A Graph-Theoretic Approach to Timbre Matching

Thomas Lysaght and Joseph Timoney

Department of Computer Science, NUI Maynooth, Maynooth, Co. Kildare, Ireland

{tlysaght,jtimoney}@cs.may.ie

Abstract. *This paper presents a novel approach to the matching problem associated with timbre morphing. In particular a graph-theoretic technique, that of subgraph isomorphism, is applied to find correspondences between graph representations of a feature set for each timbre. The features are identified from Wigner distributions of the sounds using an adaptation of the McAulay-Quatieri sinusoidal analysis techniques. These features are then interpreted as nodes in graph representations of timbre. An audio morphing application can then be implemented by the application of spatial warping and linear interpolation of the Wigner distributions based on the correspondences established.*

1. Introduction

Timbre Morphing is the blending of two individual sounds to form one new sound that exhibits the timbral qualities of both. To ensure a perceived smoothness and naturalness in the timbral transitions in the new sound it is necessary to match and align important features of both sounds. Traditional approaches to audio morphing use feature vectors that are based on spectral amplitude and harmonic location, and have relied on dynamic time warping for the matching of signal features whose accuracy is highly dependent on the path constraints imposed [Slaney 1996, Osaka 1995]. To demonstrate an alternative approach for timbre morphing this paper presents the use of graph-theoretic techniques for matching timbre features of musical tones. Graph representations of timbre enable matching based on a number of features, and an algorithm for subgraph isomorphism is adapted for this purpose. To reduce arbitrariness in the feature selection, a perceptually motivated choice of feature for musical signals, the peak of attack feature within the transient and its associated spectral values are chosen. Section 2 next describes the time-frequency representation used for timbre and section 3 details the feature identification procedure. Sections 4, 5 and 6 describe subgraph isomorphism, detail graph representations for timbre and gives results for the subgraph isomorphism approach respectively. Section 7 outlines its incorporation into morphing techniques. Finally, Section 8 discusses conclusions and future work.

2. Time-frequency Representation

To represent a signal in terms of the evolution of its frequency content over time there are a number of available options. The most widely used of these is the spectrogram. However, for such an application as audio morphing that requires accurate location both

in frequency and time of the important partials that constitute the sounds to be morphed, the resolution limitations of the spectrogram become apparent. An alternative is to use the Wigner distribution, which has been gaining popularity in signal analysis and audio processing [Janse 1983, Cohen 1995, Furlong 2001] due to its excellent time and frequency localisation properties. However, when applying the Wigner distribution a compromise must be made to render it computationally efficient and for using it to examine multicomponent signals some form of smoothing must be incorporated into its implementation. This results in what is called the Smoothed Pseudo Wigner Distribution (*SPWD*) [Pielemeier 1986]

$$SPWD(n,\theta) = 2\sum_{k=-L+1}^{L-1} e^{-jk2\theta} p(k) \sum_{l=-M+1}^{M-1} z(l) g(n,k) \qquad (1)$$

where $p(k) = w(k)w^*(-k)$ and $g(n,k) = s(n+k)s^*(n-k)$, with $w(k)$ being a suitable window function, $s(n)$ is the signal and $z(l)$ is also a smoothing window function.

In particular for the *SPWD* the time smoothing of the temporal correlation function $g(n,k)$, as shown in (1), allows for subsampling of the distribution which gives a reduction in the computational effort by as much as 100 times when compared with the full Wigner distribution. For music signals this degree of smoothing allowable is determined by the fundamental frequency of the sound.

3 Features

The categorisation of orchestral instrumental tones into three parts, namely, the `attack' portion, the `steady-state portion', and the `decay' portion offers a general guideline as to the structure of instrumental tones [Moorer 1977]. The peak of attack feature, being the culmination of the combined onset of all partials and a significant perceptual point in the attack transient of musical signals, is taken here as the principal feature for matching between partials. This feature is then used to determine spectral shapes of the sounds to be morphed from which correspondences can be established between them, and temporal alignment of the sounds before morphing occurs. Temporal alignment of the peak of attack feature is done to ensure only one such feature in the new morphed timbre.

A well-known technique, the McAulay-Quatieri sinusoidal analysis-synthesis technique [McAulay 1986] was adapted for the extraction of partials from SPWD surfaces. Figure 1 (a) shows the tracks extracted from 500msecs of a sampled trumpet sound. An iterative technique was developed for the identification of the peak of attack for each partial. The rate of attack can vary greatly from one instrumental timbre to another and, furthermore, it is not always obvious from the waveform where this peak falls. The basic procedure, therefore, was to, first, find the `global' peak of attack within the waveform transient and then to determine one for each partial (Figure 1(b)).

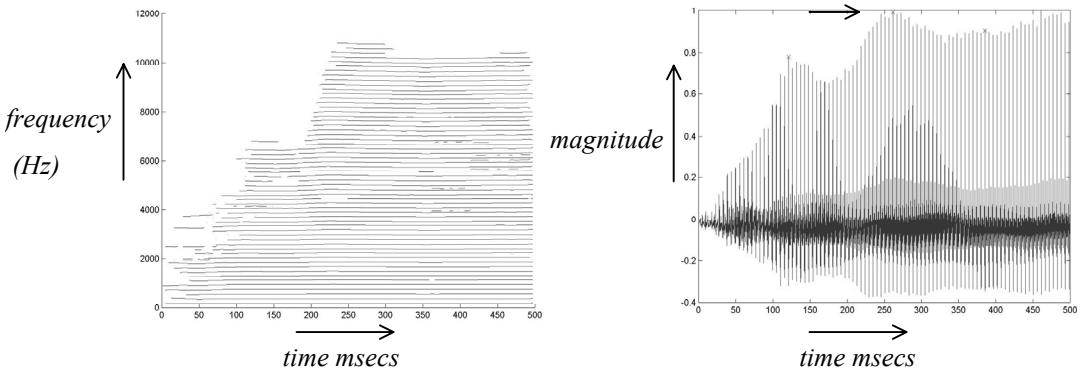


Figure 1. (a) Tracks extracted from 500msecs of a sampled trumpet sound. (b) peak of attack location on trumpet waveform.

4 Subgraph Isomorphism

The correspondence problem is central in such areas as stereo vision for establishing a left-to-right match of two views of the same scene [Redert 1999]. For audio morphing, the correspondence problem consists of identifying conjugate pairs in two 2-D spectral representations. Within the area of stereo vision, extracted image features can be organized as relational structures such as graphs that can then be matched using a number of available techniques [Messmer 1998]. In an analogous manner, the key timbral features extracted from the time-frequency representations of the sounds can also be displayed and matched using graphs. This matching problem can be very computationally intensive however if large graph structures are involved and simplifications are needed. One particular approach is based on matching similar sections of one large graph to that of another, and is known as Subgraph isomorphism. Different techniques for its computation have been proposed and specifically the subgraph isomorphism algorithm by Ullmann [Ullmann 1976] is used here to establish correspondences between the graph representations of timbre. Ullmann's algorithm is determined by means of a simple enumeration procedure with backtracking, and is designed to find all isomorphisms between a given graph G_α and subgraphs of a further graph G_β. The enumeration algorithm works by generating all possible matrices $M^{'}$, each of which is used to permute the adjacency matrix $B = \lfloor b_{ij} \rfloor$ of G_β. A further matrix, *C*, is defined as follows:

$$C = \left[c_{ij}\right] = M^{'}\left(M^{'}B\right)^{T} \tag{2}$$

where *T* denotes transposition. If it is true that:

$$\left(\forall_i \forall_j\right)\left(a_{ij} == 1\right) \Rightarrow \left(c_{ij} == 1\right)\begin{cases} 1 \le i \le p_\alpha \\ 1 \le j \le p_\beta \end{cases} \tag{3}$$

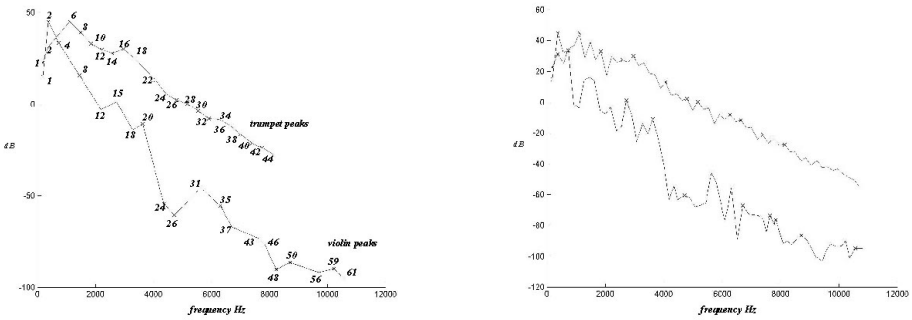
where p_α and p_β are the number of points in G_α and G_β respectively, then $M^{'}$ specifies an isomorphism between G_α and a subgraph of G_β. $A = \lfloor a_{ij} \rfloor$ is the adjacency matrix of G_α.

5 Graph Representations of Timbre

The time-frequency distributions of individual families of musical instruments exhibit characteristic properties such as formant regions in, for example, oboe tones, and spectral bandwidth concentration such as a tapering of the higher frequencies in brass instruments [Moorer 1977]. The structure of partials within the SPWD representation of timbre reflects such spectral and temporal properties. The spectral shape of the SPWD is, therefore, a suitable qualitative measure for timbre and can be approximated using the peak of attack feature as [Z★lzer 2002]:

$$S_{shape} = \{(f_1, a_1), (f_2, a_2), \ldots, (f_i, a_i)\} \tag{5}$$

where *l* is the number of partials under consideration and f_i and a_i are the frequencies and amplitudes of the peak of attack along each partial respectively.



(a) (b)

Figure 2. (a) Plot showing the spectral shape generated by 20 most significant peaks represented by partial number of peak of attack feature. (b) spectral shape generated by all spectral peaks.

Plots using the trumpet and violin sound examples are given in Figure 2. In constructing graph representations of timbre the spectral shape is represented as a connected labeled graph where the nodes encode the attributes of each partial as defined by the peak of attack amplitude. Correspondences are then established between the spectral `shapes' or graphs of both sounds, or, more particularly, between sequences of partials from both sounds. The problem of finding correspondences between timbres, therefore, is cast as a structural matching problem. A series of 1-dimensional correspondence searches between subsets of these sequences establishes correspondences between the partials of both sounds, that is, correspondences are established between sequences of predominantly low frequency partials and also between sequences of higher partials.

Subgraph isomorphism is based on a connectivity analysis of the input graphs so it is, therefore, necessary to specify the nature of connectivity in the graph representations. Each node (peak of attack) is viewed as being connected to every other node on tracks with a higher harmonic number. This constructs a directed graph where the connectivity specifies monotonically increasing sequences of partials. The graphs in Figure 3 represent the low harmonics up to about the twelfth of each sound.

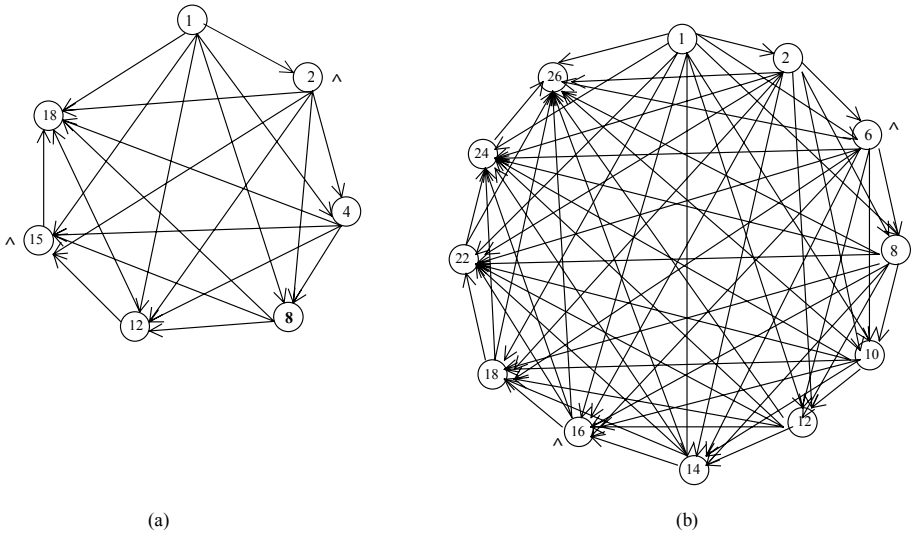


Figure 3. Graph (a) is the graph representation of the first 7 violin peaks and (b) represents the first 12 trumpet peaks both from Figure 2. The hat symbol above or adjacent to a node indicates that this peak is a maximum peak as explained in the text. The labels for each node represent the harmonic number associated with that peak .

It was decided to look for isomorphisms between sets of low harmonics only and sets of high harmonics only, thus, avoiding matching the more prominent low harmonics with lower intensity high harmonics as it was anticipated that morphing very low with very high harmonics would give rise to unrealistic instrumental sounds, while although interesting in themselves, would not reflect the properties of the input sounds. The subgraph isomorphism algorithm then tested for isomorphism based on equality of the integer matrix values between the subgraph permutation matrix A and the result permutation matrix C (equations 2 & 3).

In order to identify prominent peaks or formant regions within the spectral shape, each diagonal matrix element was allowed to hold an integer value representing any of two features:

- a 2 in a diagonal element implies that this node represents a maximum peak of attack along the spectral shape. Here, a maximum peak is specified as having neighboring peaks which are both lower in amplitude.
- a 5 added to a diagonal element represents the amplitude range for that peak. All peaks are divided into two amplitude ranges as a broad means of distinguishing high amplitude from low amplitude harmonics. The division

between the ranges is chosen arbitrarily as being above or below two-thirds the highest amplitude in the SPWD.

These restrictions enforced matches only between nodes with the same properties, that is, between nodes representing significant harmonics and formant regions and between harmonics which belong to the same amplitude range. The numbers 2 and 5 are chosen arbitrarily. A graph node for a peak within the higher amplitude range will, therefore, be denoted by 7=5+2 where the search includes both properties combined.

6 Tests

The sampled sounds used in morphing experiments are taken from the MUMS CDs consisting of 16-bit samples at a rate of 44.1kH. Experiments are run on trumpet (F#3, MUMS Vol. 2, Track 16, Index 1) and on stopped violin (F#3, MUMS Vol.1, Track 6, Index 7). Both sounds have a pitch which settles at about 193Hz and duration of about 500 msecs. In the calculation of the SPWD as per equation 1 a 2048-point frequency-smoothing window was used resulting in a bin size of 10.766Hz. This allowed for more than 50 harmonics in each sound to be used in analysis. The time smoothing window cutoff frequency was estimated to be at about $2\pi \times 180$ Hz to allow for frequency modulation in the partials.

Two test cases were used in finding correspondences between the trumpet and violin partials. In the first test a subset of the 20 most significant peaks (marked with harmonic number in Figure 2 (a)) was chosen to represent the spectral shape of each sound. This was done in order to match the broad spectral shape of each timbre. The remaining partials between these peaks were then matched separately using a heuristic approach based on peak amplitude alone. The input graphs G_α and G_β of Figure 3 respectively are the input for the subgraph isomorphism routine: G_α is the graph representation of the first 7 violin peaks and G_β represents the first 12 trumpet peaks of the spectral envelopes peaks marked with harmonic number in Figure 2 (a). In Figure 3, the hat symbol above or adjacent to a node indicates that this peak is a maximum peak indicating a possible formant region. The labels for each node represent the harmonic number associated with that peak.

The test for isomorphism represents the attempt to match the first 7 violin peaks shown with 7 of the first 12 trumpet peaks. A total of 34 isomorphisms were found between these graphs, four of which are given below:

$$\mathrm{iso}_1(\underline{vl},tr) = \mathrm{m}_{1,1},\ \mathrm{m}_{2,6},\ \mathrm{m}_{4,8},\ \mathrm{m}_{8,10},\ \mathrm{m}_{12,12},\ \mathrm{m}_{15,16},\ \mathrm{m}_{18,24}$$

$$\mathrm{iso}_2(\underline{vl},tr) = \mathrm{m}_{1,1},\ \mathrm{m}_{2,6},\ \mathrm{m}_{4,10},\ \mathrm{m}_{8,12},\ \mathrm{m}_{12,14},\ \mathrm{m}_{15,16},\ \mathrm{m}_{18,26}$$

$$\mathrm{iso}_3(\underline{vl},tr) = \mathrm{m}_{1,2},\ \mathrm{m}_{2,6},\ \mathrm{m}_{4,8},\ \mathrm{m}_{8,10},\ \mathrm{m}_{12,12},\ \mathrm{m}_{15,16},\ \mathrm{m}_{18,24}$$

$$\mathrm{iso}_1(\underline{vl},tr) = \mathrm{m}_{1,2},\ \mathrm{m}_{2,6},\ \mathrm{m}_{4,10},\ \mathrm{m}_{8,12},\ \mathrm{m}_{12,14},\ \mathrm{m}_{15,16},\ \mathrm{m}_{18,26}$$

For each match element $\mathrm{m}_{i,j}$ in the four isomorphisms above the *i* represents a node on the graph representation of the violin peaks in Figure 3 (a) and the *j* a node on the graph representation of the trumpet peaks in Figure 3 (b). The first isomorphism is chosen for morphing because this represents the closest matching of harmonics based on the proximity of harmonic numbers. However, it is possible that each of the other three isomorphism may produce equally useful morphs but they may be less perceptually

satisfying. In the second test all partials, that is, all the peak points for both instruments shown in Figure 2 (b) were used in the match. 24 isomorphisms were found two of which are given below:

$$iso_1(tr,vl) = m_{1,1},\ m_{2,2}, m_{3,3}, m_{4,6}, m_{5,8}, m_{6,9}, m_{7,10}, m_{8,12}, m_{9,13}, m_{10,16}, m_{11,17}, m_{12,18}$$

$$iso_1(tr,vl) = m_{1,1},\ m_{2,2}, m_{3,3}, m_{4,6}, m_{5,8}, m_{6,9}, m_{7,11}, m_{8,12}, m_{9,14}, m_{10,16}, m_{11,17}, m_{12,20}$$

7 Alignment, Interpolation and Signal Synthesis

To align the peaks of attack in the corresponding partials polynomial spatial warping was used. Subsequently, linear interpolation (mostly mean value) was used to morph between the warped SPWD partials. A variety of techniques for signal synthesis from SPWD representations are available to recover the morphed sounds [Melody 1998] [Bartels 1986].

8 Conclusions

A novel approach to correspondence matching of timbre for morphing applications has been offered. A method for peak of attack identification for music signals has been described from which timbral features can be extracted using a time-frequency representation. Once the features are obtained graph representations can be formed that will allow correspondence matchings based on multiple features to be made. These correspondences can then be used to perform a timbre morphing between the sounds.

Future work will explore further graph techniques such as maximal cliques and also include a perceptual evaluation of the morphed sounds.

9 References

Boudreaux-Bartels, G., "Time-Varying Filtering and Signal Estimation Using Wigner Distribution Synthesis Techniques". *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. ASSP-34, No. 3, June 1986.

Cohen, L. "Time-Frequency Analysis". *Prentice Hall*, 1995.

Furlong, D., and O'Donovan, J., "Quantitative characterization of perceptually relevant artifacts of synthetic reverberation using the earwig distribution". *Digital Audio Effects-01*, Limerick, Ireland, December 2001.

Janse, C P., and Kaizer, A. J., "Time-Frequency Distributions of Loudspeakers: The Application of the Wigner Distribution". *J. Audio Eng. Soc.,* Vol. 31, No. 4. April 1983.

McAulay R.J., and Quatieri, T.F.,"Speech analysis/synthesis based on a sinusoidal representation". *IEEE Trans. on Speech and Signal Processing*, Vol. ASSP-34, No. 4, August 1986.

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

Messmer, B.T., and Bunke, H., "Subgraph Isomorphism in Polynomial Time". *Lecture Notes in Computer Science Graph Theory - ECCV'98*, Springer-Verlag, Berlin, 1998

Moorer, J.A., "Signal Processing Aspects of Computer Music: A Survey". *Proceedings of IEEE*, Vol. 65, No. 8, August 1977.

Osaka, N., "Timbre interpolation of sounds using a sinusoidal model". *Proceedings International Computer Music Conference*, 1995.

Pielemeier, W.J., and Wakefield, G.H., "A high-resolution time-frequency representation for musical instrument signals". *Journal of Acoustical Society of America*, 99(4), Pt. 1, April 1986.

Redert, A., Hendriks, E., and Biemond, J., "Correspondence Estimation in Image Pairs". *IEEE Signal Processing Magazine*, 16(3), 1999.

Slaney, M., Covell, M., and Lassiter, B., "Automatic Audio Morphing". *Proceedings IEEE International Conference Acoustics, Speech and Signal Processing*, Vol. 2, 1996.

Zolzer, U., "Dafx-Digital Audio Effects". *John Wiley & Sons*, 2002.

Ullmann, J. R., "An Algorithm for Subgraph Isomorphism". *Journal of the Association for Computing Machinery*, Vol. 23, No. 1, January 1976.