Journal of New Music Research
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/nnmr20

Generative Structures in Improvisation: Computational Segmentation of Keyboard Performances

Roger T. Dean[a], Freya Bailes[ab] & Jon Drummond[a]
[a] University of Western Sydney, Australia
[b] University of Hull, UK
Published online: 06 Feb 2014.

Journal of New Music Research, 2014
Vol. 43, No. 2, 224–236, http://dx.doi.org/10.1080/09298215.2013.859710

Routledge
Taylor & Francis Group

Generative Structures in Improvisation: Computational Segmentation of Keyboard Performances

Roger T. Dean[1], Freya Bailes[1,2] and Jon Drummond[1]

[1]*University of Western Sydney, Australia*; [2]*University of Hull, UK*

(*Received 28 March 2013; accepted 22 October 2013*)

Abstract

We assessed the hypothesis elaborated by Pressing and by the first author that improvisation is normally based on the generation of segmental contrasts in musical features. Nine experienced improvisers performed a series of free- and 3-section referent-based pieces, each of about three minutes. Each player undertook eight improvisations with 'referent' instructions, preceded and followed by a free improvisation. A MIDI-equipped grand piano was used, and audio and MIDI were recorded. Computational analyses of the MIDI data assessed whether performers realized the referents successfully, and then determined the large-scale segmentation of the free improvisations. The performers almost always fulfilled the referents (68/72 improvisations, $p < 0.0001$), and their free improvisations were also susceptible to large-scale segmentation ($p < 0.005$). Since performers were only told of the referent structures as the session proceeded, the similar segmentation of both the first and second free improvisation suggests that such segmentation is common in solo free improvisation, in accord with the hypothesis.

Keywords: performance; music analysis; representation; tonality

1. Introduction

Improvisation in music is widespread across cultures and periods (Bailey, 1980; Smith & Dean, 1997). It is of particular interest because of the need for real-time application of a host of creative processes, with accompanying cognitive demands (Pressing, 1988). At least some of these demands are reduced or absent during interpretive performance. Most improvisation has probably taken place within quite codified structures (which we term 'referents'), such as the Indian rag, which defines scale structure and often a high level progression from sections without meter to those with highly dynamic rhythmic consistency. In Western improvisation of the last half-century, jazz has been the dominant form, also mainly reliant on pre-formed melodic and particularly, repetitive harmonic structures as the basis for improvising: for example, the 32 bar popular song, or the 12-bar blues. Many new structures have been introduced into jazz improvising, and particularly since Lennie Tristano around 1950, much free improvisation has emerged, in the sense of improvisation not so reliant on preformed structures (Dean, 1992a; Jost, 1974). Exemplars are Ornette Coleman and John Coltrane (around 1960), and the later free improvisation traditions in the US and Europe (for example, in the work of Cecil Taylor, Willem Breuker, Evan Parker, Derek Bailey).

In discussing improvisation broadly, Jeff Pressing elaborated the view that regardless of the presence or absence of preformed structure, improvisers tend to structure their performances on the relatively large time scale into distinct segments, successive parts of the piece which are differentiated by high level musical parameters (such as motivic structure, rhythmic structure, sustained acoustic intensity, tonality). Furthermore, he argued that points of transition between these segments are particularly important times, and in group improvisation, actively negotiated between participants (Pressing, 1988). The first author of this paper also developed closely related ideas slightly later in books about open-ended approaches to improvisation (Dean, 1989, 1992a). A shared implication was that musical micro-structure is used in improvisation as a vehicle for the creation of macrostructure. This might be a common and

Correspondence: Roger T. Dean, MARCS Institute, University of Western Sydney, Locked Bag 1797, Penrith DC, NSW 2751, Australia. E-mail: roger.dean@uws.edu.au

reasonable view about composed music (as argued by Xenakis (Xenakis, 1970, 1971)), but at the time was a contested view of Western improvisation, an activity which several musicologists and administrators viewed with disdain (Dean, 1992b; Smith & Dean, 1997).

From both a cognitive and musical perspective, the central questions raised by a view espousing large-scale segmentation that is generated as micro-structure becomes macrostructure concern the modes of production and organization of the improvised behavior. Pressing went on to commence analytical studies of improvised work, which were unfortunately left incomplete at his early death. One important primary article concerned pitch structures in jazz (Pressing, 1982), while in his main fully published piece of work analyzing improvisations (Pressing, 1987) he studied two free improvisations he performed himself on the DX7 synthesiser, acquiring MIDI data from its output. He provided measures of timing, dynamics, and legatoness, on the micro-level, and related these to macrostructural considerations taken from traditional music theory (tonality, rhythm, motive, pitch class). One of the improvisations had an external pulse. In all cases microstructural elements also constructed traditionally defined segments (based largely on concepts of performance phrasing), showing that one could be created from the other. For example, Pressing reveals statistical clustering of features such as attack velocities, implying that perhaps he sometimes treated them categorically in performance. Notably, in these short pieces (55 and 97 s long), the (non-computationally) analysed segments ('event clusters') were in turn very short: in one case, from 1.3 to 10.8 s (mean 5.1). As he concludes (p.172), 'microstructural recording is only sometimes…sufficiently detailed to uniquely determine the event parsing' implied by the segmentation theory.

Given this success with microstructural analysis, we hypothesized that professional improvisers would be capable of generating performances which fulfilled simple verbal referent musical structures (see below) given to them immediately prior to the performance; and that even without such instructions, during free improvisations, they would generate related segmental structure, occupying periods of many seconds. So our purpose here was to assess whether computational analysis of solo keyboard improvisations, could statistically detect segmentation at such longer durations (macrostructural) both in free improvisations, and in improvisations for which musically-defined referents were provided. Correspondingly, the study assesses the degree to which improvisers fulfil these referents (or are willing to do so). Such a referent, for example, might specify that the improvisation should comprise three successive segments, loud-soft-loud, or more subtly, tonal-atonal-tonal. Note that the first improvisation that each performer was asked to play was always free, and there was no reference to segmentation until after that had been performed. The broader context of this study is a program of research on both solo and duo keyboard mprovisations, addressing relationships between large-scale musical segmentation, continuous psychophysiological variables (notably skin conductance), and inter-performer musical relationships, such as leading and following.

Computational segmentation of symbolically encoded music (where notation is encoded or a note-centred piece is recorded as MIDI or another computable encoding) has mainly concerned segmentation at the phrase level, roughly corresponding to the micro-structural level described above. A recent review (Pearce, Müllensiefen, & Wiggins, 2010) compares segmentation based on the Grouping Preference Rules of Lerdahl and Jackendoff's (1983) Generative Theory of Tonal Music with those based on Narmour's (1989) Implication-Realisation theory, and with their own computational approaches based on n-gram and information theoretic analyses. In general, such methods can predict reasonably well the 'ground truth' of listener perceptions. For example, studying six popular songs rendered in idioms ranging amongst pop, rock and jazz, it was found that musically-trained (but not professional) listeners requested to identify 'segment boundaries' on average reported them roughly every 3 s, corresponding to short melodic phrases, and these coincided well with a subset of those identified computationally (Bruderer, McKinney, & Kohlrausch, 2012). It is also well established that listeners segment familiar tonal music (Deliege, 1987) and less familiar atonal works (Dibben, 1994) including piano music (Clarke & Krumhansl, 1990) similarly. The same is true of segmenting electroacoustic sound (Bailes & Dean, 2007a, 2007b). These studies have also implied that segmentation on larger timescales could be based on the melodic phrase groupings.

The community of music information retrieval (MIR) has developed complex approaches to segmentation of musical audio (i.e. not symbolically encoded), mainly of popular music. The most recently used audio datasets (such as MRX10V2, containing 1048 short pieces) have expanded to include familiar and fairly commercial jazz, world, live and classical items (Ehmann, Bay, Downie, Fujinaga, & De Roure, 2011). The approaches use algorithms based on multiple low level audio features, with good success, again mainly focusing mainly on short-term segmentation (see for example http://www.music-ir.org/mirex/wiki/2012:Structural_Segmentation).

Our purposes here were thus somewhat distinct from those just reviewed: while using symbolic music features, we explored improvisations which lacked the repetitive structures of popular songs and were not necessarily tonal or regularly pulsed, and with a view to identifying long-term structural segmentation. Given our interest in the processes by which these structures are generated, our methods seek to exploit the musical features requested by the referents, and then to assess whether such features can provide informative analyses of free improvisations also. So our aims were to assess whether improvisers could achieve the requested referents, which in turn depended on our establishing satisfactory methods for measuring segmentation.

2. Methods: Participants, procedures, analytical methods

2.1 Participants and procedures

Our participants were nine keyboard players (six male) experienced in improvisation at a professional level, based in and around Sydney. The first author, himself a professional improviser, was not a participant. Performers included practicing jazz musicians, composer/improvisers, and music therapists. They were primarily a subset of Western-trained and -oriented musicians, though several had considerable cross-cultural music making experience. At the time of their sessions our participants ranged in age from early twenties to early seventies. They used a Yamaha Disklavier grand piano (C5), and its MIDI and audio data were recorded using custom MAX patches, which also acquired video recordings from three angles (data not used here). We took account not only of MIDI note information, but also MIDI-pedalling information, to consider sustaining sounds. Participants were in a performance studio, and were asked to play only on the keys, to pedal only with the right foot, to clap at the start and end of each item (as a temporal alignment check for the analyses), and to move as little as possible. The requirement of playing only on the keys was imposed so that all performance data would be accessible as MIDI information. The movement and pedaling restrictions were imposed because we simultaneously measured skin conductance and leg movement by means of sensors on the left ankle. These data are not discussed here, and will be the subject of future publications.

The instructions to the participants were given verbally, by the experimenter, between the items. Each item was requested to be about 3 min long. Table 1 shows the range of performances we requested, and the descriptive words we used for the referent-based improvisations. The first and last item was always a free improvisation: no requirement was imposed other than the recommended duration. In addition, each performer undertook eight improvisations which were intended to fulfill specific referents, as described in the table. Each player performed one item from each of the eight referent rows in the table, and these were statistically distributed amongst them, so that amongst our nine performers, we obtained three renderings of every referent. For each referent type in the table (each row, for example concerning sparsity, pulsedness, tonality etc), there were three possible segmental forms. Two required ABA structures: one where B involved a change which was specified to the performer and related to the core parameter of sparseness, pulsedness etc, and one where a change was required but its basis was not specified, and at the same time, the core referent was required to be unchanged. The third did not require (nor exclude) a change, but required the core parameter to be unchanged in character throughout (e.g. sparse throughout). There was thus a clear gradient of referent precision, between those (items #2, 5, 8, 11, 14, 17, 20, 23) where the core referent and the basis of change to form ABA are specified; those where the core but not the basis of change to form ABA are specified (items $n+1$); those where the core but not whether there is to be change (items $n+2$) are specified; and the free improvisations (where

Table 1. Improviser Tasks.

Core Referent Character (item number and coding)	Change required: basis of change specified	Change required: basis of change not specified	Change not required
Free improvisation (always done first and last : item #1; coding fru)			
Sparseness	Sparse-> dense-> sparse (#2, spt)	Sparse throughout (#3, spu)	Sparse throughout (#4, spn)
Denseness	Dense-> sparse-> dense (#5, dnt)	Dense throughout (#6, dnu)	Dense throughout (#7, dnn)
Single hand melody	Single hand throughout, Low register -> high-> low (#8, oht)	Single hand throughout, no change of register (#9, ohu)	Single hand, no change of register (#10, ohn)
Pulsedness	Pulsed -> unpulsed-> pulsed (#11, put)	Pulsed throughout (#12, puu)	Pulsed throughout (#13, pun)
Quietness	Quiet-> loud-> quiet (#14, qut)	Quiet throughout (#15, quu)	Quiet throughout (#16, qun)
Staccato	Staccato -> sustain -> staccato (#17, tnt)	Staccato throughout (#18, tnu)	Staccato throughout (#19, tnn)
Tonality (loose sense)	Tonal -> atonal -> tonal (#20, att)	Tonal throughout (#21, atu)	Tonal throughout (#22, atn)
Textural (rather than pitch motive based)	Textural -> pitch based -> textural (#23, txt)	Textural throughout (#24, txu)	Textural throughout (#25, txn)
Free improvisation (Always done last and first : #26, fru)			

nothing is specified). The referents were chosen to reflect the main range of parameters that we anticipated solo improvisers using routinely (as discussed in earlier work: Dean, 1989), and to include some redundancy (that is, repetition of a parameter in different task contexts) to assess the reliability of their ability to use the feature.

The fulfillment of a referent is thus a test of our hypothesis. But additionally, a key aspect of this experimental design was to permit potential falsifiability in relation to the hypothesis. If our analyses of segmentation of a parameter (e.g. sparseness) were mistaken, or if performers were merely making arbitrary or randomised changes in relation to referents rather than those requested, then the referent types (n + 1 and n + 2) would be revealing. In these cases, falsification of the hypothesis would occur if segmentation even in relation to the supposedly conserved parameter was still observed. Some pairs of referents, such as 2 and 5 in Table 1 can be construed as ABA versus BAB, and this sought to establish durability of performance even when the context changed. Our professional improvisers expressed no difficulty or confusion in response to the referent characters we proposed. In the case of the tonal–atonal–tonal we indicated that we expected loose interpretation of these terms; and in the case of textural-pitch-textural we merely indicated a desire for a distinction between a texture of notes and a discernible melodic component.

The instructions to participants, our professional improvisers, were purposely couched in the qualitative musical language with which they are familiar. We describe below how we associate a quantitative measure with each qualitative referent parameter; and note that there is inevitably some arbitrariness involved in this.

Performances done by each performer involve one from each row, so including two free improvisations (the first and last items); thus 10 pieces in all. We cycled through the choices in each row with successive participants as follows: Participant A undertook 2, 6, 10, 11, 15, 19, 20, 24; Participant B undertook 3, 7, 8, 12, 16, 17, 21, 25; etc. Participants' data were anonymised by coding them as P1–P9, while three letter codes xxx were used to define the item (1–26 from Table 1) being performed. The first two letters represent the primary feature, e.g. sp for sparse, fr for free. The third represents the required musical change type: t, change required and specified; u, required but unspecified while main referent feature retained; and n, no change required but main referent feature retained. Thus the overall task codes are: 1 fru; 2 spt; 3 spu; 4 spn; 5 dnt; 6 dnu; 7 dnn; 8 oht; 9 ohu; 10 ohn; 11 put; 12 puu; 13 pun; 14 qut; 15 quu; 16 qun; 17 tnt; 18 tnu; 19 tnn; 20 att; 21 atu; 22 atn; 23 txt; 24 txu; 25 txn; 26 fru. Performer 8 undertaking referent 17 would thus appear as P8tnt.

2.2 Analytical methods

We hypothesized that performers would be able to fulfill the referents, and hence that these referents suggest suitable parameters for analysis. The parameters discussed here are our quantifiable interpretation of the referent components, and a test of their fulfillment. Thus a group of musical feature algorithms was developed in R for the purpose of identifying segmentation based on the core referent characteristics requested in most of the items of Table 1, initially using solely the MIDI data acquired during performance. The algorithms are summarized here; some represent procedures that have been applied previously to the analysis of popular and of some classical music, available in software such as the MIDI Toolbox (Eerola & Toiviainen, 2004). However, previous applications normally involve computing an overall parameter representing the feature over a segment or piece. In contrast, most of our analyses involve large numbers of measures of quite short windows rolling across the piece, where means over a chosen window length of the performed time series describe the potentially changing parameter, and generate new time series, which can reveal transitions. Only in the case of the pitch range study was simple inspection (or measurement) of the series of individual data points (pitches) sufficient to define abrupt transition points; in all other cases, the windowing approaching determines that the performed transition will be analysed as a relatively smooth one. In most cases our window-lengths were based on a chosen number of successive notes (or events, see below). In a few cases, such as analysis of 'event (temporal) density', that is how many events there are per unit time, the analysis window must by definition be based on time. In all cases, we have represented the results with respect to the time axis, so that they can be understood in relation to musical flow, and analyses of a given performance can be compared with each other.

The initial bases of the musical feature algorithms used simple quantifiable hypotheses as to how a performer might fulfill each referent, as follows. Sparseness/denseness: temporal frequency of note-ons (windowed note-on density) or windowed inter-onset interval between events, where chords are treated as single events, and a chord is defined as a group of notes whose MIDI IOI is less than 30 ms (see below); low-high register: MIDI note number (pitch range with or without windowing); pulsedness: the degree to which events could be ascribed to durations with simple integer relationships; quietness: MIDI-on key velocity; staccato: note-duration. For some texture vs melody analysis, the proportion of chordal vs single note events, and the average number of pitches per event were both found useful. For tonality and free improvisation analyses see Section 3.

The window sizes chosen for the measures were based on a few precepts, and some empirical testing. The hop size was normally one note or event. We noted that most chords had no more than six notes (only about 3% of chords had more across the total data set). Especially when creating event series, by distinguishing chords from individual notes, this immediately implied that windows of ≥ 30 notes would be required to ensure obtaining 5–10 event phrases. Window size was also informed by the fact that we requested three segments in c. 3 minute performances, and hence modest length segments would be required within a 1 minute section to show significant variation. Beyond this, we assessed the variance obtained across the constructed time series for windows between 30 and 100 events, generally finding 80 events useful. For some analyses where the window had to be constructed on the basis of time, for example in measuring event density, we chose 10 s windows, to correspond to one or several plausible musical phrase lengths, and for these the hop size was one second. For analysis of pitch entropy or pitch class entropy, the R package Entropy (Hausser & Strimmer, 2009) was used, estimating Shannon's H by maximum likelihood. For an alphabet size z, and associated cell probabilities $p_1 \ldots p_z$ this is given by $H = \sum p(\log(p))$.

Once we had constructed a time series representing the extracted feature, designed to relate closely to the core referent character, we used a principled changepoint analysis to detect whether statistically significant segmentation was achieved. For this we used the R Package 'Changepoint (v0.6)' (Killick & Eckley, 2012). At the core, testing for a changepoint involves establishing whether the maximum log likelihood is achieve by a model of the sequential data with or without any particular changepoint. We used the Segment Neighbour search algorithm, which searches the entire segmentation space for significant changepoints, and we required it to search for a maximum of two internal segmentation points, with the discriminating penalty (creating the probability criterion) being the Bayesian Information Criterion. The default method assumes a normal distribution of data, and applies a default penalty to provide a stringent criterion. Thus the method can support or contradict the hypothesis that there are segmentation points. In most cases, the results were unambiguous. For confirmation of the results of the Segment Neighbour search, when desired, we used a Binary Segmentation method which does not require an assumption as to the distribution of the data, the non-parametric cumulative sum of squares method. In this case, an Asymptotic Penalty was used, where the chosen value (≤0.05) is the theoretical Type 1 error. All presented segmentations are significant at this probability level or better. Segmentations were primarily made in relation to changes in the windowed means of the measured parameters, but variance was also considered in a few cases (each noted below), for example when inspection of the time series suggested variance change without change in mean, or where s.d. parameters consistently varied more than mean. The Changepoint package allows assessing segmentation on the basis of mean or variance alone, or taking them jointly. Thus Changepoint analysis provides a quantitative principled tool to test the hypothesis that a referent-related segmentation is achieved.

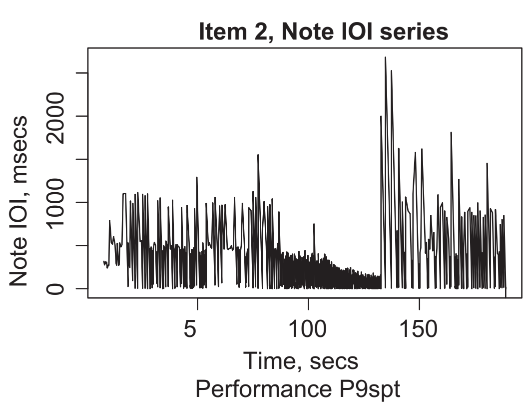

Fig. 1. A raw Inter-onset interval time series for the Sparse–Dense–Sparse referent (Item 2).

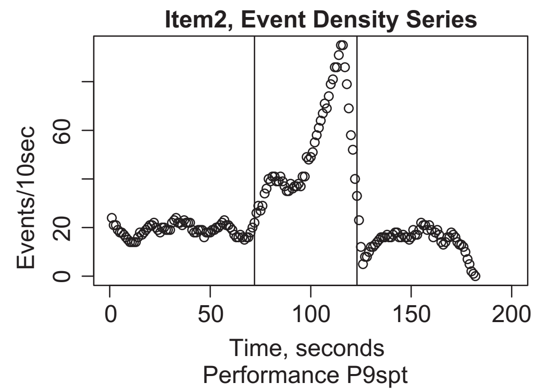

Fig. 2. The event-density time series for the Sparse–Dense–Sparse referent (Item 2).

Representative analyses of individual performances are shown in the figures. Vertical lines are added to the figures to define the position of the segmentation points detected by the rigorous Changepoint procedure. The tables summarise the results of the analyses of all the performances.

3. Results and discussion

3.1 Computational segmentation of the referent improvisations, based on quantitative musical features

3.1.1 An example: The simple case of sparse-dense-sparse (item #2)

This analysis is based on event density (frequency of chord or melodic single note-ons with respect to time). First, Figure 1 illustrates the time series of raw IOI (inter-onset interval) values for all sounded notes (without discriminating chords into events) in one performance of Item 2. The ABA Sparse-Dense-Sparse form is already apparent. The distribution of IOIs and their estimated probability density function showed that 30 ms was a reasonable cut off for associating notes with an individual chord, in agreement with previous work. For example Gabrielsson (1973) and others have shown that 50 ms IOIs are ample to confirm that the notes are perceived as separate rather than belonging to a chord, while Pressing (1987) arrived at 30 ms for defining chords in analyses of his own keyboard improvisations, exactly as we did). Thus Figure 2 shows the time series of Event IOI values, where notes that belong to a chord constitute single events, as do single melodic notes. The three part segmentation, Sparse-Dense-Sparse is now clear. All three participants who undertook this task achieved it clearly, and Table 2 (see later) summarises the data representing all performances of each referent.

3.1.2 Methods and their refinement for segmenting each core referent: Focusing on the ABA core-change referents

This section describes the range of methods developed to achieve segmentation for each referent, and thereby demonstrate whether or not the improvisers were able to fulfill them. As discussed already, the Changepoint analysis also provides repeated opportunity to provide evidence inconsistent with our hypothesis. The method outlined in Section 3.1.1 for referents focused on sparse-dense contrast (Items 2–4), was immediately applicable to Items 5–7 focussed on dense-sparse contrasts, and performers again fulfilled the referent.

Figures 3(a) and (b) show the results of analyses of referents from the Items 8–10 group, concerned with pitch range/singled handed performance. Vertical lines represent the computationally defined segmentation point and it can be seen that the pitch sequence itself was enough to decipher the fulfillment of the low-high-low one handed referent (Figure 3(a)) and the referent in which pitch range was not required to change (Figure 3(b)). We interpret the referent 'constant register' to mean that the majority of the pitches in each segment occupy the same pitch range, and contrarily for 'changing register'. In Figure 3(a) the high register (middle segment) is completely separated from the low. In Figure 3(b) the registral range changes in the middle section, but the majority of events (>70%) are within the range circumscribed by the outer segments taken together.

The analysis of referents 11–13, focused on 'pulsedness', was more complex. Attempts to segment indirectly, based on patterns of accentuation (MIDI-on velocity) rather than timing features, were unsuccessful. Consequently, we focused on more direct, timing based measures. Using the event time series (in other words, amalgamating chord-notes into single events as discussed above), we measured duration ratios between successive events, given that tempo varies locally in performances (review: Collier & Collier, 1996). As might be expected, the vast majority of ratios were around 1–2 (events were of similar durations); there was usually a subsidiary peak in the histogram around 8–10: thus we considered that windows of time in which the mean ratio moved away from a small number and became more variable might characterise the unpulsed segments. We term the standard deviation of the IOIRatio (windowed over 30 successive events) 'pulsedness', where low values indicate strong pulse character, and high values indicate unpulsed character. This parameter was quite discriminating (Figure 4) and it defined 'unpulsed' regions which were apparent to expert listeners (the authors). Refinements to this method could use a longer window to estimate local tempo, but there would be serious trade-offs in such approaches, and they were not necessary. It was apparent that IOIRatios were much more homogeneous in the performances of the referents which were required to be pulsed throughout than those requiring pulsed-unpulsed-pulsed changes.

Items 14–16 required a focus on Quietness, with a Quiet-Loud-Quiet referent for 14. In most cases the analysis could be done simply and very clearly using mean MIDI-on velocity for all notes (without a need to isolate chord-notes into events), windowed over 50 notes. There was one exception, where the windowed mean did not show the expected pattern. The only alternative method by which a player might achieve the required dynamic contrast would be by controlling the number of notes sounding at any instant: for a given average MIDI-on velocity, increasing 'polyphony' (for example more chords instead of single notes) would increase the loudness in a monotonic (but not necessarily linear) way. We call this aspect of polyphony, which is simply the number of notes sounding at any moment, chordality. This term and 'chordalness' have specific meanings within graph theory that we do not imply here. We assessed such polyphony first by measuring the chordality at each note entry and the corresponding mean MIDI velocity, multiplying the two for simplicity and then averaging across 10 s windows (1 s hop). We do not claim that this multiplication is a direct measure of produced acoustic intensity, but that the latter rises monotonically with this product, and we also confirmed this approach by direct measurements of the audio amplitude. The case in question is shown in Figure 5.

Staccato playing was the core referent for Items 17–19, with 17 requiring a Staccato-Sustain-Staccato ABA form. We considered that the impact of a staccato chord on the computational evaluation of the degree of staccato quality should be the same as that of an individual staccato note, and thus we used the event series rather than the note series (as above). Staccato literally refers to 'detaching' notes from each other; with notated music, any durational value may be played either legato or staccato, and it is the separation between notes/events which is most diagnostic (for example, gap as proportion of IOI). In improvisation, we assumed that instead staccato playing would be focused on very short notes. This is because it is difficult to distinguish a note of three slow beats in length followed by a rest of one beat from a staccato realization of a note whose notated duration is four beats. Our assumption was correct, in that windows of 50 events allowed a time series of mean event duration which revealed patterns clearly, and

Table 2. Summary of Computational Segmentation of Referent-based and Free improvisations.

Referent (detailed in Table 1)	Number of Participants fulfilling the parameter in focus/ Number undertaking Referent. Note that in 2,5,8,11,14,17,20,23 this required contrast in the parameter. In others it required constancy.	Parameters showing largest Coefficient of Variation in the normalized time series (number of performances showing this feature)	Comments
1. Free Improvisation	NA	Chordality(5), tonalness(2), eventdensity(1), pitchrange(1)	Performed First
2. Sparse-Dense-Sparse	3/3	Eventdensity(2), pitchrange(1)	
3. Sparse throughout (some other change required)	3/3	Tonalness(1), chordality(1), pitchrange(1)	
4. Sparse throughout (other change permitted)	2/3	Chordality(2), pulsedness(1)	One performance became more dense in a middle segment contrary to the referent
5. Dense-Sparse-Dense	3/3	Eventdensity(3)	
6. Dense throughout (some other change required)	3/3	Eventdensity(2), Chordality(1)	
7. Dense throughout (other change permitted)	2/3	Tonalness(2), pitchrange(1)	
8.Low-High-Low pitch (change hands)	3/3	Chordality(3)	
9. Single hand no pitch register change (some other change required)	3/3	Event density (2); chordality (1)	
10. Single hand no pitch register change (other change permitted)	3/3	Chordality(2), eventdensity(1)	One person expanded the pitch range, though still performing with one hand
11. Pulsed-Unpulsed-Pulsed	3/3	Chordality(2), tonalness(1)	Note that the pulsedness parameter itself was not amongst the highest cvs. See text for discussion.
12. Pulsed throughout (some other change required)	3/3	Tonalness(1), chordality(1), eventdensity(1)	
13. Pulse throughout (other change permitted)	3/3	Chordality(2), eventdensity(1)	
14. Quiet-Loud-Quiet	3/3	Chordality(2), eventdensity(1)	2 People achieved this by varying average MIDI velocity of note onsets; 1 by varying density of note events per unit time. Loudness cvs were only slightly smaller than those for the reported parameters.
15. Quiet throughout(make some other change)	3/3	Chordality(2), eventdensity(1)	
16. Quiet throughout (other change permitted)	3/3	Chordality(2), MIDI-on velocity (1)	
17. Staccato-Sustain-Staccato	3/3	Chordality(2), eventdens(1)	
18. Staccato throughout (make some other change)	2/3	Chordality(2), pitchrange(1)	There was one deviation from the staccato fixity.
19. Staccato throughout (other change permitted)	3/3	Chordality(2), eventdensity(1)	
20. Tonal-Atonal-Tonal	3/3	Tonalness(1), chordality(1), eventdensity(1)	
21. Tonal throughout (some other change required)	3/3	Tonalness (2), chordality(1)	

(*Continued*)

Table 2. (*Continued*).

Referent (detailed in Table 1)	Number of Participants fulfilling the parameter in focus/ Number undertaking Referent. Note that in 2,5,8,11,14,17,20,23 this required contrast in the parameter. In others it required constancy.	Parameters showing largest Coefficient of Variation in the normalized time series (number of performances showing this feature)	Comments
22. Tonal throughout (other change permitted)	3/3	Chordality(2), MIDI-on velocity (1)	
23. Textural-Pitchbased-Textural	3/3	Chordality(2), event density(1)	Two People performed this by moving from a narrow to a wide pitch range; one by moving from chordal to more single line melody construction.
24. Textural throughout (make some other change)	2/3	Chordality(1), eventdensity(1), pitchrange(1)	One deviation from the textural fixity
25. Textural throughout (no other change)	3/3	Eventdensity(2) pulsedness (1)	
26. Free Improvisation	NA	Eventdensity(4), chordality(2), MIDI-on velocity(2), tonalness (1).	Performed last
Referent (detailed in Table 1)	Number of Participants Successful in the parameter in focus/ Number undertaking Referent	Parameters showing largest Coefficient of Variation in the normalized data series	Comments

confirmed that referent 17 was achieved this way by each participant who undertook it. Mean durations for the staccato sections were 80 ms, varying slightly between performers, while those for the sustained sections varied far more on the basis of the musical features being used, and were mainly in the range 500–1500 ms.

The issue of 'tonal-atonal-tonal' transitions, which we raised for Items 20–22 is a more complex one. We gave no definitions of these terms, and as professional musicians, our performers were not challenged by them (as judged by questioning after the complete experiment set), nor did they ask any questions about them when the task was presented. We hypothesized that improvisers would have a pragmatic approach to the real-time application of ideas of controlling 'tonalness' (Temperley, 2007): that they would consider atonal features (which we construe more loosely as close to non-tonal) to be represented primarily by simultaneous and immediately successive intervals of minor second, augmented fourth, and major seventh (and of course their equivalences such as minor ninths); while the remaining intervals would represent the tonal. We also argued that all the intervals within a chord need to be considered in this context, and that the performers would do this. Given that even with these particular referents there were few chords with more than six notes (<10% of all chords in the renderings of Item 20, for example, as well as <3% across all pieces), we measured pitch intervals between each sounded note and the succeeding five, and expressed these as an 'atonal:tonal ratio'. This ratio is measured as the number of 1, 6, 11 semitone intervals when the data are considered as pitch class, divided by the number of occurrences of the other intervals. Including all the notes (chords and isolated notes), we assessed the window (total number of notes) that provided the best resolution as that with maximum coefficient of variation of the ratio, and greater than 40 notes (to accommodate a possible succession of six six-note chords at the minimum) and arrived at a window of 80 successive notes as suitable for the analysis. While this approach is well supported by discussions of pitch class usage and perception (Krumhansl, 1990) we also used a measure of Shannon's pitch class entropy in R. This gave results consistent with our simpler analysis, but with slightly lower resolution (as judged by coefficient of variation and the resulting segmentations), and is not shown.

The results support our hypothesis of pragmatic behaviour, and show that improvisers could mostly fulfill this referent request. Figure 6 gives an illustration of a slightly ambiguous response to the Tonal-Atonal-Tonal referent (20). While the central section is more atonal than either surrounding, it is not fully delineated by our standard Changepoint analysis, which only identifies segmentation c. 10 and 80 s, because of the differences in mean between the segments either side of it. It is possible that the performer did conceive a 10 s tonal section followed by a 70 s atonal one, but it is perhaps more likely that the intention was represented by the segmentation at 40 s (added by repeating our standard analysis and criteria over the range 0 to 80 s). In either case, this performance fulfills the

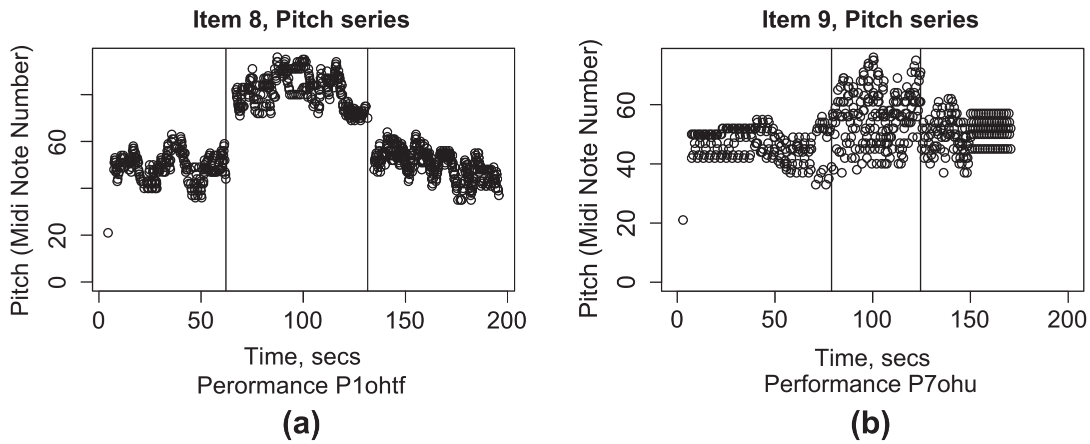

Fig. 3. (a) Referent 8, Required Register Change. (b) Referent 9, Register constant, some other change required.

referent (which made no specification of the relative levels of tonality in 1 and 3). The other two performances are more clear-cut instantiations, having similarly tonal segments 1 and 3, allowing the atonal segment to stand our more clearly in statistical terms, with higher peak atonal: tonal ratios (0.5–0.6). It seems likely that the relatively simple but practical and achievable way of characterizing tonality vs atonality we hypothesised is relevant to our performers.

'Textural' playing was the core referent for Items 23–25, and was contrasted with 'pitched melody' playing. It was clear from the raw pitch contours of several performances that segments were delineated on the basis of pitch range. This in itself would not fulfill the referent, but it suggested that considering aspects of pitch would be a suitable starting point. Bearing in mind that performers were not allowed to produce texture by playing directly on the strings, and that issues of tonality were not a required feature, we hypothesized that textural playing would be achieved by the more chordal, higher pitch entropy sections and pitched melodies (largely, single-note passages) by lower values of these parameters. In other words, for texture, there would be more notes sounding at once (in chords; pedaling would also influence this), and there would be wider ranges in both pitch range and class, than for melodies. So we assessed both the degree of chordality and the (raw) pitch entropy as windowed parameters across the performances. As above, we use the term chordality to refer to the degree of chordal character. Figure 7 shows a somewhat ambiguous example of a Texture-Melody-Texture, where alternative interpretations of the segments are apparent, but the referent was fulfilled.

3.1.3 Performances for which an unchanging core referent feature was required

As noted above, all performances of the primary ABA referent for each core feature fulfilled the referent

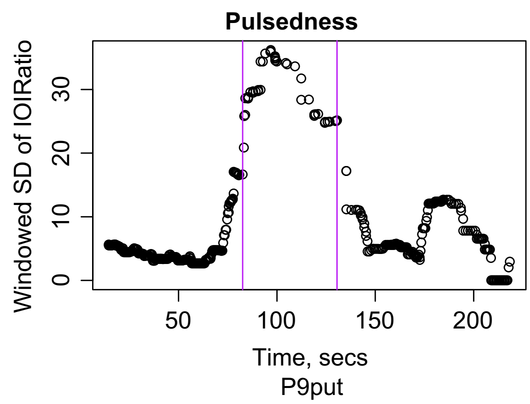

Fig. 4. Referent 11, Pulsed–Unpulsed–Pulsed.

requirement. For referents 3, 4, 6, 7, 9, 10 …, one of the tasks of the improviser was to retain relative homogeneity in the core referent character (for example sparsity in 3, density in 6, etc.). Thus we next tested explicitly whether this was achieved, as judged by the methods already described for each referent group; this constituted a key opportunity for falsification. Assessments were based on the changepoint segmentation analysis developed already, taken together with visual assessment of the plots; now we tested for a breach of the referent, i.e. whether there was a changepoint in a parameter for which the specification required there to be no change. As shown in Table 2, there were only three out of a possible 16 cases in which there was a clearcut breach of the call for constancy; there was only one other failure to fulfill the referent. Figure 3 shows an intermediate performance of Item 9, requiring continuous playing in a single register, and requiring a separate change: while there was no abrupt or complete change of register (so the required constancy of the core-referent was achieved), there was a segment in which the range of pitches contained in the performed register was expanded. In the cases of complete failure to retain the core referent,

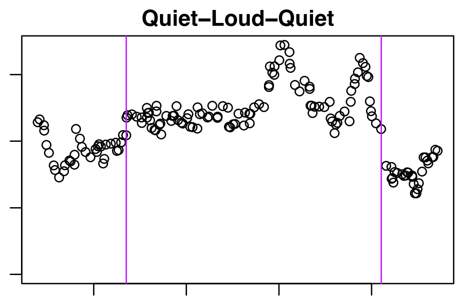

Fig. 5. Quiet–Loud–Quiet created by combination of chordality and MIDI-on velocity.

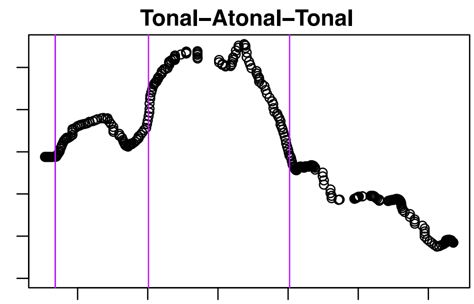

Fig. 6. A performance of Item 20, Tonal–Atonal–Tonal, with tonal segments distinct.

it may be that uncertainty as to whether to make changes, and/or in what feature, reduced performer attention to the demand of homogeneity of the core referent feature. This possibility is supported by later data in this paper suggesting that improvisers have a strong tendency to produce segmentation when improvising freely. The question of the impact of uncertainty in real-time planning of an improvisation deserves focussed investigation.

We do not detail examples of analysis of the features changed in 3, 4, 6, 7… for reasons of space, and because they will be addressed in future studies correlating them with skin conductance changes. Rather, in the next section, we address the question of uncertainty of the extent and nature of change, and how to assess it through our analyses of free improvisations.

3.1.4 Free improvisations: neither a core nor a change referent is specified

Given that the data so far show that our methods are suitable for detection of segmentation, and that our improvisers were capable of fulfilling the referents and did so quite consistently, we proceeded to analyse the free improvisations. Our purpose here was to assess the hypothesis that segmentation is common in free improvisations. Free improvisation Item 1, prior to any referent instructions, was always the first item, and so we considered that if these improvisations were generally segmented this would support the hypothesis. In contrast, Free Improvisation Item 26 might well be conditioned by the demands of the intervening referents. We considered that the range of analyses developed for Items 2–25 would encompass many of the likely ways of segmenting the free improvisations, given that they were solo, and that all sounds had to be made by playing the keys.

We did not aim to determine every possible segmentation point, but rather to test, using our standard Changepoint criterion, whether the free improvisations contained at least one. Thus we analysed each free improvisation with a logical sequence of the individual analyses, taking the occurrence of at least two segments revealed by an analysis as evidence in agreement with our hypothesis. Given a positive outcome of this kind, we did not necessarily continue with the remaining pre-planned sequence (this may be useful in future work). Falsification of our hypothesis was again feasible, in that some free improvisations could lack segmentation.

We arrived at the sequence of analyses to undertake by normalizing every parameter of the performances (using all the methods developed) so that they spanned the range 0–1 in each case. This then permits useful comparison of the coefficients of variation (CV = s.d./mean, a measure of variability) of the different parameters for an individual piece. Unless this is done, the absolute value of the mean grossly perturbs the CV for a given s.d., rendering comparisons between different parameters uninterpretable. The results of this approach are shown in Table 2 (column 3), and reveal that the parameters which most commonly have the highest CV after normalization for an individual piece are chordality, tonalness, and event density. During the chordality measurement we also made windowed measures of the distinct parameter chord:note frequency ratio. This frequency ratio is that between the number of chords and the number of individual (separated) notes in a passage, as distinct from chordality, a measure of the average number of notes sounding per event. Far less common amongst parameters with the highest CV are pitch range, pulsedness, and MIDI-on velocity. Of these results, perhaps only the low ranking of MIDI-on velocity would not have been anticipated: but this is explicable as discussed above for

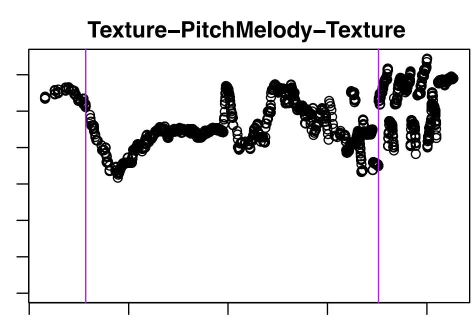

Fig. 7. The Texture–Pitch Melody–Texture Referent (Item 23): an assessment by pitch entropy.

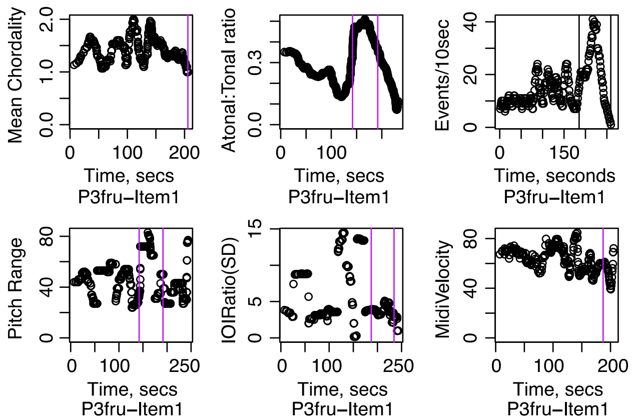

Fig. 8. Multiple approaches to segmenting a Free Improvisation from the Item 1 set.

the Quiet-Loud-Quiet referents, by performers often using chordality more extensively than MIDI-velocity when controlling loudness. Our sequence of analyses was thus set in the order just given. Because there was no referent here, there was no way to anticipate whether what might be controlled to produce segmentation would be reflected in changes in mean values or parameter variance, or a combination. Thus while retaining the standard changepoint criterion we allowed the application of the methods testing mean, variance, or the two in conjunction. When anything other than the standard mean method is used, it is noted below.

All our free improvisations, both from Item 1 and Item 26, were unambiguously segmented (nine performances, one from each player, for each free improvisation). To illustrate this, we show the complete sequence of analyses, each with the detected changepoints, on one performance from the Item1 set (Figure 8), and the first successfully segmenting analysis in the sequence for another Item 1 and for an Item 26 (Figures 9 and 10), by different players (one who is primarily a concert improviser, one who is a music therapist). We did not assess rigorously whether there was more segmentation for 26 than 1, but this was not obviously the case. Figure 8 shows interesting congruence between segments detected by Tonalness and PitchRange, indicating quite clearly how this segment was improvised. Similarly, there is good concordance between the EventDensity, NoteOn velocity and Pulsedness segmentation. Chordality was uninformative, but chord: note ratio, measured at the same time, was useful (see Figure 9 for the latter).

Performer 8 showed several segmentation features, of which the first detected in our analytical sequence using the basic change point mean analysis, chord: note ratio, is shown in Figure 9.

Similarly, for Item 26, Figure 10 shows that besides other segmenting features not shown, EventDensity defined a coda section in the performance of Participant 9.

Parameters studied for the information in column 3, with windowing, were event-density, chordality (which here also includes chord : note frequency), pitch-range, pulsedness, MIDI-on velocity, tonalness, pitch class entropy, and raw pitch entropy. Given that only 4 out of a total of 72 referent tasks were not achieved (as judged computationally), a chi-square analysis indicates $p < 0.0001$ for this to occur by chance. Similarly, $p < 0.005$ for segmentation to occur by chance in the free improvisations (3 performances lack segmentation, 15 include it).

3.2 General discussion and summarised conclusions

Our overall purpose was to test the hypothesis that our participants could fulfill referent instructions for improvisations, that our analytical methods were suitable to determine this, and that free improvisations are also commonly segmented in related ways. The data presented show clearly that professional improvisers can process high level musical resources with enough statistical consistency to develop computationally identifiable musical segmentation over periods of the order of a minute per segment, complementing earlier data on segmentation over periods around 5 s from a self-analysis study (Pressing, 1987). They also show

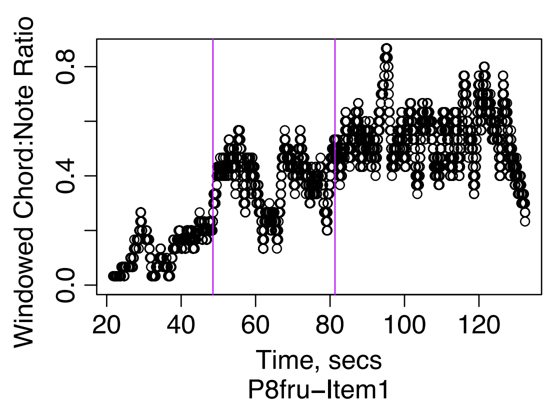

Fig. 9. Free Improvisation Item 1: by Performer 8.

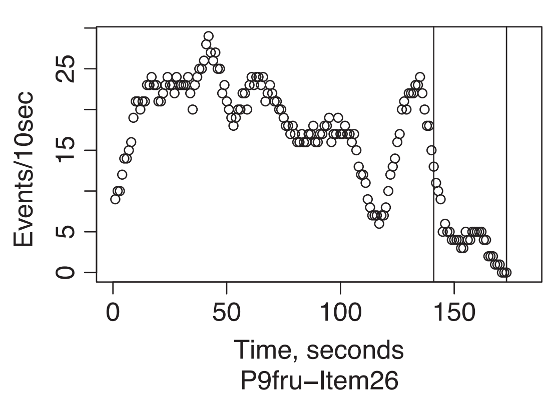

Fig. 10. Free Improvisation Item 26: by Performer 9.

that our methods are suitable to detect this. The detected segments were all heard as entirely plausible from a listener's perspective by three musicians not involved in the performances: the three authors. By this we mean that we consider that the segments were either produced by design by the performers, or would be recognized by them as constituting segments if asked. We are not documenting our (potentially biased) perceptions of this issue in depth here, because in further studies we have worked on the comparisons between the perceptions of musical structure by the performers themselves (listening back to some of their own work from the experimental session) and by uninvolved non-musicians. These latter data, being unbiased by knowledge of the statistical results, nor additionally the purposes (respectively), are more informative, and remain generally supportive of the interpretation that the statistically detected segments are related to perceptually impacting changes.

So professional improvisers have the ability to generate segmentation in high-level musical structure. A corollary of our hypothesis is that they would be prone to do this whatever the context and without prompting. This might be because of a propensity to form structure so as to create musical argument, agency or meta-narrative (see, e.g. Lewis, 1996; Maus, 1997): this requires the juxtaposition of distinguishable structures, i.e. segments. For example, the entry and exit of a soloist in a concerto with orchestra may provide segments which differ in perceived agency; while our improvisers may introduce such contrasts through structural segmentation. Our results show correspondingly that in the first (free) improvisation each participant undertook, segmentation was indeed detectable, in support of the hypothesis. The relative similarity of the definition of segments in both the first and last free improvisation is also supportive of the view that this process is normal in improvising contexts. Future analyses from our duo improvisers' performances will also address this question.

It is interesting to consider the range of musical parameters that the improvisers seem to have used to achieve segmentation in those segments for which this was not specified. Table 2 shows the results of our analysis of variability in the eight major parameters we analysed, and thus for items 3, 4; 6, 7; etc., as well as for the free improvisations, reveals aspects of their choices. Chordality and event density were important in the referent pieces at large, and in both free improvisations, and it is notable that so was tonalness: thus an emphasis on control of tonality, even if loosely conceived by the improvisers from the stance of music-theory, was important even before they had performed referents which specified its involvement, and no more so afterwards. A simple interpretation of the importance of chordality and event density is that they are means towards control of overall density and acoustic intensity, while correspondingly the note velocities are somewhat less influential. More broadly, it is apparent that the microstructural segmenting features such as introduced by Pressing, are also at work here at the macrostructural level, in accord with the referents, but also in the free improvisations. Our evidence thus supports the view proposed initially that microstructure is used to create macrostructure.

It should be freely admitted that the choice to emphasise the parameters with the highest CVs after data normalization is a somewhat arbitrary one, though successful. For the future, possibilities exist to optimize the free improvisation segmentations statistically using multivariate approaches, and to compare the resultant segmentations with those derived from individual parameters (as illustrated in Figure 8). For example, each parameter could be standardized (to have mean 1, and standard deviation 1), and a multivariate Mahalanobis distance computed between each event and the performance mean. This distance series could then be the basis for Changepoint segmentation. Future work will assess the utility of such approaches in the context of the data we have relating skin conductance and leadership changes to the segmentation.

Such a multivariate approach brings to mind some of the now conventional MIR approaches to audio segmentation, using barrages of low level features measured automatically, be it with Sonic Visualiser, or Echonest, as mentioned above. It would not be plausible that an improviser could consciously operate on some of these features, nor that they could operate simultaneously on a large number of them; though they might still have psychological impact. But the MIR approaches, as discussed above, are particularly pertinent to compositions, such as those of popular music. When applied to electroacoustic sound, such approaches are again the main ones available at present, and so there is a need for sparse sampling to provide an acoustic gesture alphabet which can be analysed from an information theoretic viewpoint, so as to relate the acoustic processes more closely to the note-based event series of keyboard and classical music (cf. Bloch, Dubnov, & Assayag, 2008; Pearce et al., 2010). Studies on popular music structure identification are also providing leads towards this reconciliation, which might also permit an assessment of how many distinct musical features an improviser can simultaneously control over extended segments, such as those of the macrostructures we define here.

Acknowledgements

The participation of our improvisers is greatly appreciated. Naturally, their data are not individualized above, but most were happy to be listed, and many of their recordings and performances, and other aspects of their work, are accessible internationally. They included: Diana Blom, Leonie Cohen, the late Robin Howat, Kevin Hunt, Mark Isaacs, Dave MacRae, Matt McMahon, and Alister Spence. We also thank Dr. Mary Broughton for assistance in developing the methods for the experimental study, and Dr Rebecca

Killick (University of Lancaster) for advice on Changepoint and other helpful comments. Dr David Temperley kindly suggested the use of pitch class entropy in the context of gradients of tonalness.

References

Bailes, F., & Dean, R. T. (2007a). Facilitation and coherence between the dynamic and retrospective perception of segmentation in computer-generated music. *Emprical Musicology Review, 2*(3), 74–80.

Bailes, F., & Dean, R. T. (2007b). Listener detection of segmentation in computer-generated sonic texture: An exploratory experimental study. *Journal of New Music Research, 36*(2), 83–93.

Bailey, D. (1980). *Improvisation: Its nature and practice in music*. Ashbourne: Moorland (republished Prentice Hall; and revised edition, British Library). This book also formed the basis of a series of British Broadcasting Corporation films (1991), directed by J. Marre.

Bloch, G., Dubnov, S., & Assayag, G. (2008, August). *Introducing video features and spectral descriptors in the Omax improvisation system*. Paper presented at the International Computer Music Conference, Belfast, N. Ireland.

Bruderer, M.J., McKinney, M.F., & Kohlrausch, A. (2012). Perceptual evaluation of musicological cues for automatic song segmentation. Psychomusicology: Music, Mind and Brain, *22*(1), 3–17.

Clarke, E.F., & Krumhansl, C.L. (1990). Perceiving musical time. Music Perception: An Interdisciplinary Journal, *7*(3), 213–251.

Collier, G. L., & Collier, J. L. (1996). *Microrhythms in jazz : a review of papers* (pp. 117–139). VIII: Annual Review of Jazz Studies.

Dean, R. T. (1989). *Creative improvisation : Jazz, contemporary music and beyond*. Milton Keynes: Open University Press.

Dean, R. T. (1992a). *New Structures in Jazz and Improvised Music since 1960*. Milton Keynes: Open University Press.

Dean, R. T. (Ed.) (1992b). *Eleven Views of Music Improvisation (Vol. 32)*. Sydney: Australian Music Centre.

Deliege, I. (1987). Grouping conditions in listening to music: An approach to Lerdahl & Jackendoff's grouping preference rules. *Music Perception: An Interdisciplinary Journal*, *4*(4), 325–359.

Dibben, N. (1994). The cognitive reality of hierarchic structure in tonal and atonal music. *Music Perception: An Interdisciplinary Journal, 12*(1), 1–25.

Eerola, T., & Toiviainen, P. (2004). MIR in MATLAB: The MIDI toolbox. Proceedings of the International Conference on Music Information Retrieval, Barcelona, Spain, October (pp. 22–27, http://www.ismir.net/). Canada: ISMIR.

Ehmann, A.F., Bay, M., Downie, J.S., Fujinaga, I., & De Roure, D. (2011). *Music structure segmentation algorithm evaluation: Expanding on MIREX 2010 analyses and datasets. Proceedings 12th International Society for Music Information Retrieval Conference (ISMIR 2011), Miami, FL, USA*, *October* (pp. 561–566, http://www.ismir.net/). Canada: ISMIR.

Gabrielsson, A. (1973). Similarity ratings and dimension analyses of auditory rhythm patterns. *I. Scandinavian Journal of Psychology, 14*(2), 138–160.

Hausser, J., & Strimmer, K. (2009). Entropy inference and the James-Stein estimator, with application to nonlinear gene association networks. *The Journal of Machine Learning Research, 10*, 1469–1484.

Jost, E. (1974). *Free Jazz* (English ed.). Graz: Universal.

Killick, R., & Eckley, I.A. (2012). Changepoint: An R package for changepoint analysis. R package version 0.6.

Krumhansl, C.L. (1990). Cognitive Foundations of Musical Pitch USA: Oxford University Press.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.

Lewis, G. (1996). Improvised music since 1950: Afrological and Eurological perspectives. *Black Music Research Journal, 16*(1), 91–122.

Maus, F.E. (1997). Narrative, drama, and emotion in instrumental music. *Journal of Aesthetics and Art Criticism, 55*(3), 293–303.

Narmour, E. (1989). *The analysis and cognition of basic melodic structures*. Chicago: University of Chicago Press.

Pearce, M., Müllensiefen, D., & Wiggins, G. (2010). Melodic grouping in music information retrieval: New methods and applications. In Z.W. Ras & A.A. Wieczorkowska (Eds.), *Advances in music information retrieval* (pp. 364–388, Studies in Computational Intelligence Series, Vol 274). Berlin: Springer.

Pressing, J. (1982). Pitch class set structures in contemporary jazz. *Jazz Forschung, 14*, 133–171.

Pressing, J. (1987). The micro-and macrostructural design of improvised music. *Music Perception: An Interdisciplinary Journal, 5*(2), 133–172.

Pressing, J. (1988). Improvisation: methods and models. In J. Sloboda (Ed.), Generative processes in music. The psychology of performance improvisation and composition (pp. 298–345). Oxford: Clarendon Press.

Smith, H., & Dean, R. T. (1997). *Improvisation, hypermedia and the arts since 1945*. London: Harwood Academic.

Temperley, D. (2007). *Music and probability*. Cambridge, MA: MIT Press.

Xenakis, I. (1970). Towards a metamusic. *Tempo, 93*, 2–19.

Xenakis, I. (1971). *Formalized Music*. Bloomington: Indiana University Press.