Multidimensional timbre analysis of shakuhachi honkyoku

Deirdre Bolger, Niall Griffith

▶ To cite this version:

Deirdre Bolger, Niall Griffith. Multidimensional timbre analysis of shakuhachi honkyoku. Conference on Interdisciplinary Musicology (CIM05), Society for Interdisciplinary Musicology, Mar 2005, Montreal (Canada), Canada. hal-03991696

HAL Id: hal-03991696
https://amu.hal.science/hal-03991696

Submitted on 16 Feb 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Multidimensional timbre analysis of shakuhachi *honkyoku*

Deirdre Bolger, Department of Computer Science and Information Systems, University of Limerick, Ireland
deirdre.bolger@ul.ie

Niall Griffith, Department of Computer Science and Information Systems, University of Limerick, Ireland
niall.griffith@ul.ie

Proceedings of the Conference on Interdisciplinary Musicology (CIM05)
Actes du Colloque interdisciplinaire de musicologie (CIM05)
Montréal (Québec) Canada, 10-12/03/2005

Abstract

A method of analysing timbre in melody that takes account of its multidimensional characteristic is described. The aim of this analysis technique is to uncover the role of timbre as a structural element in melody. Gaining an understanding of how an abstract, multidimensional sound phenomenon such as timbre is used in a structured manner can lead to important insights into potential cognitive abilities; our ability to form mental representations of abstract phenomena so that they can be conserved and, subsequently, used in an organised or structured manner. This method of analysis comprises two main components, a first which carries out a time-frequency analysis of the melodic signal and processes the signal data using DSP techniques that model aspects of auditory processing before calculating the timbre descriptors, and a second that represents the melody in terms of its timbral changes rather than its absolute timbral values. A self-organising feature map is used to reduce the timbral detail and the dimensionality of the timbral representation, and to contrast-enhance the timbral changes. An example of the implementation of this analysis technique is presented using an extract from a Japanese shakuhachi *honkyoku* melody, chosen because of its accepted exploitation of timbre.

Introduction

Traditionally, analyses of the structure of recorded music have been notation-based in their focus, concentrating on the musical elements of pitch and rhythm. If the musical style being analysed is structured principally through an interaction of these two musical elements, such an analyses can yield a very comprehensive account its structure. The fact that pitch and rhythm based analyses of musical structure have been so popular is due, in no small part, to the fact that both these musical elements are of central importance in Western music. However, by focussing on pitch and rhythm, we are neglecting the possible contribution of another, very important, musical phenomenon, that of timbre.

The timbre or sound quality of a sonic event plays a very important role in our perception of sounds and is considered by some as the most important sound phenomenon due to its contribution to sound identification and discrimination (Schellenberg *et* al, 1999; Menon *et al*, 2001). However, despite its importance, the comprehensive analysis of its use in musical structuring has been hampered by its, now, well accepted multidimensionality. Unlike pitch and rhythm, which can be well described by changes along a single physical dimension, the timbre of a sound results from the interaction of several physical attributes, particularly frequency and amplitude and can be described by several descriptors. To use a sound element as a structuring device in creating music, it has to be possible to structure that element in a systematic manner. Pitch structures take the form of scales, which consist of orderings of discrete pitch entities according to their frequency and rhythmic structures consist of a hierarchy of relative durations. These basic structures allow one to build up complex pitch and rhythmic relationships that lend coherence to musical compositions. Attempts to build timbre structures that show the same hierarchical characteristics as those of pitch and rhythm are fraught with difficulty given its abstract, multidimensional nature (Lerdahl, 1987). However, our ability to conserve, identify and discriminate timbre strongly suggests that we must employ some mechanism of structuring it and these mechanisms may also apply in musical contexts.

The extent to which timbre is used in a structured manner in music creation differs depending on the musical style in question. The dominant feature of Western music is its reliance on pitch and harmonic pitch relations as the principal carrier of musical structure, the contrast between the timbres of individual instruments is exploited but, as a structural element, it plays a lesser role than pitch and rhythm. In contrast, melodic music, which does not rely on harmony for structure, is more likely to exploit timbre in this capacity. Given that a vast number of musical cultures worldwide are primarily melodic, non-Western melodic musical styles provide a very good platform for the investigation of timbre structures. This paper employs an example from a Japanese shakuhachi *honkyoku* melody to describe the implementation of a multidimensional analysis of timbre in a melodic context.

Japanese shakuhachi *honkyoku*

In carrying out an investigation of timbre as a structural element in melody, it was important to base it in a melodic style in which timbre plays an important role. For this reason Japanese shakuhachi melodies of the *honkyoku* tradition were chosen. The shakuhachi is a bamboo end-blown flute that was traditionally played by mendicant Buddhist monks called *komuso*. The *honkyoku* style of shakuhachi melody was pioneered by the *komuso* as a meditative tool or *hoki*. For this reason the melodies are solo, rhythmically free and have a structure that is governed in no small part by the breathing patterns of the musician. The central role of breathing in the execution of *honkyoku* melodies has given rise to "tone-cells" known as "one-breath-tones" or *issokuon* (Gutzwiller and Bennett, 1991) that compose all pieces of this style. These tone-cells are executed in a single breath and are separated by a pause during which the musicians inhales and prepares for the followng tone-cell, thus, they can be perceived as distinct structural elements.

As a form bearing musical element, timbre plays a role in the internal structure of tone-cells. Gützwiller and Bennet have shown that individual tone-cells have a tri-partite structure caused by variations in pitch, tone quality and dynamics. Such variations are a result of the execution of the techniques of *meri* and *kari*, which flatten and sharpen the tone respectively. Both are executed by movements of the head. A *meri* tone has a sound that is softer and less stable in terms of pitch. In contrast, a *kari* tone is strong, with a more stable pitch and higher amplitude and, as a result, is considered the "main sound" of a tone-cell[1]. Gützwiller and Bennett describe the evolution within a tone-cell as generally constituting a change from a *meri* to a *kari* and back to a *meri*. As will be shown in the section on the timbre descriptors, this *meri-kari* evolution has implications for the timbre of the tone-cells and the change in their timbre over time. Thus, using Tsang's concept of timbre segmentation, each of the three phases of a tone-cell may constitute a timbral segment. Given this pre-defined timbral structure, it provides a very good platform for investigating the effectiveness of the timbre analysis technique in extracting significant changes. In describing the implementation of the multidimensional timbre analysis, use is made of an extract from a recording of a famous *honkyoku* melody, "Kokû", performed by shakuhachi master, Kozan Kitahara.

Timbre analysis technique

The timbre analysis technique is composed of three main parts,

1. The input melodic signal data is analysed using short-time-Fourier-transform (STFT) and processed to take account of frequency weighting (equal loudness curves), spectral masking.
2. The timbral descriptors are calculated and the time-dependent timbre representation is integrated over time.
3. A SOFM is used to cluster the time-dependent multidimensional timbre data allowing the calculation of timbre differences. The calculated timbre differences are used to extract significant timbral changes in the melodic signal.

Part 1: Analysis and processing of melodic signal

The melodic signal is analysed using a STFT, which yields a time-dependent representation. Shakuhachi *honkyoku* melodies can often be in the order of 5 to 10 minutes, which means that the analysis of an entire melody is not possible due to computational constraints. However, the fact that *honkyoku* melodies are composed of distinct melodic phrases known as "tone cells" makes it possible to isolate sections for analysis. For the following analysis, a single tone-cell has been isolated from Kitahara's performance of "Kokû". This tone-cell was chosen as it constitutes a motif that is evident in all performances of "Kokû" and, from listening, appears to play an important role as a perceptual

[1] From communication with shakuhachi player, Jürg Zürmuhle (9/9/01).

anchor in this, often very long, melody. The transcription and waveform of this motif is shown in figure 1.

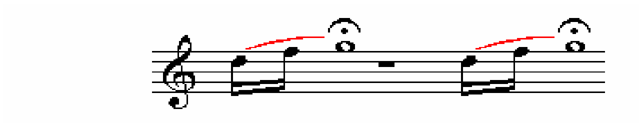

Figure 1: Transcription of tone-cell motif from recordings of "Kokû" by Kitahara.

The STFT is implemented using a sample frequency of 44.1kHz and a window length of 2048 samples, thus giving a frequency resolution (Δf) of 43Hz and a time resolution (Δt) of 0.0116 seconds. This time-dependent representation of the melodic signal will be preserved throughout subsequent stages of the analysis. Before calculating the timbre descriptors, the signal data undergoes the following processes:

1. A **peak detection routine** is applied to each time frame (t) of the input signal, which discards spurious components that arise as a result of the windowing applied in the STFT calculation.
2. **Equal-loudness curves** (Fletcher and Munson, 1933) are applied to each of the spectral components for every t. This converts the "sound intensity level" (dB) of a component to "loudness level" (dB) by taking account of its frequency. Thus, values of "loudness level", also known as "phons", are measured on a frequency-compensated decibel scale.
3. A **simultaneous masking** algorithm is implemented for each value of t, thus isolating those components that are most perceptually significant. Auditory masking is not simply a function of the relative intensity of components, but also relates to the frequency resolving characteristics of the peripheral auditory system, which is evident in the frequency response of the basilar membrane (BM). The frequency sensitivity of the BM is tonotopically organised so that certain points along it display a maximum responses to particular frequencies and a progressive decay in the response to these frequencies as one either increases or decreases the input frequencies. The limit of the response to a particular frequency is governed by the width of the frequency band, referred to as the "critical band", around a point of maximum response. Thus, if two components that are so close in frequency that they fall in the same critical band are sounded simultaneously, their BM response will interact. Based upon this understanding, this masking algorithm involves the generation of the BM response to the input frequencies, known as the "excitation pattern" (Moore and Glasberg, 1983, 1986 and 1987). To represent the excitation pattern, an approximation of the shape of an auditory filter response is applied to the energy calculated at each point along the BM, broken into steps of ¼ critical-band. This approximation is generated using the linear gammatone filter function (Johannesma, 1972, Patterson *et al*, 1995). It is from the excitation pattern that the unmasked components are isolated before the calculation of timbre descriptors.

Part 2: Calculating timbre descriptors

Four timbre descriptors were selected to describe the timbre space of of the melodic extract: spectral centroid, irregularity, harmonicity and roughness, all four of which are spectrally-based. While it is accepted that a comprehensive account of timbre should include timbre dimensions that are temporal, such as the attack time, it was decided to confine this analysis to four spectral descriptors as it is based on melody and also because of certain characteristics of shakuhachi melody. Grey (1977) concluded, on the basis of experiments on the discrimination of timbres in a melodic context, that the spectral aspects of a tone are emphasised in a melodic context, while the temporal aspects of a tone are emphasised in an isolated context. This finding along with the fact that shakuhachi *honkyoku* is generally characterised by long, sustained tones that feature pitch and timbral changes is taken as a justification for the concentration on spectrally-based timbre descriptors in this analysis. However, future implementations of the analysis could very easily incorporate both spectral and temporal timbral measures.

Spectral centroid

This descriptor accounts for the concentration of energy in the spectrum of the input tone is correlated with the attribute of brightness. A high spectral centroid value implies a concentration of energy at the higher end of the spectrum, while a low spectral centroid value indicates that the energy is concentrated in the lower frequencies. It is calculated for each value of *t* as follows:

$$centroid(t) = \frac{\left(\sum_{n=1}^{N} f_n a_n\right)}{\sum_{n=1}^{N} a_n}, \qquad \text{equation 1}$$

where t is the current time (in seconds), n is the partial index, N is the total number of partials, f_n is the frequency of partial n and a_n is the amplitude in linear intensity (W/m^2) of partial n. The resulting *centroid*(t) value is normalised by the fundamental, f_o, calculated for the spectrum at time t.

Irregularity

The calculation used here defines irregularity as the sum of the squares of the difference in the amplitude of adjacent components (Jensen, 1994). In other words, the irregularity value is based on the magnitude of amplitude change between adjacent partials and is, therefore, related to the spectral envelope at each time interval, t. Thus, a tone whose spectrum at time, t, is highly fluctuating will yield a high irregularity value. Due to its reported perceptual significance (Krimphoff, 1994; Jensen, 1994), it is included among the timbre descriptors used in this analysis. It is calculated as follows:

$$irregularity(t) = \frac{\sum_{n=1}^{N} \left(a_n - a_{n+1}\right),}{\sum_{n=1}^{N} a_n^2} \qquad \text{equation 2}$$

where t is the current time (seconds), n is the partial index, N is the total number of partials and a_n is the amplitude of the n^{th} partial (W/m^2). This irregularity calculation yields values that are generally below 1 and never any higher than 2.

Harmonicity

This attribute is generally described as the balance between the harmonic and inharmonic components of a spectrum. The harmonicity calculation is composed of two parts, a first which calculates the percentage frequency deviation from harmonic values, $\%fdiff\_harm_n$, for each component, n, of a complex at time, t, and a second part that calculates the proportion of the total intensity of the spectrum relating to harmonic, $\%harm_n$, and inharmonic, $\%inharm_n$,components for every n and time, t, and which is is also expressed as a percentage. The $\%harm_n$ is weighted by the $\%fdiff\_harm_n$ for each n and a total inharmonicity value for time, t, is found through summation, which is expressed as a percentage giving $\%harm\_fdiff_{total}$. A total value of $\%inharm_n$ is also found through summation for time, t, and expressed as a percentage giving $\%inharm_{total}$. The final harmonicity value is derived by expressing the relationship between between $\%harm\_fdiff_{total}$ and $\%inharm_{total}$ as a signal-to-noise ratio (SNR), which is defined in dBs, as follows:

$$\text{SNR} = 10\log_{10}\left(\frac{\%harm\_fdiff_{total}}{\%inharm_{total}}\right)(dB), \qquad \text{equation 3}$$

The SNR of a complex can have a maximum value of 20dB. This maximum SNR value implies that 99% to 100% of the total intensity of a complex for time, t, corresponds to the energy of harmonic components. On the other hand, a SNR of 0dB implies an equal division of intensity between harmonic and inharmonic components in the spectrum for time, t.

Roughness

Like spectral centroid, irregularity and harmonicity, the degree of roughness of a sound is based on the spectrum of a sound. However, a significant difference between roughness and the other three timbre descriptors is that while they can be calculated on the basis of the acoustic data only, the calculation of roughness requires that one takes account of both the acoustical signal data and perceptual processes. The reason for this is that the sensation of roughness is a result of the inability of the auditory system to resolve spectral components that are very close in frequency. Thus, the presence of stochastic components in the spectrum will have implications for the its calculated roughness value. In relation to shakuhachi melody, a consideration of roughness is important as the presence of noise in the shakuhachi sound is a desired characteristic. The standard "big sound" in shakuhachi blowing techniques is produced by directing about 80% of air over the top of the flute, resulting in the characteristic "windy" sound of shakuhachi music.

The calculation of roughness is based on a model of roughness developed by Vassilakis (2001), which integrates a psychoacoustical model of dissonance (Sethares, 1993), which focuses on frequency distance and a calculation of roughness that concentrates on the temporal aspect of roughness, i.e. the phenomenon of beating. The calculation of temporal roughness used is that of Vassilakis (2001) and equates the energy of the amplitude fluctuations caused by the interaction of components in the time domain with the perceived roughness.

Given two components with amplitudes A_1 and A_2, the roughness sensation resulting from the amplitude fluctuations, R_{temp}, is calculated as follows:

$$R_{temp} = (A_1 A_2)^{0.1} \times 0.5\left(\frac{2A_2}{A_1 + A_2}\right)^{3.11}, \qquad \text{equation 4}$$

where $\frac{2A_2}{(A_1 + A_2)}$ is the degree of amplitude fluctuation, AF_{deg}.

Sethares' psychoacoustic dissonance calculation uses a parameterisation of Plomp and Levelt characteristic V-curves, which relates consonance values to the relative distances of spectral components within a single critical-band. The final expression for the dissonance of two pure tones with frequencies, f_1 and f_2 with amplitudes, A_1 and A_2 is as follows:

$$d(f_1, f_2, A_1, A_2) = A_{12}\left(e^{-as|f_1 - f_2|} - e^{-bs|f_1 - f_2|}\right), \qquad \text{equation 5}$$

where a=3.5, b=5.75, $s=\frac{d^*}{(s_1 f_1 + f_2)}$, s_1=0,0207, s_2=18.96 and d^* is the frequency point of maximum dissonance, which is set at 0.24.

Integrating the two elements of the final roughness calculation gives the the following expression for the two components in a spectrum:

$$R(f_1, f_2, A_1, A_2) = (A_1 + A_2)^{0.1} \times 0.5\left(\frac{2A_2}{A_1 + A_2}\right)^{3.11} \times \left(e^{-as|f_1 - f_2|} - e^{-bs|f_1 - f_2|}\right). \qquad \text{equation 6}$$

In comparing the four timbre descriptors it has proved very useful to normalise the timbre values. This normalisation is carried out by dividing the timbre values at every time interval, t, by the maximum timbre value of all values of t. This limits the timbral range between 0 and 1 and allows for easy comparison of the timbral measures.

Time integration

The timbral descriptor calculations described in the previous sections were calculated for a value of Δt of 0.0116 seconds, giving very detailed time-dependent representations that exceed the level of granularity at which we are thought to perceive owing to limitations in temporal resolving ability of our auditory system. Therefore, it was decided to apply a temporal integration to the time-dependent timbral representations to yield representations that are more perceptually valid. This temporal integration involves the "smearing" or averaging of temporal detail. The degree of the averaging is determined by the chosen constant of integration, τ, over which the averaging is performed. The method of integration used here is taken from the process of intensity summation employed in the "long-time" theories of temporal integration (Munson, 1947; Green, 1960; Green and Swets, 1966; Zwislocki, 1960 and 1969), which presume values of τ in the order of hundreds of milliseconds. The value of τ employed here is set at 0.2 seconds. The "smoothing" effect of the temporal integration can be seen in the figure 2 for a single shakuhachi tone.

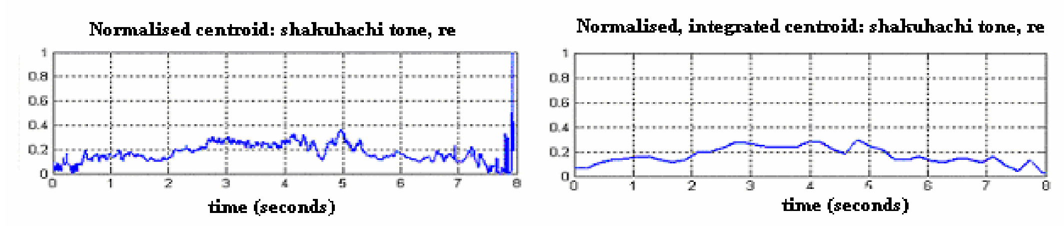

Figure 2: Comparison of non-integrated and integrated spectral centroid of shakuhachi tone, *re* or G4.

Part 3: Extraction of timbral changes

Once the timbre descriptors have been calculated using the processed signal data and the resulting time-dependent timbral representations have been time integrated, they proceed to the final stage of the analysis, the extraction of points of timbral change at certain values of *t* in the melody with the aim of investigating the interaction between timbre and significant elements of melodic structure.

Tsang (2002) and Malloch (2004) both tackled this question in the context of 20th century compositions. Their analysis techniques, while different from that described here, offer very useful insights into issues that need to be taken into account when analysing the role of timbre in a musical context. The most crucial point emphasised by both Tsang and Malloch is the importance of relative timbral values or timbral change. "Change" is considered an important phenomenon in cognition, if not more significant than absolute values (Grossberg, 2000). In both analysis techniques, different degrees of timbral change are related to structural points of varying significance. Both Tsang and Malloch are concerned with reducing the timbral data with a view to investigating patterns of timbre use that coincide with elements of musical structure. To this end, Tsang introduces the notion of "timbral segments" that are isolated on the basis of their contrasting timbre and their relative salience. Similarly, Malloch defined contrasting textures as **A-textures** and **B-textures**. Thus, in the current analysis, it was considered necessary to find a method of clustering timbres on the basis of some criteria of similarity to reduce the timbral data.

Grey (1975 and 1977) employed a multidimensional scaling technique (MDS) to create a timbre space of different instrumental timbres that is representative of a cognitive map of timbre similarity. However, this timbre classification technique required performing listening tests in which listeners gave verbal ratings of similarity. For analysis purposes, what is desired is a purely analytical method of deriving clustering timbres within a multidimensional timbre space but in a manner that is cognitively relevant. Attempts to do this have led to the use of neural networks, particularly self-organising maps (de Poli and Tonella, 1993; de Poli *et al*, 1993; Cose, de Poli and Lauzanna, 1994; de Poli and Prandoni, 1997). Cosi *et al* (1994) used a combination of auditory modelling and Self-Organising Feature Maps (SOFM) to classify instrumental timbres in a multidimensional space and achieved results that compared very well with Grey's subjective timbre space generated using MDS. The analysis method described here employs a SOFM (Kohonen, 1989) to cluster the timbre values for every time frame, which due to time integration, has a value of Δt of 0.2 seconds. The similarity between the timbral events is measured by the Euclidean distance. In implementing this classification two issues have to be carefully considered as they will strongly affect the final representation of timbral variation within the melody analysed:

1. The number of categories or classes into which the timbral events will be clustered.

A high number of categories implies that each category will be represented by a smaller portion of the timbre space and will capture the finer details of timbral change within the melody. However, with too many categories it will be more difficult to derive a structured use of timbre in the melody. Conversely, a small number of categories will mean that a larger portion of the timbre space will be represented by each cluster, which will have the effect of smoothing the noisiness in the timbre representation but may also cause smaller but potentially structurally significant timbral changes to be discarded. To deal with this problem it was decided to carry out a number of classifications using a different number of categories as part of every analysis, with the hope of achieving a hierarchical relationship between the timbral change representations resulting from the use of a different number of categories. As an initial test, three types of categorisations were used; a 10-by-10 (10x10) yielding

100 categories, 5-by-5 (5x5) yielding 25 categories, a 3-by-3 (3x3) categorisation yielding 9 categories and a 2-by-2 (2x2) giving 4 categories. It was hypothesised that the lower resolution categorisations, such as 2x2 would highlight only the most significant changes while the higher resolution categorisations, for example 10x10 and 5x5, would capture the smaller changes. Because of the simplification of the timbral representation resulting from this classification, it was decided to refer to the final representations of timbral change in the melodies as "**timbral reductions**".

2. The number of timbral dimensions (1, 2, 3 or 4D) on which the classification will be based.

As well as clustering the timbral data on the basis of similarity, the implementation of the SOFM also has the effect of reducing the dimensionality of the timbre space by mapping the 3D or 4D space onto a 2D feature map. In this analysis, the timbral data is presented to the SOFM in two-dimensional (2D), three-dimensional (3D) and four-dimensional (4D) form and the results of each compared.

Timbral reduction of standard motif from "Kokû"

Once the SOFM classification has been carried out, the result is 100, 25, 9 and 4 timbral clusters that will be used to represent the timbral changes in the timbral reduction. At this stage each time-dependent timbral event is assigned to a particular timbre cluster by finding the cluster which yields the lowest Euclidean distance between it and the timbral measure at time, *t*. However, the focus is on the magnitude of the change between these timbral clusters so we need to find a way of extracting and classifying these changes. It was decided to measure the timbral change by comparing the distance between categories, *d*, with the maximum Euclidean distance between the categories, d_{max}, in the following way:

$$x = \frac{d}{d_{\max}} \cdot \qquad \text{equation 7}$$

To categorise the values of *x* we applied an adapted version of Lerdahl's characterisation of timbral change described in his "timbral hierarchy theory" (Lerdahl, 1987). Lerdahl defined three degrees of timbre change, "strong prolongation" or an exact repetition, a "weak prolongation" or a change in a single timbre dimension and a "progression" a change on all timbre dimensions. As an initial step it was decided to define the change between timbral categories using the following three rules:

1. If $0<=x<=0.1$, the timbral change is classed as a "strong prolongation" (SP).
2. If $0.1<x<0.7$, the timbral change is classed as a "weak prolongation" (WP).
3. If $0.7<=x<=1.0$, the timbral change is classed as a "progression" (P).

It may be necessary to adjust these thresholds depending on the range of the timbral space of the melodic extract being analysed. In the timbral reductions, the timbral change categories range from 0 to 1. A SP is assigned a value of 0.25, a WP a value of 0.5 and a P a value of 1.

Figure 3 presents a 4D 3x3 timbral reduction, in green, of the standard motif of the *honkyoku* melody, "Koku" (notated in figure in figure 1) performed by Kitahara. In each case it is plotted against the pitch contour of the motif, in blue.

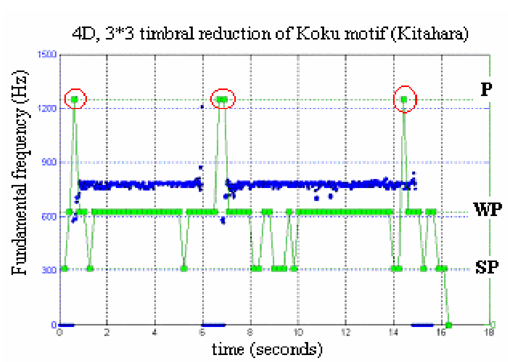

Figure 3. 4D, 3x3 timbral reduction of standard against pitch contour motif from "Kokû" performed by Kitahara.

The 4D reduction takes account of the inter-relationship between all four timbre descriptors and this is reflected in the resulting timbral reduction. Due to the interaction of all four descriptors in the classification, only the most significant changes are highlighted (circled in figure 3). Unlike 4D case, the 2D and 3D timbral reductions cannot take account of all the timbre descriptors in a single representation. For example, a 2D timbral reduction can only take account of pairs of descriptors at a time, implying that one would have to compare six 2D timbral reductions for each type of classification, 2x2, 3x3, 5x5 and 10x10. In order to make the 2D and 3D timbral reductions easier to interpret, it was decided to integrate all the timbral reductions that result from both dimensionalities. The resulting representations are called "summary timbral reductions" (STR). A 2D STR, for example, comprises the mean of the timbral reductions of the pair-wise timbral classifications. The STR emphasises features of timbral change that are consistent across all of the component timbral reductions, thus, if a "progression" occurs in all six of the 2D timbral reductions, it will feature as a "strong prolongation" in the STR. Figure 4 presents an example of a 2D and 3D, 3x3 STR for the motif performed by Kitahara. The significant points of timbral change evident in the 4D timbral reduction in figure 3 are also evident in the STRs, with the further indication of their relative significance.

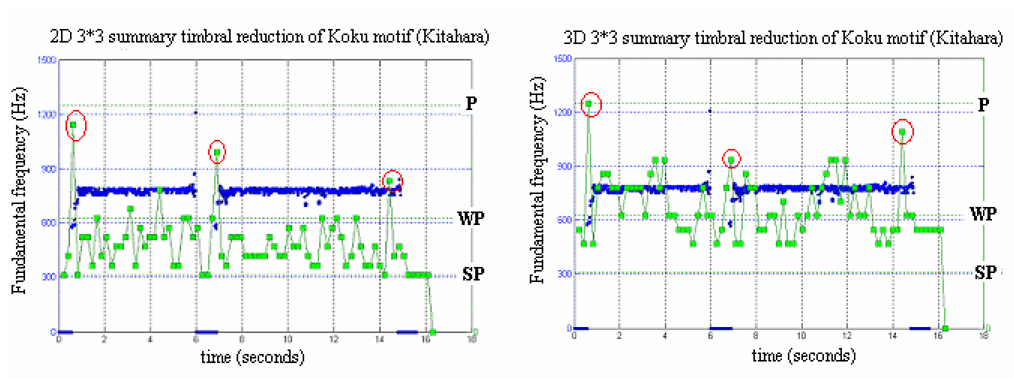

Figure 4. 2D, 3x3 STR (left) and 3D, 3x3 STR (right) of standard motif from performance of "Kokû" by Kitahara.

To get a clearer picture of the relationship between the derived STR and the evolution of the timbre measures over time in this motif, the 2D, 3x3 STR is plotted against the "mean timbral contour", shown in figure 5a, and the first-order-difference (1OD) of the "mean timbral contour", shown in figure 5b. The "mean timbral contour" is calculated as the mean of the four normalised, time integrated timbre descriptors. A comparison of the STRs and both timbral contours indicates that the reduction is effective in isolating points of significant timbre change.

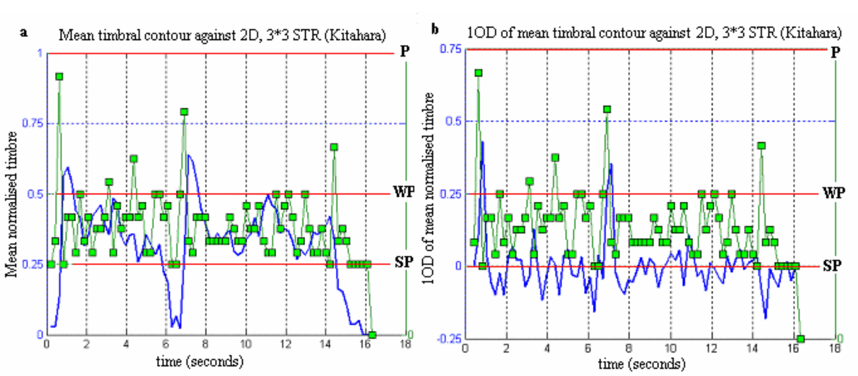

Figure 5. Mean timbral contour (a) and 1OD of mean timbral contour (b) against 2D, 3x3 STR of motif in "Koku" performed by Kitahara.

A comparison of the 4D, 3x3 timbral reduction in figure 3 and the 2D and 3D, 3x3 STRs in figure 4 reveals a consistency in the occurrence of points of significant timbral change. These common points, which have been highlighted with red circles, occur at the attack points of each repetition of the basic motif and at the end of the second motif. A clearer picture of the relationship between these three timbral reductions is given in the outline "summary timbral reductions" in figure 6 (left), which were derived by hand from the timbral reductions and STRs. Figure 6 (right) shows the relationship between the 2D, 3x3 and 3D,5x5 STRs and a 4D, 10x10 timbral reduction.

A comparison of the 2D and 3D outline STRs and the 4D outline timbral reductions in both examples in figure 6 indicate a strong relationship between them, especially in terms of the consistent presence of significant points of timbre change (highlighted by blue circles). The relationship between the three dimensionalities can be described as hierarchical due to the decreasing level of change detail evident from the 2D STR to the 4D timbral reduction. The 4D timbral reduction provides the clearest picture of an overall change structure but does not indicate the relative significance of the changes. In the 2D and 3D STRs for both the 3x3 and 5x5 conditions, a relative significance between the three significant points of change can be observed, with the first "progression" as the most dominant.

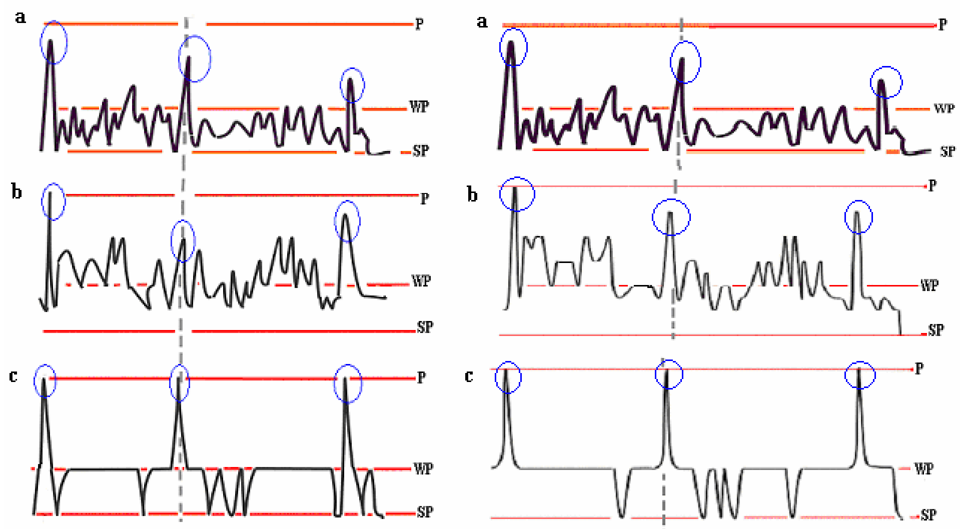

Figure 6: (Left) A comparison of 2D,3x3 STR (a), 3D, 3x3 STR (b) and 4D, 10x10 timbral reduction (c).
(Right) A comparison of 2D,3x3 STR (a), 3D,5x5 STR (b) and 4D, 10x10 timbral reduction(c) of motif.

In terms of using the timbral reductions to uncover a role of timbre as a structuring element, an interesting question relates to the extent to which these timbral change representations account for the tripartite, *meri-kari*, form of tone-cells in *honkyoku* melodies (Gutzwiller and Bennett, 1991). In terms of timbral change, a progression from *meri* to *kari* involves a movement from an area of less stability (*meri*) to an area of greater stability (*kari*). The problem with attempting to isolate the *meri* and *kari* portions of a tone cell is that they do not have a prescribed relative duration. The dominance of the *meri* or *kari* sound in a tone-cell is governed by the interpretation and technical ability of the musician. In their present form, the outline STRs and timbral reductions do not show a clear pattern of timbral change within the tone-cells, indeed, the timbre of the tone-cell appears to be in a constant state of flux with change values oscillating around the category of "weak prolongation". However, one could consider the significant points of timbre change at the attack and decay portions of each tone-cell as constituting the *meri* portion of the tone, with the lesser changes that are characterised as "weak prolongations" in the 4D timbral reductions being considered as the more stable *kari* portions.

Conclusions

We have outlined a method of analysing timbre use in melody with a view to relating values of timbre change to aspects of melodic structure and to implement this analysis in a perceptually relevant manner. The preprocessing of the signal data before calculating the timbre descriptors, the subsequent time integration of the time-dependent timbre measures and the use of a SOFM to cluster the timbre events, to reduce the dimensionality of the timbre representation and to reduce the timbral information for easier interpretation all give perceptual and cognitive relevance to this analysis technique. The most promising result gained from applying this analysis method to the investigation of timbre in an extract of a shakuhachi melody is the hierarchical relationship between the timbral reductions of differing dimensionalities. Such a representation can indicate the role of timbre at different structural levels within a melody. For example, the 4D timbral reductions give prominence to the most significant changes while smoothing out the smaller ones, thus highlighting its role in the higher level structuring of the melody. The 2D and 3D STRs, on the other hand, do not indicate as clearly these significant points of change but, in detailing the smaller changes indicate the contribution of timbre to finer melodic structure. In the case of shakuhachi *honkyoku*, the higher granularity of the timbre change representation provided by the 2D and 3D STRs can indicate the *meri-kari* structure of individual tone-cells.

The analysis technique presented here is still undergoing development and its further development will require taking into account certain issues. Firstly, given that shakuhachi *honkyoku* is an oral musical tradition a particular melody can show differences from one performance to another. Therefore, in attempting to uncover a structured use of timbre in shakuhachi *honkyoku* melodies, a single motif from one version of a melody is not sufficient. What is required is an analysis of the same motif in different versions of the same melody to obtain an understanding of the general characteristics of timbre use. Bolger (2004) carried out such an analysis on two characteristic motifs, including the one used here, from three different performances of "Kokû". This initial investigation revealed a pattern of similarities at a higher level of structure along with variations at a finer level. Secondly, although the application of the SOFM reduces timbre data, there still remains a need to identify and classify the less significant patterns of change revealed by the 2D and 3D STRs. Lastly, any conclusions regarding the role of timbre in melody have to be supported by input from native musicians to ensure that they have *emic* relevance. To this end,further development of this work will require the input of ethnomusicological investigation and experimentation.

References

Bolger, D. (2004). *Computational Models of Musical Timbre and the Analysis of its Structure in Melody*, unpublished PhD thesis, University of Limerick.

Fletcher, H. and Munson, W. A. (1933). "Loudness, Its Definition, Measurement and Calculation", *Journal of the Acoustical Society of America*, 5, 82-108.

Fletcher, H. (1940). "Auditory Patterns", *Rev. Mod. Phys.*, 12, 47-61.

Green, D. M. and Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*, New York: Wiley.

Green, D. M. (1960). "Auditory detection of a noise signal", *Journal of the Acoustical Society of America*, 36, 570-575.

Grey, J. M. (1975). "Exploration of Musical Timbre", Standford University Dept. of Music Technical Report, STAN-M-2.

Grey, J. M. (1977). "Multidimensional perceptual scaling of musical timbre", *Journal of the Acoustical Society of America*, 61, 1270-1277.

Grossberg, S. (2000). "The Complementary Brain", *Technical report CAS/CNS-TR-98-003*, Boston University.

Gutzwiller, A. and Bennett, G. (1991). "The World of a Single Sound: Basic Structure of the Music of the Japanese Flute Shakuhachi", *Musica Asiatica*, 6, 36-59.

Jensen, K. (1994). *Timbre Models of Musical Sounds*, Ph.D. Thesis, University of Copenhagen.

Johannesma, P. I. M. (1972). "The pre-response stimulus ensemble of neurons in the cochlear nucleus", In: *Symposium of Hearing Theory*, Eindhoven, Holland, 58-69.

Kohonen, T. (1989). *Self organisation and the associative memory*, Berlin, Springer-Verlag.

Krimphoff, J. S., McAdams, S. and Winsberg, S. (1994). "Caractérisation du timbre des sons complexes. II, Analyses acoustiques et quantification psychophysique", *Journal de Physique IV*, Colloque C5, 4.

Lerdahl, F. (1987). "Timbral Hierarchies", *Contemporary Music Review*, 2(1), 135-160.

Malloch, S. (2004). "Timbre and Technology: An analytical partnership", *Contemporary Music Review*, 19(2), 53-79.

Menon, V., Levitin, D. J., Smith, B. K., Lembke, A., Krasnow, B. D., Glazer, D., Glover, G. H. and McAdams, S. (2002). "Nueral Correlates of Timbre Change in Harmonic Sounds", *NeuroImage*, 17, 1742-1754.

Moore, B. C. J. and Glasberg, B. R. (1983). "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns", *Journal of the Acoustical Society of America*, 74(3), 750-753.

Moore, B. C. J. and Glasberg, B. R. (1986). "The use of non-simultaneous masking to measure frequency selectivity and suppression", In: *Frequency Selectivity in Hearing*, 170-250, Moore, B. C. J. (Ed.), London: Academic Press.

Moore, B. C. J. and Glasberg, B. R. (1987). "Formulae describing frequency selectivity as a function of frequency and level and their use in calculating excitation patterns", *Hearing Research*, 28, 209-225.

Munson, W. A. (1947). "The growth of auditory sensation", *Journal of the Acoustical Society of America*, 44, 187-203.

Patterson, R. D., Allerhand, M. and Giguère, C. (1995). "Time-domain modelling of peripheral auditory processing: A modular architecture and software platform", *Journal of the Acoustical Society of America*, 98, 1890-1894.

Schellenberg, E.G,. Iverson, P., and McKinnon, M. C. (1999). "Name that tune: Identifying popular recordings from brief excerpts", *Psychon. Bull. Rev.*, 6(4), 641-646.

Sethares, W. (1993). "Local consonance and the relationship between timbre and scale", *Journal of the Acoustical Society of America*, 94(3), 1218-1228.

Sethares, W. (1998). *Tuning, Timbre, Spectrum, Scale*, London: Springer Verlag.

Takeuchi, A. H., and Hulse, S. H. (1993). "Absolute pitch", *Psychol. Bull*, 113(2), 345-361.

Tsang, L. (2002). "Towards a theory of timbre for music analysis", *Musicae Scientiae*, 4(1), 23-52.

Vassilakis, P. N. (2001), *Perceptual and Physical Properties of Amplitude Fluctuation and their Musical Significance*, PhD. thesis, University of California, Los Angeles.

Zwislocki, J. J. (1960). "Theory of temporal auditory summation", *Journal of the Acoustical Society of America*, 32, 1046-1060.

Zwislocki, J. J. (1969). "Temporal summation of loudness: An analysis", *Journal of the Acoustical Society of America*, 46, 431-440.

Discography

Japon, Sankyoku, Ensemble Yonin no Kaï, Ocora C560070, Radio France, 1995.

View publication stats