AN EVALUATION OF WARPING TECHNIQUES APPLIED TO PARTIAL ENVELOPE ANALYSIS

Joseph Timoney, Tom Lysaght
Dept. of Computer Science,
NUI Maynooth, Maynooth,
Co. Kildare,
Ireland

Lorcan MacManus
Sch. of Elec. and Comms. Eng.,
DIT Kevin St.
Dublin,
Ireland

Victor Lazzarini
Department of Music,
NUI Maynooth, Maynooth,
Co. Kildare,
Ireland

ABSTRACT

Traditional methods for partial envelope analysis run into difficulties in the presence of tremolo and noise. This paper investigates an alternative method of analysis that uses a variety of non-linear warping techniques to match a synthetic envelope template to the partial envelope. Generation of suitable templates is discussed and the variety of implementations for the warping algorithm is examined. A complete procedure for matching the template is then explained and justified. Application of this procedure to partial envelope analysis of flute sounds is described and the results presented demonstrate the best choice of warping implementation. This technique allows for better envelope segmentation thus enabling improved modelling in representations such as the Timbre Model [1].

1. INTRODUCTION

One of the most important timbral attributes of a sound is the envelope [1]. The envelope is the evolution over time of the amplitude of a sound. Assuming an additive model for a sound, the envelope can be more exactly defined as the sum of the envelopes of the partials or additive components that make up the sound. Modelling of partial envelopes is of great interest to those constructing models of musical timbre [1] as an accurate and compact description of an envelope is a necessary prerequisite to further processing. Envelopes are generally categorised into three or four parts namely: the 'attack' portion, the 'decay' portion, the 'steady-state' or 'sustain' portion, and the 'release' portion [2], typically known the ADSR (Attack-Decay-Sustain-Release) description. Although simple this 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].

Automatic classification of the time evolution of partial envelopes into ADSR portions has not been a subject of intense study, however. Often the identification of these time segments is done by eye [4]. Of the algorithmic techniques available the most well known uses a piecewise linear approximation-based model [5]. Another two methods outlined in [1] are termed the percent method and the slope method. These are based on identifying a peak in the envelope or its smoothed derivative respectively and then working backwards to determine the start points of attack segments based on an amplitude threshold. A similar idea is used to find the release. All of these methods were examined in [6] for the analysis of the partial envelopes of Irish tin whistle sounds. It was found that they suffered from drawbacks particularly when the partial envelope was not well defined due to contamination with tremolo or noise. The piecewise-linear approximation blindly searches for four segments but unfortunately the segments it identifies do not necessarily correspond to the ADSR portions especially when the envelope exhibits tremolo. The percent and slope methods require many heuristics in their implementation to prevent incorrect segment identification, making it difficult to have complete confidence in their results when applied automatically.

To overcome the limitations of these techniques it was proposed to use a segment identification procedure based on template matching. Given that the ADSR description is a simple model and the partial envelopes are very variable, it was felt that a template ADSR curve, stretched or compressed to fit the partial envelope in some best sense should give a good approximation to the hypothesised underlying segments. As the segment boundaries of the template are known the warping path should automatically locate the corresponding segments on the partial envelope once the matching is performed. To stretch and compress the template so that it fits the partial envelope required a non-linear warping. This was achieved using the technique of Dynamic Time Warping (DTW), an algorithm that was the cornerstone of many isolated word recognition algorithms [7].

Initial experiments using a method that employed DTW for the analysis of the partial envelopes of tin whistle sounds suggested that it produced more reliable results [6]. In that work, however, only one template was used. While sufficient for such a preliminary investigation, the significant variability of partial envelopes suggests that one template is inadequate and it would be better to have a selection of templates of various shapes. Another consideration is that there are a number of possible non-linear mapping techniques, and it would be worthwhile to investigate which one gives the best results. This paper intends to address these issues and presents both a more studied and comprehensive approach to applying non-linear mapping techniques to partial envelope analysis along with experimental results that illustrate which techniques provides the best results.

2. DYNAMIC TIME WARPING – AN OVERVIEW

Classic Dynamic Time Warping, as described by [7] aligns a template data series $\mathbf{x} = \langle x_i \rangle$, $i = \{1,\ldots,M\}$ with a reference data series $\mathbf{y} = \langle y_k \rangle$, $k = \{1,\ldots,N\}$ by finding the lowest cost path through the $\langle N \times M \rangle$ field, **F**, defined as:

$$\mathbf{F}_{i,k} = \mathrm{d}(x_i, y_k) \tag{1}$$

where $\mathrm{d}(x_i, y_k)$ is some appropriate distance metric. Classic dynamic time warping yields the first eight variations on the DTW algorithm employed in this paper. They are henceforth referred to by number: 1, 2, 3 and 4 referring respectively to the step constraints of Figure 1(a-d) in symmetric form; 5,6,7 and 8 referring respectively to the step constraints of Figure 1(a-d) in asymmetric form [7].

More recently, new warping techniques have been proposed as an alternative to DTW, they are Correlation Optimized Shifting (COS) and Correlation Optimized Warping (COW) [8], [9]. COS is a technique to simply shift a vector 'left-right' to get a maximum correlation.

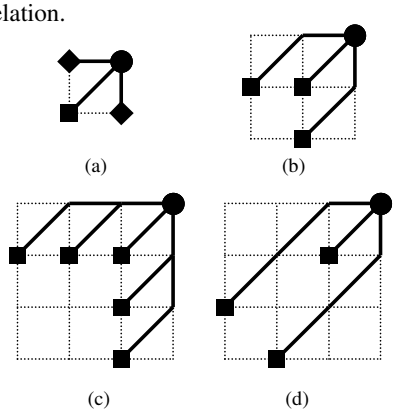

Figure 1. DTW step constraints [7]

In COW, the template data series $\mathbf{x} = \langle x_i \rangle$, of length *M* and the reference data series $\mathbf{y} = \langle y_k \rangle$ of length *N* are first globally aligned by stretching or compressing $\mathbf{y}$ to length *M* using linear interpolation to produce the vector $\mathbf{y}'$. Both vectors are then broken into *p* segments, each of length *q*. Each segment in $\mathbf{y}'$, is then further stretched or compressed by $\pm\Delta$ points ($\Delta << q$). For each point $\{q-\Delta,\ldots,q+\Delta',\ldots,q+\Delta\}$, the cross correlation is found with the $q+\Delta'$ points taken from the beginning of the corresponding segment of the reference template. A dynamic programming technique is used to maximise the sum of cross correlation coefficients across the entire set of *p* segments, thus achieving the desired warping.

COW and COS are henceforth referred to as techniques 9 and 10 respectively.

3. ALGORITHM

To model partial envelopes the following procedure was devised. First, a good synthetic template for each envelope was found. Next, variability in the partial envelope was reduced in such a manner that its coarse shape was not affected. Lastly, the synthetic template was warped so that it fitted the partial envelope.

3.1 Generating Synthetic Templates

The ADSR amplitude envelope of the patch diagrams for 24 well known instruments as given in [10] were modified to give 10 separate sustain levels resulting in 240 templates. As relative timings were used in the patch diagrams the length of the templates was varied to equal the length of the partial envelope under analysis. Each template envelope was then correlated with the envelope under analysis to determine the best match.

3.2 Reducing partial envelope variability

Visual inspection determined the partial envelopes to be highly variable. Although in most cases the underlying trend could be discerned by eye, the shape was distorted due to noisy fluctuations and low-frequency periodic fluctuations or tremolo. Filtering was found not to be ideal for removing these disturbances as a number of problems arose. If the filter was designed to have a sharp cutoff to remove the tremolo, too much smoothing of other portions of the envelope occurred which resulted in a smearing of events. Furthermore, a significant delay was introduced to the envelope. Broadening the filter passband appeared to be ineffective. Nonlinear median filtering was effective at attenuating the tremolo, but it also distorted the peak of attack significantly and so was found not to be useful. It was then discovered that an efficient and straightforward method of preserving the overall shape and position of significant transitions while reducing the variability was to quantize the partial envelope. Therefore, before warping of the best-match synthetic envelope to the partial envelope quantization of the partial envelope was carried out.

3.3 Synthetic Envelope Warping

Once the partial envelope was quantized, the best-match synthetic envelope was distorted to fit it using a warping procedure. Before warping is carried out, normalisation of the partial envelopes is required. Furthermore, trailing zeros or waveform values of very low amplitude value, on the order of 10^{-2}, at the start and end of the signal are removed to ensure they do not unduly influence the warping procedure. The warping path returned from the warping algorithm is used to adjust the synthetic template. Figure 2 plots an example of the warped template from the output of the DTW algorithm (dotted

line) superimposed on the original partial envelope (solid line) for the second harmonic of a flute note. It can be seen that the warped synthetic envelope provides a very good match to the shape of the partial envelope.

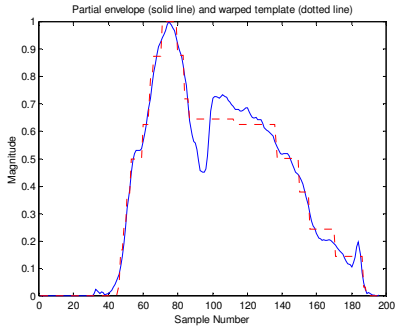

Figure 2. Warped Synthetic Template (dotted line) superimposed on the original partial envelope (solid line).

3.4 Iterative Quantization and Warping

To improve the overall matching it was reasoned that the quantization and warping of the partial envelope should be carried out in an iterative manner. The procedure commenced with a low-resolution coarse quantization of the partial envelope followed by a warping of the synthetic envelope. A second quantization at a higher resolution of the partial envelope was then performed, which was again followed by a warping of the synthetic envelope. This process can be repeated as often as is deemed necessary. The block diagram in Figure 3 illustrates the complete procedure.

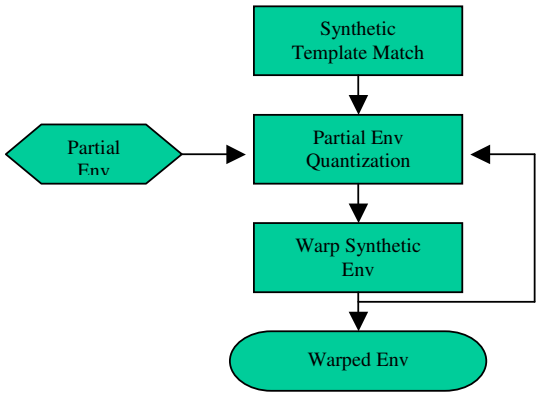

Figure 3. Flowchart of iterative warping procedure.

In the implementation described here, two stages were found to be sufficient. The first quantization was to 4 bits on the normalised partial envelope while the second quantisation was to 6 bits.

4. RESULTS

To test the performance of the various warping algorithms a set of partial envelopes from real sounds had to be obtained first. A complete set of samples for a flute were obtained from the University of Iowa Electronic Music Studios website [11], consisting of notes B3 to Db7 at three dynamic levels: *pp*, *mf*, *ff*. Prior to analysis of the notes the endpoints were detected using the algorithm of [12] to remove any leading or trailing zeros. Each note was then analysed using the well-known McAulay and Quatieri Sinewave analysis/synthesis technique [13] using an implementation provided by [14]. The algorithm produced a set of magnitude envelopes and frequency tracks for each note. Post-processing was performed on the algorithm output to retain only the most significant partial envelopes. The power of each envelope normalised by its maximum amplitude value was computed and only those whose power exceeded a threshold of 0.2 were retained. This was done to prevent very weak noisy envelopes corrupting or biasing the evaluation of the warping algorithms. It was found that the implementation used did not produce useable envelopes for very high-pitched quiet sounds. Approximately 5 useful envelopes for each of the 112 notes extracted were stored.

Each partial envelope was passed through the template matching procedure to find the most correlated synthetic envelope. This partial envelope was then input to the two-stage quantization and warping algorithm which output a warped synthetic envelope matched to the shape of the original partial envelope. This was carried out for all ten warping procedures. To determine which warped synthetic envelope best matched the original partial envelope the mean square error between them was computed. This is summarised by equation (4)

$$MSE_{n,i,w} = \sum_{t=0}^{L} \left(env_{n,i}(t) - synth\_env_{n,i,w}(t) \right)^2 / L \quad (4)$$

where n is a integer index variable representing the number of the note being analysed, i is the index of partial envelope being analysed, $env(t)$ is the partial envelope, $synth\_env(t)$ describes the warped synthetic template resulting from using the warping procedure indexed by w, and L is a variable representing the length of each partial $env_{n,i}(t)$.

For all the warping procedures the width of the adjustment window size was large relative to the signal length. In the cases of the correlation optimised warping and the correlation optimised shift the segment length was selected to be approximately 5% of the total length. In a few cases this segment length was too long and these procedures failed. To overcome this an iterative procedure was included in the code to keep reducing the segment length one sample at a time until the procedure worked.

In Figure 4, the solid line with diamond markers plots the average of the mean square error for the different warping procedures over all the partials for each note. It can be seen that the three best warping

procedures in the order of lowest average MSE first are 1, 5 and 6 namely: the DTW with symmetric constraints 1, DTW with asymmetric constraints 1 and DTW with asymmetric constraints 2.

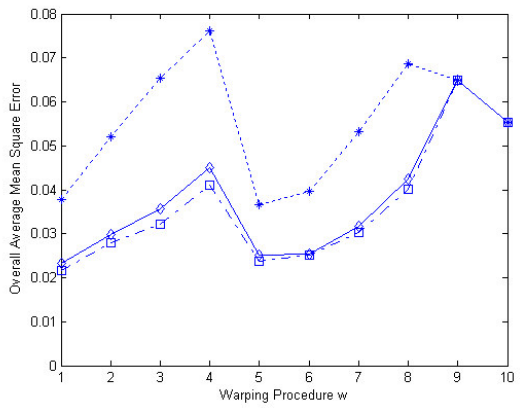

Figure 4. Average mean square error for each warping procedure for the first partials only.

Both the correlation optimised warping and the correlation optimised shifting algorithms gave the worst performance. It appears then that from an overall perspective the DTW algorithm with the simplest constraints is best. To investigate whether different warping procedures were favoured for particular partial indices the dash-dot line with square markers gives the average MSE of the different warping procedures over all the notes for the first partial only, while the dotted line with asterisk markers presents the same information but for the fourth partial only. The first partial appears to be almost identical to the average. Again, the best three warping procedures in order are the DTW with symmetric constraints 1, DTW with asymmetric constraints 1 and DTW with asymmetric constraints 2. The fourth partial shows different behaviour. According to the plot the order of the best three warping procedures are the DTW with asymmetric constraints 1, DTW with symmetric constraints 1 and DTW with asymmetric constraints 2. Another anomaly is that the performance of the correlation optimised shift and the correlation optimised warping procedures is relatively better than before.

5. CONCLUSIONS

This paper has described a comprehensive method for the application of template matching and non-linear warping techniques to the analysis of partial envelopes. The generation of suitable synthetic templates was investigated. A number of implementations of the Dynamic Time Warping algorithm along with Correlation Optimised Shifting and Correlation Optimised Warping were applied to the matching of the synthetic templates to real partial envelopes. The results suggested that the best implementation of DTW was technique 1 with symmetric constraints. This procedure could be readily applied in any application that requires timbral analysis such as the Timbre model decomposition proposed in [1], [15].

Future work will extend these experiments to partial envelopes extracted from other instrumental sounds. Another development under consideration is the generation of more realistic templates using hand-segmented prototypical envelopes extracted from real partial envelope data.

6. REFERENCES

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

[2] Helen, M., and Virtanen, T., "Perceptually motivated parametric representation for harmonic sounds for data compression purposes", *DAFX03*, Queen Mary University of London, London, UK, Sept. 8-11, 2003.

[3] Grey, J. and Gordon, J., "Perceptual effects of Spectral Modifications on Musical Timbres", *JASA*, vol 63, no. 5, 1978.

[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] Bernstein, A., and Cooper, E., "The piecewise linear technique of electronic music synthesis". *Jnl. of the AES*, vol 24, no. 6 1976.

[6] Timoney, J., Mac Manus, L.,, Lysaght, T., and Schwarzbacher, A., "Dynamic Time Warping for Tin Whistle Partial Envelope", *Irish Signals and Systems conference 2004*, Belfast, Northern Ireland, July 2004.

[7] Sakoe, H., and Chiba, S., 'Dynamic programming algorithm optimization for spoken word recognition', *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 26, no. 1, Feb. 1978.

[8] Tomasi, G., *DTW (Dynamic Time Warping) and COW (Correlation Optimized Warping) demo*, Quality and Technology, Dept. of Food science, KVL, Denmark, available at http://www.models.kvl.dk/source/DTW_COW/index.asp

[9] Tomasi, G., and Van Der Berg, F., and Andersson, C., 'Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data', *Journal of Chemometrics*, vol. 18, no. 5, Jul. 2004.

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

[11] University of Iowa Musical Instrument Samples available at http://theremin.music.uiowa.edu/MIS.flute.html

[12] Rabiner, L., and Sambur, M., 'An algorithm for determining the endpoints of isolated utterances', *Bell Syst. Tech Jnl.*, vol. 54, no. 2, Feb. 1975.

[13] McAulay, R., and Quatieri, T., 'Speech analysis/synthesis based on a sinusoidal representation', *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 34, no. 4, Aug. 1986.

[14] Ellis, D., *Sinewave and Sinusoid+Noise Analysis/Synthesis in Matlab*, available at http:/labrosa.ee.columbia.edu/matlab/sinemodel

[15] Timoney, J. et al. , "Timbral attributes for objective quality assessment of the Irish tin whistle", *DAFX 2004*, Naples, Italy, Sept. 2004.