Dynamic Time Warping for Tin Whistle Partial Envelope Analysis

Joseph Timoney$^{\phi}$, Thomas Lysaght, Lorcán Mac Manus* and Andreas Schwarzbacher

$^{\phi}$*Department of Computer Science,*
NUI Maynooth, Maynooth,
Co. Kildare, IRELAND.
E-mail: $^{\phi}$ jtimoney@cs.may.ie

* School of Electrical and Electronic Engineering.
Dublin Institute of Technology, Kevin St,
Dublin 8, IRELAND
*E-mail: *lorcan.macmanus@dit.ie*

_____

Abstract – **This paper examines techniques for the analysis of the amplitude envelope of harmonic partials in a tin whistle sound. Due to the nature of the envelope, a robust technique is required to estimate envelope parameters. This is achieved using a dynamic time warping based method. Furthermore, a heuristic scheme for addressing difficulties in envelope modelling is also introduced. The results presented show performance improvements over more established methods.**

Keywords – **Timbre model, Partial Envelope Analysis, Dynamic Time Warping.**

_____

I   INTRODUCTION

The penny whistle, the tin whistle or the Irish tin whistle is an instrument most often associated with Irish folk music. While the oldest surviving specimens are the 12th-century bone whistles recently unearthed in the old Norman quarter of Dublin, various types of whistle flutes that were the progenitors of the modern tin whistle are frequently mentioned in legends of ancient Irish society [5].

The tin whistle breaks many of the norms of musical instruments. In particular, there is no consensus among musicians on the merits of more finely crafted, expensive instruments. Indeed many of the great tin whistle players actually prefer to play inexpensive whistles, believing the sound to be more in keeping with the traditions of folk music than the 'sweeter' tones of the more expensive varieties [7].

In its basic form, the whistle consists of a tube (usually metal), which is either cylindrical or conical in shape. A whistle is attached to one end. Different notes can be sounded by blowing through the whistle, and using different finger pattern on the six holes on the shaft of the instrument. The most common whistle type is in the key of D, which is also the key of most Irish folk music. Air is passed through the whistle opening and strikes the lip of the whistle as it exits the fipple-hole (See Figure 1) [8].
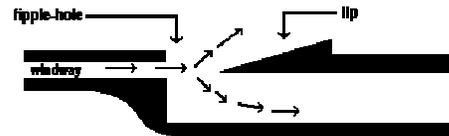
Figure 1: Air jet striking the fipple hole of a tin whistle.

The air jet then rapidly switches between one of two paths: through the bore of the instrument and passing out to the atmosphere, thus setting up a series of acoustic oscillations in the shaft. The effective length of the shaft can be altered by appropriate fingering as shown:
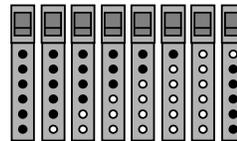
Figure 2: This shows the fingering positions for the tin whistle, playing Doh in the lower register on the left hand side, up to Doh in the upper register, on the right hand side. The black dots indicate closed finger holes, while the white dots indicate open holes.

Altering the effective length of the instrument in this way changes the resonance frequency. This ensures that only the relevant fundamental tone and partials are amplified, to produce the required note. As can be seen, the whistle is a naturally diatonic

instrument. Chromatic notes may also be played, however, such details will not concern us here.

1.1 *Characterising Musical Instruments*

Various approaches towards the characterisation of musical instruments are possible. Acoustic impedance measurements are applied by [9] to classical and modern flutes. In this approach, an experiment is carried out that characterises the ratio between the air pressure and (the complex) airflow rate, to give a complex impedance measure, which is a function of frequency. This acoustic impedance is a characteristic of the instrument that is independent of the musician. In this paper, a signal processing approach is taken using the timbre model [1] on tin whistle sounds. The sounds were recorded in uncompressed format in a quiet room using a standard microphone with a mini-disc recording system at 44.1kHz (CD quality), played by a professional Irish traditional musician.

II ADSR

The envelopes of instrumental tones are generally categorised into three parts namely: the 'attack' portion, the 'steady-state' portion, and the 'decay' portion as a general guideline as to the structure of sounds. Furthermore, research on timbre perception has identified the existence of perceptual attributes associated with these categorizations, such as the temporal pattern of the attack and decay [4]. Often the identification of these time segments is done by eye [11]. The ADSR (attack-decay-sustain-release) envelope is widely used as a way of modelling these envelope segments [2].
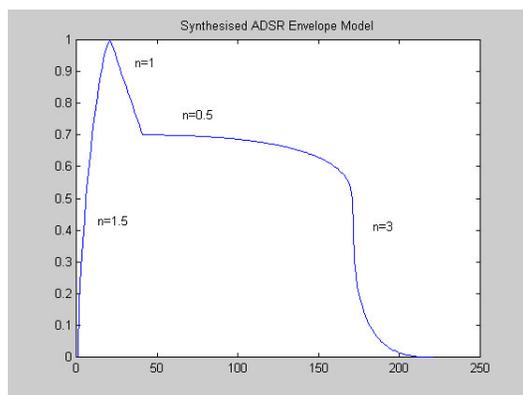
Figure 3: Synthesised version of the ideal ADSR curve.

It has been shown by [1] that each segment may be modelled using equation (4), which can be altered using a single model parameter, *n*. Figure 3 shows four versions of this curve being used to create a template of the four segments in the ADSR model,

using the indicated values of *n* [2]. Automatically identifying each of the four individual segments is unfortunately not a trivial matter, as the real envelope data does not always correspond very well with the ideal ADSR model. The task of identifying the segments is also the task of identifying the start and end points of each segment. This is termed Split Point Identification, and is the main subject of this paper.

III SPLIT POINT IDENTIFICATION

This section is composed of two parts. In Section 3.1 the existing methods for split point identification are discussed. Section 3.2 introduces the dynamic time warping method, which is the main contribution of this paper.

3.1 *Percent based method and piecewise linear method*

An envelope model presented in [1] identifies 6 split points:

SOA = Start Of Attack
MOA = Middle Of Attack
EOA = End Of Attack
SOR = Start Of Release
MOR = Middle Of Release
EOR = End Of Release

A method called the percent method is used to find these split points. This consists of finding the maximum of the curve to be modelled, and then finding the first and last time in the curve where the value is above a constant percent of the maximum. The SOA is then defined as the first time the partial amplitude is above 10%, and the EOR is defined as the last time the amplitude is above 10%, the EOA as the first time the amplitude is above 90% and the SOR as the last time the amplitude is above 70%. With this method it is not possible to find good estimates of split points for all partials. A more accurate method is to identify maxima and minima from the derivative of a smoothed version of the envelope amplitude.

Traditional piecewise linear approximation assumes linearity between segments but this assumption has no real psychoacoustic basis. The piecewise linear segmentation algorithm depends on a user-defined input of a desired number of segments. It works by creating a fine-segmented representation and then merging the lowest cost segments, in terms of the mean squared error, until only the desired number remain [3, 12].

3.2 *Dynamic Time Warping*

A better approach to more accurately identify the split points would be to take the ADSR model envelope and try to fit it to the envelope split points. A dynamic time warping approach is used to fit the synthetic envelope to the actual envelope of each partial.

Dynamic time warping is a method of aligning corresponding points in two time series of points. It is typically implemented using dynamic programming and results in matching each point in one time series to corresponding points in the other. The standard Euclidean distance metric is used as a similarity measure between points in each series. The dynamic time warping distance between two time series $\vec{x}, \vec{y}$ is given by:

$$D_{DTW}^2\left(\vec{x}, \vec{y}\right) = D^2\left(First\left(\vec{x}\right), First\left(\vec{y}\right)\right)$$
$$+ \min \begin{cases} D_{DTW}^2\left(\vec{x}, \text{Rest}\left(\vec{y}\right)\right) \\ D_{DTW}^2\left(\text{Rest}\left(\vec{x}\right), \vec{y}\right) \\ D_{DTW}^2\left(\text{Rest}\left(\vec{x}\right), \text{Rest}\left(\vec{y}\right)\right) \end{cases} \quad (1)$$

An $n \times n$ matrix is constructed to align the time series $\vec{x}^{n}$ and $\vec{y}^{m}$ with cell (i, j) representing the alignment of element $\vec{x}^{n}$ with element $\vec{y}^{m}$. A warping path, P, from cell $(1,1)$ to (n, m) corresponds to a particular alignment, element by element, between $\vec{x}^{n}$ and $\vec{y}^{m}$:

$$P = p_1, p_2, \ldots, p_L = \left(p_1^x, p_1^y\right)\left(p_2^x, p_2^y\right), \ldots, \left(p_L^x, p_{L1}^y\right) \quad (2)$$
$$\max(m,n) \le L \le n + m - 1$$

where $p_t^x, p_t^y, t = 1, 2, \ldots, L$ are the position numbers of $\vec{x}^{n}$ and $\vec{y}^{m}$ respectively in the alignment. The distance between $\vec{x}^{n}$ and $\vec{y}^{m}$ on the warping path P is the distance between $x_{p_t^x}$ and $x_{p_t^y}, t = 1, 2, \ldots, L$. The constraints on the path P are:

$$P \text{ must be monotonic:}$$
$$p_t^x - p_{t-1}^x \ge 0 \text{ and } p_t^y - p_{t-1}^y \ge 0$$
$$P \text{ must be continuous:}$$
$$p_t^x - p_{t-1}^x \le 1 \text{ and } p_t^y - p_{t-1}^y \le 1 \quad (3)$$

The number of possible warping paths grows exponentially with the length of the time series. The distance that is minimized over all paths is the Dynamic Time Warping distance. With dynamic programming the computational time is of the order $O(mn)$. Empirical application of dynamic time warping of the synthetic ADSR envelope to the partial envelopes of the tin whistle did not always provide a successful result. It was found that to obtain an improved performance, quantization of the partial envelope was required. The effect of quantization is to reduce variability of the data values in the envelope and therefore force a reduction in the suitable candidate matching points. Furthermore, a two-stage quantization was found to be more effective, as the first stage mostly gave a good matching for the attack while the second stage helped to improve the warping of the release portion of the synthetic envelope. Thus, fitting the synthetic ADSR envelope to the actual envelope of each partial is done in a sequence of steps as follows:

1. A Hanning smoothing window is used to generate a filtered envelope of the original partial. The impulse response of the window is 250ms long.
2. Following normalisation, the filtered envelope is quantized to a resolution of 6-bits in order to obtain the coarsest features only, in a manner unobtainable by simple linear smoothing.
3. The dynamic time warping algorithm is used to warp the synthetic envelope to this signal.
4. The normalised unfiltered envelope is then quantized to a resolution of 4-bits.
5. The dynamic time warping algorithm is used again to warp the signal resulting from Step 3 to the signal resulting from Step 4.
6. The split points on the original ADSR model will then have been translated to the best estimate for these split points on the amplitude envelope of the partial.

IV   CURVE FITTING

The segments between split points are modelled by a curve either in exponential or logarithmic form to create a synthetic ADSR envelope. Each segment is modelled using the following curve model

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

where x represents the time evolution of the curve normalised between 0 and 1, v_0 is the starting value of the curve, v_1 is the final value, and n is the curve

coefficient that must be estimated. This is done by first using a least squares regression on the logarithm of the normalised envelope:

$$\log[\widehat{env}_{norm}(x)] = \log\left[\left(1-(1-x)^n\right)^{1/n}\right] \quad (5)$$

$$\text{where } \widehat{env}_{norm}(x) = \frac{\widehat{env}(x)-v_0}{(v_1-v_0)}$$

$$= \frac{1}{n}\log\left[\left(1-(1-x)^n\right)\right]$$

Using the Binomial expansion on $(1-x)^n$, gives:

$$= \frac{1}{n}\log\left[\left(1-\left(1-nx+\frac{n(n-1)}{2!}x^2-\frac{n(n-1)(n-2)}{3!}x^3+\ldots\right)\right)\right] \quad (6)$$

$$= \frac{1}{n}\log\left[\left(nx\left(1-\frac{(n-1)}{2!}x+\frac{(n-2)(n-2)}{3!}x^2-\ldots\right)\right)\right]$$

$$= \frac{\log(nx)}{n}+\frac{1}{n}\log\left[\left(1-\frac{(n-1)}{2!}x+\frac{(n-2)(n-2)}{3!}x^2-\ldots\right)\right]$$

$$= \frac{\log(n)}{n}+\frac{\log(x)}{n}+\frac{1}{n}\log\left[\left(1-\frac{(n-1)}{2!}x+\frac{(n-2)(n-2)}{3!}x^2-\ldots\right)\right]$$

(7)

Assuming that the third term in the above is small, it is possible to use a regression procedure to form an estimate for *n* using either of the first two terms. This estimation is refined using the Levenberg-Marquardt optimisation routine [10]. However, it was found to be necessary to add some heuristics when dealing with the envelopes of the tin whistle sounds because of some unpredictable deviations from the ideal ADSR curve. Without these heuristics, the optimisation routine can fail. They consisted of:

1. If the envelope has a sufficiently high crest factor a non-linear filtering routine is applied to reduce this figure.

2. If any values of the envelope curve are less than zero because of the filtering routine, they are replaced with a small positive constant to enable the logarithm to be taken of the entire data set.

The regression produces two estimates for the curve coefficient. By default, the second estimate is chosen. However, if this estimate is found to be of value zero or to have the incorrect sign, the first estimate is chosen. If both estimates have the incorrect sign, then the regression is repeated on a median filtered version of the envelope.

V   RESULTS

Figure 4 illustrates the results of applying the three split point estimation techniques to the envelope of the third partial of the note G played on a Generation-D tin whistle.
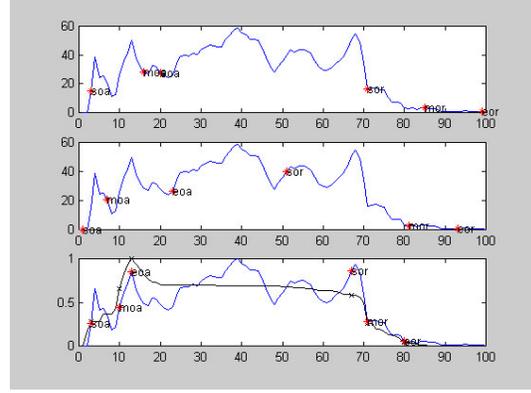
Figure 4: Third partial of Generation-D G5

From the evolution of the partial it can be seen to conform in some sense to the ADSR model but the attack is not well defined, and over the decay and sustain portions of the envelope there is some variation in envelope amplitude. In the first panel, the split points as estimated by the algorithm contained in [1] are shown marked by asterisks. It can be seen that the estimation of the correct location of these points is not good in all cases. The second panel gives these points estimated using the piecewise linear segmentation technique. While they can be said to be reasonably true for the attack, for the release the SOR appears too early in the envelope while the MOR and the EOR appear too late. In the third panel, a comparison between the normalised envelope and the fitted envelope using dynamic time warping is made. Visually, a good match appears between the two. Furthermore, the estimates of the points is excellent.
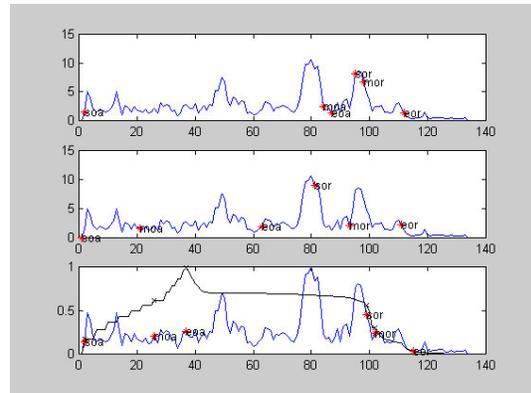
Figure 5: Fifth partial Generation-D E5

Figure 5 illustrate the results from applying the techniques to the envelope of the fifth partial of the note E played on a Generation-D tin whistle. The envelope of this partial is weak and cannot be said to resemble the ADSR envelope model. Furthermore, because of their close proximity, the split-points estimate using the technique of [1] is not good for the MOA, EOA and the MOR. Using the piecewise linear segmentation, the split points estimate is good for the SOR and the EOR. In the third panel, the synthetic envelope after warping does highlight the release of the sound, however, it is not so certain in establishing where the attack lies. This case is possibly very difficult for all three techniques.

Figure 6 shows the envelope of the first partial of the note G played on the Feadóg tin whistle (solid line), the resynthesised envelope using the curve modelling procedure, with the split points estimated using the algorithm of [1] (dashed line) and using the piecewise linear approximation (x's).

piecewise linear approximation and the o's to the dynamic time warping based technique.
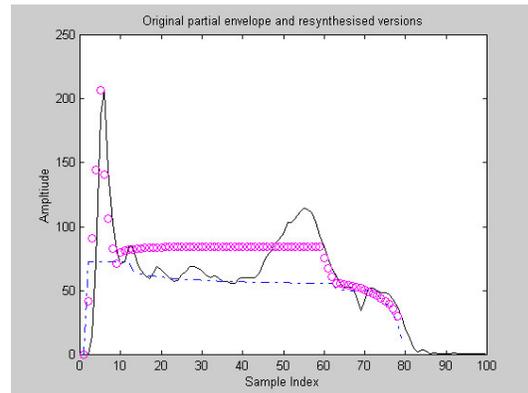
Figure 7: Reproduced envelope using the split points estimated by dynamic time warping.

Figure 6: First partial of note G5
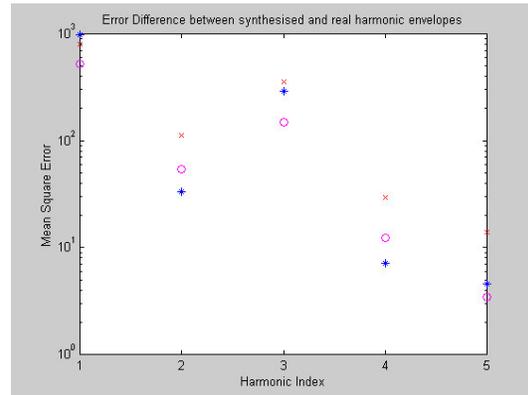
Figure 8: Comparison of the three algorithms

It can be seen that the envelope using the split points estimated from [1] can capture the shape from sustain to release but misses the significant peak of attack. This is also true of the split points estimated using the piecewise linear approximation and furthermore, this method reproduces an incorrect release portion of the original envelope.

Figure 7 shows the reproduced envelope using the split points estimated by dynamic time warping (o's). As can be seen, it correctly captures the attack portion of the sound envelope.

Figure 8 shows the logarithm of the mean square error difference between the envelopes of five harmonic partials of the note G and their resynthesised versions, using the curve model determined over segments whose split points were found by the various techniques. The asterisks corresponds to the algorithm of [1], the x's to the

The error of the resynthesised envelopes is lowest for the method of [1] in two cases and the dynamic time warping-based method provides the best estimate in the other three cases.

VI   CONCLUSIONS

This paper demonstrates that the method of dynamic time warping provides a superior way to identify the split points in the ADSR envelope of the partials than more established techniques. Examples for which the technique did not work well generally involve envelopes that exhibit significant deviations from the ADSR model. This difficulty is evident from Figure 7 and all techniques discussed fail to satisfactorily resolve this difficulty. However, the comparison of the three techniques in Figure 8 shows that split points provided by the dynamic time warping method is an improvement over previous techniques.

Future work will focus on further improving the dynamic time warping split point identification technique on the tin whistle and on other families of instruments. We will also examine a more thorough application of the timbre model to the tin whistle.

REFERENCES

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

[2] M. Helen and T. Virtanen. "Perceptually motivated parametric representation for harmonic sounds for data compression purposes.". *6TH International Conference on Digital Audio Effects (DAFx03)*, pp. 177-179, queen Mary University of London, London, England, September 8-11, 2003.

[3] A. Bernstein and E. Cooper, "The piecewise linear technique of electronic music synthesis". *Journal of the AES*, Vol 24 No. 6 1976.

[4] J. Gray and J. Gordon, "Perceptual effects of Spectral Modifications on Musical Timbres" *Journal of ASA*, Vol 63 No. 5 1978.

[5] L.E. McCullough, "The Complete Irish Tin Whistle Tutor", *ISBN 0825603404*, March 2002.

[6] Yunyue Zhu and Dennis Shasha, "Warping Indexes with Envelope Transforms for Query by Humming", *ACM SIGMOD/PODS* 2003 San Diego, California, June 9-12, 2003.

[7] www.chiffandfipple.com/tutorial.html

[8] www.geocities.com/SoHo/Museum/4915/ 02.htm

[9] Wolfe, J., Smith, J., Tann, J. and Fletcher, N.H. (2001), "Acoustic Impedance of Classical and Modern Flutes", *Journal of Sound and Vibration*, 243, 127-144.

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

[11] S. McAdams, J. W. Beauchamp, and S. Meneguzzi "Discrimination of musical instrument sounds resynthesised with simplified spectrotemporal parameters" *Journal of Acoustic Society of America*, 105(2), Pt.1, February 1999.

[12] www.cs.ucr.edu/~eamonn/TSDMA/ time_series_toolbox/