Partial Envelope Analysis with Nonnegative Matrix Factorization

Joseph Timoney[1], Tom Lysaght[1], Victor Lazzarini[2] And Lorcan MacManus[3]

[1] *Dept. of Computer Science,* [2] *Department of Music, NUI Maynooth, Maynooth, Co. Kildare,*
[3] *Dublin Institute of Technology, Kevin St., Dublin, Ireland*
contact author: jtimoney@cs.may.ie,

Abstract. This paper investigates the application of the technique of Nonnegative Matrix Factorisation (NMF) to the analysis of partial envelopes extracted from a variety of instrumental sounds. Firstly, the initialisation of the NMF method is shown to be important for obtaining useful results. This is followed by an experimental analysis of the bases produced by NMF that demonstrates how they represent prototypical signal features.

I. Introduction

Nonnegative Matrix Factorization (NMF) is an algorithm designed to factorise a data matrix under non-negativity constraints [1]. It is an unsupervised technique that represents a high-dimensional input by the linear superposition of a user-defined number of basis function in which both the linear coefficients and the basis functions are constrained to be non-negative. Because the non-negativity constraint allows only an additive combination of bases, NMF is said to give a parts-based representation of the original dataset [1]. For example, applied to images of faces, NMF learns basis functions that correspond to eyes, noses and mouths, or applied to handwritten digits, it learns basis functions that correspond to cursive strokes. Furthermore, it has also been used for the decomposition of spectrographic information to identify and separate contributions from a number of harmonic sources of different pitches. Such a parts-based decomposition can be appealing when analysing certain types of data as it can be in keeping with a general interpretation held by a large number of observers. The time-varying magnitude envelopes of the partials that result from a sinusoidal analysis [2] of musical instrument sounds fit into this category of data, and therefore NMF may prove to be a useful technique when applied in an analytical context. This paper intends to investigate the output of the NMF algorithm when the input data consists of partial envelopes. Also, to ensure that the results can be generalised the analysis is applied to a number of different instrumental sounds. The structure of the paper is as follow. Firstly, the analysis of partial envelopes is introduced and then the NMF algorithm is described. This is followed by the experimental procedure and the results of the analysis. Lastly, conclusions regarding NMF are drawn.

II. Partial Envelope Analysis

The envelope can be defined as the slowly time-varying evolution of the amplitude of a sound. It is one of the most defining timbral attributes of any sound [3]. Under the assumption that a sound can be modelled as an additive combination of more elementary partials, the overall sound envelope can be interpreted as the additive combination of the envelopes of the partials. The characterisation of partial envelopes is a key feature of models of musical timbre [3]. Envelopes are generally categorised into four parts namely: the 'attack' portion, the 'decay' portion, the 'steady-state' or 'sustain' portion, and the 'release' portion [3], typically known as the ADSR (Attack-Decay-Sustain-Release) description.

Depending on instrument characteristics it is usually appropriate to truncate the model to be ADR or ASR. An ADR model is more suitable to instruments that produce an initial energy burst followed by damping, such as in plucked string sounds, while an ASR model is more suitable for wind and bowed instruments. Although simple the ADSR model is a well-accepted model and is valid because research on timbre perception has identified the existence of distinct perceptual attributes associated with these categorizations [3]. Figure 1 illustrates a prototypical ADSR envelope.

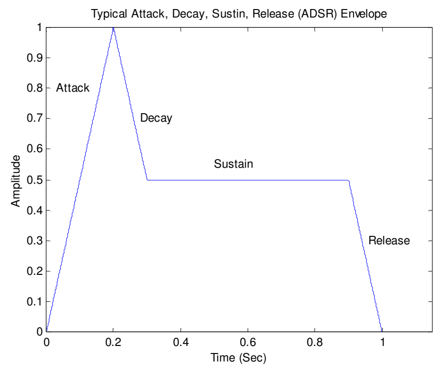

Figure 1: *Prototypical ADSR envelope generated using linear segments*

Because the ADSR model can be interpreted as a part-based representation for partial envelopes, combined with the fact that partial envelopes are magnitude envelopes and thus have the property of positivity, suggests that they would be very suitable for the application of NMF.

III. Nonnegative Matrix Factorization

The goal of NMF is that given a set of observations $\mathbf{y}_t$ at time *t*, the aim is to compute a set of basis functions $\mathbf{W}$ and linear coefficients $\mathbf{x}_t$ such that the reconstructed vectors $\overline{\mathbf{y}}_t = \mathbf{W}\mathbf{x}_t$ best match the original observations [2]. The observations, basis functions and coefficients are all constrained to be non-negative. Firstly, concatenate the vectors $\mathbf{y}_t$, $\overline{\mathbf{y}}_t$ and $\mathbf{x}_t$, separately and denoting them as $\mathbf{Y}$, $\overline{\mathbf{Y}}$ and $\mathbf{X}$ respectively. The dimensions of $\mathbf{Y}$ and $\overline{\mathbf{Y}}$ are $M \times T$, where *M* is the number of observations at each time instant and *T* is the total time. The dimensions of $\mathbf{W}$ and $\mathbf{X}$ are set to be $M \times r$ and $r \times T$ respectively, where *r* is the rank of the factorisation and is representative of the number of parts. Generally it must satisfy the condition that

$$(M+T)r < MT \tag{1}$$

The NMF algorithm is normally set to start from non-negative initial conditions for $\mathbf{W}$ and $\mathbf{X}$, which are then iteratively updated by the following rules [2]:

$$W_{\alpha\beta} \leftarrow W_{\alpha\beta}\left[\sum_t X_{\beta t}\left(\frac{Y_{\alpha t}}{\overline{Y}_{\alpha t}}\right)\right] \quad , \; X_{\beta t} \leftarrow X_{\beta t}\left[\frac{\sum_\alpha W_{\alpha\beta}\left(Y_{\alpha t}/\overline{Y}_{\alpha t}\right)}{\sum_\gamma W_{\gamma\beta}}\right] \tag{2}$$

where α, β and γ denote matrix indices.

Iteration of these update rules converges to the local maximum of the objective function

$$F = \sum_\alpha \sum_t \left(Y_{\alpha t}\log(WX)_{\alpha t} - (WX)_{\alpha t}\right) \tag{3}$$

Monotonic convergence of NMF can be proven using techniques similar to those used in proving the convergence of the EM algorithm [1].

IV. Application of NMF to Partial Envelopes

To study the application of NMF to partial envelope analysis it was decided to discover the extent to which the basis functions characterise the input envelope data. For sound envelopes the most important features are the attack and release, that can be described both by their duration relative to the overall envelope length and their curve coefficient *n* which is a parameter of the envelope segment curve model given by [3]

$$env(x) = v_0 + (v_1 - v_0)\left(1 - (1 - x)^n\right)^{1/n} \tag{4}$$

where *x* represents the time evolution of the envelope segment normalised between 0 and 1, v_0 is the starting value of the segment, v_1 is the final value, and n is the curve coefficient. The value for *n* is first estimated using a least squares regression on the logarithm of the normalised envelope that is then refined using the Levenberg-Marquardt optimisation routine [3].

Analysis data consisted of .AIFF files for Altosaxophone, Eb clarinet and flute from the University of Iowa Electronic Music Studios, sampled at 44100Hz, and recorded at *pp*, *mf* and *ff*. Note separation was required on each file to obtain the individual notes which were then analysed using SMS of Serra to extract the partial envelopes [3]. Post-processing was done to convert the partial envelopes from dB to linear scaling and to remove very low-power envelopes. Furthermore, it was observed that in the case of very quiet or very high-pitched notes SMS on occasion returns corrupted envelopes, and so these were also ignored. This resulted in 85 envelopes for each instrument each of which were then analysed by eye to locate their ASR segments. It was observed that, in general, because the instruments analysed were wind instruments, the most suitable envelope description was ASR. As the NMF algorithm requires the data vector at each time instant to have an equal number of points, it was necessary to stretch/compress the envelopes. This was achieved by applying interpolation or decimation to the sustain portion of the envelopes only [4].

When first applying NMF to the envelopes it was decided to specify random non-negative initial conditions for both **W** and **X**. The value of *r* was set to 3 to reflect the representation of ASR. The basis functions for the Altosaxophone envelopes determined by the NMF algorithm are shown in Figure 2(a). To enhance the display they are normalised by the maximum value of all the bases. Although it is possible to see that each basis function outlines the shape of its respective ASR segment there is no obvious compaction of energy into three time-separated regions. This was improved by specifying templates with the regions over which the basis functions should be computed.

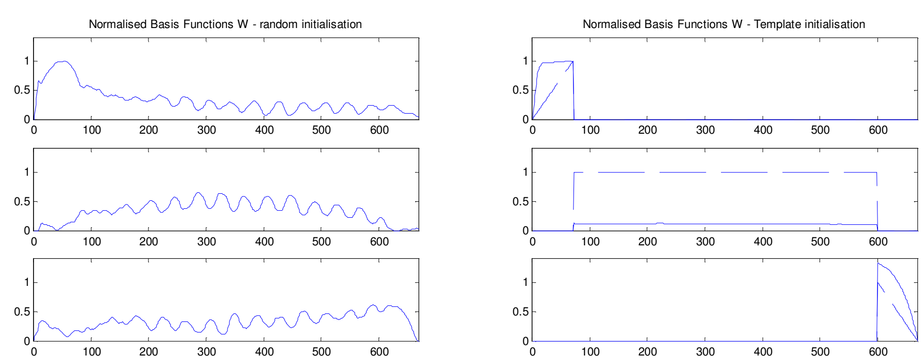

Figure2: *Basis functions found by NMF with (a) Random initialisation (b) Specified initialisation*

The resulting normalised bases for the Altosaxophone envelopes along with the templates used are shown in Figure 2(b). In the Figure, the solid line shows the bases and the dashed line the templates. The basis templates for each instrument were created based on ASR descriptions given in [5].

To examine the relationship between the bases for the three instruments and the characteristics of their various envelopes, the curve coefficient (*n*) and percentage duration of the segment length (*%length*) was computed for each envelope. The mean of these values after the removal of outliers for each set of instrument envelopes was then found. In a similar manner, the attack and release segments of the first and third basis computed by NMF for each instrument sound were identified and the curve coefficient and percentage segment length were calculated. Table 1 illustrates the results.

	Altosax (*n*, *%length*)	Clarinet (*n*, *%length*)	Flute (*n*, *%length*)
Actual – Attack	(1.82, 3.9)	(1.74, 4.34)	(1.73, 5.72)
Actual – Release	(0.8, 11.5)	(0.69, 6.89)	(0.95, 4.23)
Bases (**W**) – Attack	(1.84, 3.88)	(1.53, 5.31)	(1.89, 7.88)
Bases (**W**) – Release	(0.76, 8.36)	(0.65, 9.14)	(0.97, 7.44)

Table 1: *Values of curve coefficients and percent segment duration for attack and release of actual data and NMF basis functions.*

From Table 1 it can be seen that in all cases the curve coefficient of the NMF bases is a close match to the average of those extracted from the actual data. For the percentage segment duration, the match is not as good with a tendency for the percent segment duration of the basis function to be larger. This is most likely the result of the approximation imposed on the data by the algorithm.

V. Conclusions

This paper has investigated the application of NMF to partial envelope analysis. It has shown that to obtain useful basis functions, templates should be used. The curvature on the basis functions that model the attack and release segments is close to that of the mean for the partial envelope data. However, the basis functions do tend to overestimate the length of these segments. This NMF has application in sound synthesis algorithms such as audio morphing where data templates can be applied to create hybrid sounds from a multiplicity of different timbres.

References

[1] Lee, D., and Seung, S., "Learning the parts of objects by non-negative matrix factorization", *Nature*, Vol. 401, pp. 788-791, Oct. 1999.

[2] Sha, F., and Lawerence, K., "Real-Time pitch determination of one or more voices by non-negative matrix factorisation", in *Advances in Neural Information Processing Systems 17*, MIT Press, 2005.

[3] Jensen, K., "Timbre Models of Musical Sounds", *Ph.D dissertation*, Department of Computer Science, University of Copenhagen, 1999.

[4] McAdams, S., and Beauchamp, J.W., and Meneguzzi, S., "Discrimination of musical instrument sounds resynthesised with simplified spectrotemporal parameters", *JASA*, Vol. 105, No.2, Feb. 1999.

[5] Massey, H., et al., *A synthesist's guide to acoustic instruments*, Amsco publications, New York, USA, 1987.