TIMBRE MORPHING USING THE MODAL DISTRIBUTION

Thomas Lysaght and Joseph Timoney

Department of Computer Science
National University of Ireland, Maynooth
Tom.Lysaght@may.ie, Joseph.Timoney@may.ie

ABSTRACT

We present techniques for timbre morphing between two audio signals based on the Modal distribution time-frequency representation of music signals. A signal synthesis method is described which resynthesises signals from Modal distributions. Direct resynthesis from the original signal produces a timbre that is almost indistinguishable from the source. In deciding which salient features to morph a relational graph representation of timbre is used and linear interpolation and non-linear warping are applied in performing the morph between Modal distributions.

1. THE MODAL DISTRIBUTION

The Modal Distribution [1] in Eqn. (1) was developed as a tool for music signal analysis and is based on Cohen's class of bilinear time-frequency distributions. These distributions, which include the Wigner distribution (WD), have superior resolution to the traditional spectrogram. The Modal distribution is related, more specifically, to the smoothed pseudo-Wigner distribution (SPWD) and was designed for the analysis of signals based on the sum of sinusoids model.

$$M(t,\omega) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} W_S(\tau,\xi) h_{LP}(t-\tau)\xi \times G_{LP}(\omega-\xi) d\tau d\xi \qquad (1)$$

where:

$h_{LP}(\tau)$ is the time smoothing window and $G_{LP}(\xi)$ is the frequency smoothing window. In particular, the two Modal distribution kernel filter functions consist of auto-correlated time and frequency smoothing windows giving a 42dB side lobe attenuation. For music signals, the time smoothing window is usually chosen with a cutoff frequency a little less than the fundamental frequency. An important consideration with the Wigner and Modal distributions is that the Nyquist sampling frequency is four times the value of the highest frequency component in the signal. This necessitates sampling at twice the Nyquist rate or using the analytic version of the signal. The Modal distribution is computed by taking an FFT of each time slice of the smoothed temporal correlation function. As well as achieving cross term suppression, an added advantage of the time smoothing is that an FFT need only be taken at time intervals equal to the smoothing window length meaning that the Modal distribution is practically useful from a computational viewpoint. For these reasons, the Modal distribution is ideally

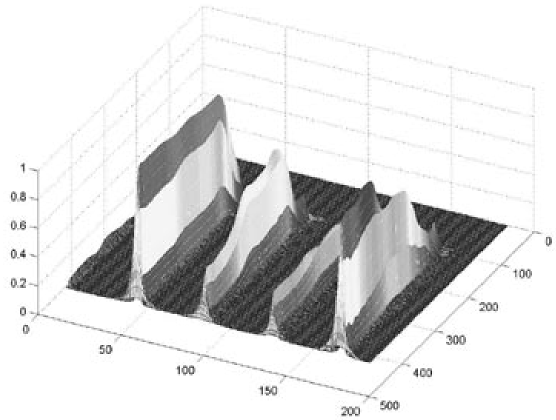

Figure 1 : *Modal distribution of first 4 harmonics trumpet sound*

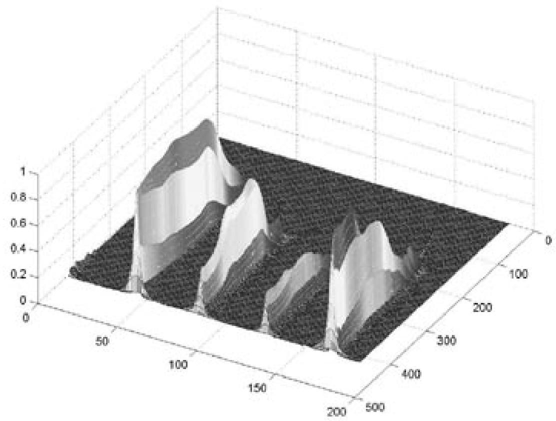

Figure 2 : *Modal distribution of warped harmonics in Figure 1.*

suited to music signal processing, including morphing, where accurate estimates of features such as spectral peaks, onset times and vibrato cycles are required. Modal distribution analysis has been applied successfully to event detection in instrumental musical signals [4].

1.1. Sounds

The sampled sounds for morphing are taken from the MUMS CDs consisting of 16-bit samples at a rate of 44.1kH. Initial experiments are run on 500ms trumpet (A4, MUMS Vol. 2, Track 16, Index 16) and stopped violin (A4, MUMS Vol. 1, Track 1, Index 16) notes of equal pitch. A 2048-point frequency-smoothing window results in a bin size of 10.766Hz. The time smoothing window cutoff frequency is about $2\pi \times 420$ Hz. Up to 24 harmonics in each sound are used in the morphing process.

2. CORRESPONDENCE

In any matching problem a correspondence needs to be established between features. The main features for timbre morphing include the peak of attack, loudest point, vibrato cycles, maximum extent of harmonics, and start of decay. In the examples given only the initial attack portion of the signal is used, consisting mainly of the peak of attack (POA) feature. Frequency tracks are extracted from Modal distributions by applying the peak-tracking method of McAuley / Quatieri [2]. Features are identified by locating extrema along each track. The POA feature along each harmonic is then found from maxima along each track The time position of the POA feature will be different for different harmonics as can be seen in Figures 1 & 2. This POA feature is used for temporal alignment of corresponding harmonics described in section 4.

2.1. Subgraph Isomorphism

Graph theoretic techniques are employed in establishing correspondences between Modal distributions. Sets of features are represented as nodes in relational graphs and a subgraph isomorphism algorithm [3] determines which subset of nodes in each graph correspond. The subgraph isomorphism algorithm describes a brute-force depth-first tree search algorithm with backtracking, designed to find all subgraphs between a given graph and subgraphs of a further graph. Subgraph isomorphism belongs to the class of NP-comlete problems and the brute force approach is O(n!) for a graph with n nodes. A refinement procedure reduces the number of searches. The algorithm is based on a connectivity analysis of each graph represented as adjacency matrices. Connected nodes indicate sequences of tracks for correspondence. A match matrix specifies the isomorphism found. The adjacency matrices can hold binary or integer values to represent the type of connectivity between nodes.

For timbre morphing, graph nodes represent features such as the magnitude of POA, harmonic number and maximum extent of harmonics. The structure of the graphs show the connectivity between features, thereby representing, for example, the spectral shape in the Modal distribution of each sound. The connectivity of each graph determines which sequence of tracks from each MD are combined for morphing. Using directed graphs will ensure only ascending or descending sequences of tracks for correspondence. Example matrices for the trumpet (G_{trp}) harmonics and violin (G_{vln}) harmonics are given in Eqn. 2. Integer values represent the type (magnitude) of connectivity between the POA features. For example, a value of 1 at position *G(row,col)* indicates that the POA features are within the same magnitude range. The match matrix *M*, represents just one of the isomorphisms found and specifies that harmonics 1,2,3,4,5, & 6 of the violin correspond to harmonics 1,2,3,4,6 & 9 of the trumpet.

$$G_{trp} = \begin{pmatrix} 011112222 \\ 001112222 \\ 000112222 \\ 000012222 \\ 000002222 \\ 000000122 \\ 000000012 \\ 000000000 \end{pmatrix} G_{v\ln} = \begin{pmatrix} 011122 \\ 001122 \\ 000122 \\ 000022 \\ 000002 \\ 000000 \end{pmatrix} \tag{2}$$

$$M = \begin{pmatrix} 100000000 \\ 010000000 \\ 001000000 \\ 000100000 \\ 000001000 \\ 000000001 \end{pmatrix}$$

3. INTERPOLATION AND WARPING

Non-linear warping is used to align key features such as the peak of attack. Peaks along corresponding partials are temporally aligned to yield one peak in the morphed timbre. A low order polynomial is sufficient to effect the warp but where multiple features are to be warped then a higher order polynomial is used. Linear interpolation is used to effect the morph between warped surfaces. This includes methods such as mean value and local average interpolation (Eqn 2):

$$S(f,t) = \frac{1}{2}\left\{N_{tr}(f,t) + N_{vl}(f,t)\right\} \quad (3)$$

where $N_{tr}(f,t)$ and $N_{vl}(f,t)$ are the local average of values on e.g., the warped trumpet and violin MD surfaces respectively. A gradient-weighted interpolation function is also defined which examines the effect of gradient on morphing. A morph step of 0.1 is used to determine the relative contribution of each instrument to the morph; a step of 0.1 indicates a 10% contribution from the first sound and a 90% contribution from the second sound. Furthermore, the morph window size also effects the quality of the morph. This has a greater effect on the local average interpolation where the bandwidth of the morphed harmonics are influenced by window size.

4. SYNTHESIS FROM MODAL DISTRIBUTIONS

We present a method of signal synthesis from Modal distributions based on estimates of instantaneous amplitude (Eqns. 3 & 4) and frequency (Eqn. 5) calculated along each track similar to that described in [4].

$$A(n) = 2\sqrt{P(n)} \quad (4)$$

where

$$P(n) = \sum_{l=L1}^{L2} M(n,l) \quad (5)$$

$$F(n) = \sum_{l=L1}^{L2} \frac{lM(n,l)}{P(n)} \quad (6)$$

Linear interpolation is used to compute frequency and amplitude samples between frame hop periods. As no phase information is available in the MD, the integral of the instantaneous frequency is used to recover the phase. Extensive experimentation shows the recovered signal to be almost indistinguishable from the original. The timbre quality can be improved by increasing the number of tracks for synthesis.

4.1. Results

Initial results for local average morphs between trumpet and violin sounds indicate a distinctive 'in-between' quality of timbre emerging as the morph step changes from 0.1 to 0.9. Because of the greater prominence of the trumpet harmonics, the change from violin timbre to trumpet timbre occurs when the violin contributes significantly to the morph (a morph step of 0.1 or 0.2). Correspondence here was based only on harmonic number. The morphed sound's Modal distribution is shown in Figure 3 with the original violin MD shown in Figure 4.

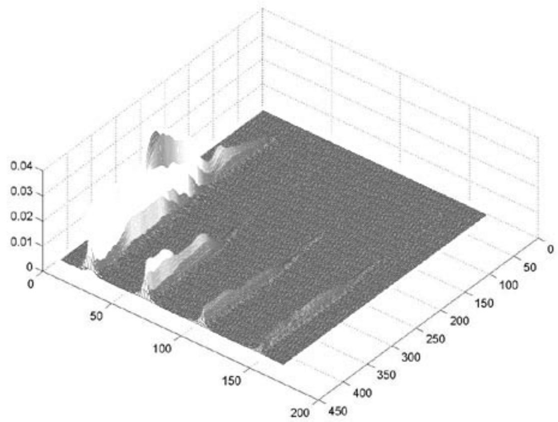

Figure 3 : *MD showing harmonics 5-8 of morphed sound*

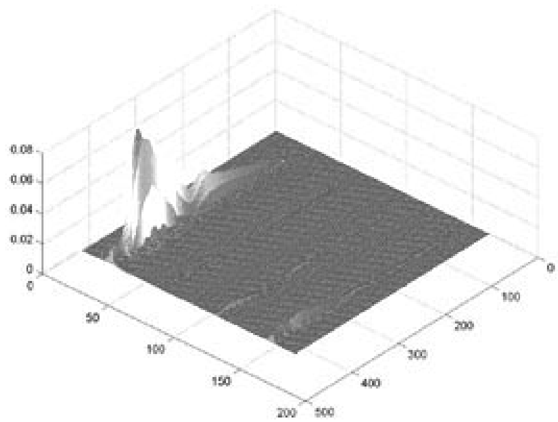

Figure 4 : *MD showing harmonics 5-8 of violin sound*

5. CONCLUSIONS

A model of timbre morphing has been presented based on the Modal time-frequency distribution which gives accurate estimates of key timbral features. Non-linear warping has been implemented as a means of successfully aligning temporal features in Modal distributions. A graph theoretic approach to the correspondence problem has been applied using subgraph isomorphism. Results prove successful in creating an in-between sound quality for the morphed trumpet and violin timbres. Future research will focus on morphing longer sounds including tones with vibrato cycles and identification of correspondences between multiple features. Furthermore, future work will include a perceptual analysis of the newly created timbres based on previous psychoacoustics studies of timbre.

6. REFERENCES

[1] W. J. Pielemeier and G. Wakefield, "A high-resolution time-frequency representation for musical instrument signals". *Journal ASA*, 99(4), Pt. 1, April 1986.

[2] Robert J. McCaulay and Thomas F. Quatieri, "Speech Analysis/Synthesis Based on a Sinusoidal Representation", *IEEE Trans. On Acoustics, Speech, and Signal Processing*, Vol. ASSP-34, No. 4, August 1986.

[3] J. R. Ullmann, "An Algorithm for Subgraph Isomorphism", *Jorunal of the Association of Computing Machinery*, 1035, 373-382, 1976.

[4] M. Mellody and G. Wakefield, "A Modal Distribution Study of Violin Vibrato", *Proceedings SPIE, Advanced Signal Processing Algorithms, Architectures, and Implementations VII*, July 1998.