", Journal of New Music Research 27 (3), 1998, 194-210.