Relationships between generated musical structure, performers' physiological arousal and listener perceptions in solo piano improvisation

Journal:	*Journal of New Music Research*
Manuscript ID	NNMR-2015-0067.R1
Manuscript Type:	Research Paper
Keywords:	emotion, expression, performance, Improvisation; segmentation; professional improvisers; skin conductance; movement; affect.

Abstract. What musical structures do improvisers produce, and how do these relate to their physiological arousal while performing and to listener perceptions? Nine professional improvisers performed both structured and free improvisations. We hypothesised that increases in performers' arousal and attention during structural transitions would be reflected by changes in skin conductance. Consistent with the hypothesis, skin conductance changed particularly around transitions. Improvisers then listened to their improvisations, continuously rating musical change. Fourteen non-musicians also rated change, and separately rated perceptions of affect. Their perceptions related to structural parameters, though these were less influenced by musical features than those of the performers.

Key words: Improvisation; segmentation; professional improvisers; skin conductance; movement; affect.

1. Introduction

Studies on cognitive and electrophysiological correlates of musical improvisation are currently accelerating. Most work considers highly simplified tasks, which may not involve many of the complex demands of professional improvised performance, even though in some cases professional improvisers are participants. For example, Large, Palmer and Pollack (1995) asked pianists to improvise variations on given melodies, requiring the simple elaboration of highly structured material. While more recent neuroimaging studies have accorded slightly greater improvisational freedom, their ecological validity is compromised by the need for tight experimental control (e.g. Bengtsson et al., 2007; Donnay et al., 2014; Limb & Braun, 2008). The most sophisticated professional abilities brought to bear in these studies include those of the conventional jazz improviser (Donnay et al., 2014; Keller et al., 2011), used to regular repeating harmonic chord sequences with no or limited modulation; and those of the classical concerto soloist (Berkowitz, 2010) who may to some degree improvise a cadenza. In spite of the limited task demands of these experiments, the neuroscience observations of brain regions of interest are disparate (Beaty, 2015), and as Beaty discusses, this probably results both from methodological differences and analytical deficiencies.

Our interest is in professional solo keyboard improvisers who are faced with complex improvisation task demands such as are routine in free improvisation. By 'free improvisation', we refer to making music with no or minimal pre-specified demands or materials, for example no pre-agreed chord sequence, rhythmic structure, or tonality. In the two free improvisations each performer undertook in our study, the only musical restrictions imposed by the experimenters were the request for no left leg pedalling (explained below), and for no playing directly on the strings or body of the piano (since this information would not be captured by our MIDI-recording). This is not to suggest that free improvisation occurs in a constraint-free vacuum: an improviser is necessarily influenced by background, experience, expertise, and cultural and immediate contexts. On the other hand, it is worth noting that some free improvisers advocate what Derek Bailey called 'non-idiomatic' improvising, meaning active avoidance of familiar musical conventions; and others advocate what we have termed 'non-sensory' improvising, meaning active avoidance of cognising or being influenced by other on-going musical activities. Naturally these are idealised positions, but they are far from those of the mainstream jazz or Indian music improviser, strongly utilising preformed structures and conventions. These issues are elaborated in several books such as Bailey (1992 revised edition), Dean (1992), and Smith and Dean (1997).

For our experiments, professional keyboard improvisers performed a range of improvisations, all without preformed musical motivic material. Throughout, their ultimate purpose was to create pieces of music, but under the improvising constraints we requested. Such constraints are typical even of free improvisation performances, for which the improviser either provides the constraints for themselves, or if playing with others, may also use the constraints provided by them. From a procedural and structural point of view, the range of pieces we requested spans free improvisations, and 'referent' improvisations in which the performer is constrained to create their own version of an ABA structure, where A and B are characterised in relation to a simple core musical feature (such as event density) or a more complex one (such as tonality). Within the set of ABA referent structures we request, some involve transforming A in order to create B (e.g. sparse-dense-sparse, pulsed-unpulsed-pulsed or tonal-atonal-tonal). In others B has to be created without changing A (which continues during the

B section), or it is optional for B to change, yet A must still remain throughout. There is no demand for fixed or regularly repeating harmonic or metrical structure.

In our preceding two studies on improvisers (Dean & Bailes, 2015; Dean, Bailes & Drummond, 2014), we first showed how skin conductance (SC) can be measured during performance providing account is taken of the performer's movements (which can influence SC dramatically). We obtained evidence that musical segmentation during very simple performances and during imagining them can be detected in the SC traces. In the second study, developing from work by Pressing (1987) on the generation of musical micro-structure during improvisation, we showed that our improvisers mostly realised the macro-structures demanded by the task, and could use the musical referent features specified in order to do so. We also showed that music-computational segmentation of the performances of what we will refer to as 'fully-specified' referents could be achieved on the basis of the prescribed changes. By fully-specified, we mean those ABA referents in which the nature of the distinction between A and B was defined by our instructions (e.g. sparse-dense-sparse). ABA structures in which the B section involved a continuation of the A feature, and the construction of contrast and transition by means of a different musical feature, are termed 'partially-specified'. This term also covers the referent tasks in which we left it optional for the performer whether there was a B section (in which A continued), or whether they chose effectively to realise an AAA structure with no clear middle segment. In the partially-specified tasks, as well as in two free improvisations which sandwiched all the others (one free improvisation first, before any of our referents or implied purposes had been mentioned; and one at the end after all the referents had been performed) we found that a considerable diversity of music computational features could segment the pieces, and that even the first free improvisations (prior to any task demand instructions influencing the performers) were comparably segmented.

In the present work we assess the relationship between computational segmentation and performer SC. Since attentional demands are expected to be maximal at transitions between segments, we hypothesized that SC should reflect the segmentation, with enhanced SC during the periods in which transitions are generated between A and B and vice versa. We found previously that there were few if any 'canonical' skin conductance responses (SCR) in relation to the generation of musical segmentation (Dean & Bailes, 2015), where a canonical SCR is a substantial rise and fall of SC during a period of around 10 s, normally associated with a specific external stimulus, such as a loud sound (Bach et al., 2009; Lim et al., 1997). Clearly there were no such external stimuli in our study, and we anticipated that continuous SC changes might reflect performers' change in attention and/or arousal as they manipulate the transitions, but with few SCR. SC, whether or not displaying SCR, comprises both phasic (short-term adjustment or oscillation) and tonic (long-term trend) changes, like many physiological processes. Both for our performers and for an independent group of non-musicians not involved in the improvisations, we assessed how musical change was perceived continuously, in relation to the computational features of the music, and investigated factors that can contribute to models of such continuous perception of change. For the non-musicians only, we also assessed how the musical features of the improvisations related to their continuous perception of affect expressed by the music as they listened to recordings, following on recent contributions to an extensive line of research on modelling and interpreting such continuous responses (Bailes & Dean, 2012; Dean & Bailes, 2010, 2014; Dean, Bailes & Drummond, 2014). Key aspects of the earlier literature are covered in (Madsen &

Fredrickson, 1993; McAdams et al., 2004; Schubert, 1996), and reviewed by (Schubert, 2006). The present work is probably the first to measure continuous response perceptions of unfamiliar improvised music, but the situation of hearing a new genre or style of music and a completely unfamiliar piece can occur for us all.

2. METHODS and ANALYTICAL APPROACHES

2.1 Participants.

Seven male and two female professional keyboard improvisers from Sydney, Australia, performed improvisations and undertook a continuous perception of change task. There were two participants whose professional improvising was in music therapy. The remaining seven musicians had particular experience in contemporary jazz, where referents (compositions for improvisation) are primarily provided by the musicians themselves, rather than by popular songs or older jazz compositions. Of these, all had professional experience of free improvising, and in the opinion of author RTD (an improvising peer and colleague of these musicians), this experience was very considerable for four. As in every country, the number of professional improvising pianists in Australia who solely participate in free improvisation is very low, and none of these pianists fit that description.

Three males and eleven females undertook the continuous perception of change and separately a continuous perception of affect task while listening to recordings of the improvisations. They were undergraduate psychology students from the University of Western Sydney who participated in exchange for course credit and were aged from 18-45, $M = 23.1$, $SD = 7.0$. They completed the Ollen Musical Sophistication Index questionnaire, which provides an index ranging from 0-1000 that indicates whether the participant is musically sophisticated (index >500) or not

(Ollen, 2006). The highest score amongst our participants was 398, their mean was 132, and *SD* = 124, so we describe them in this paper as 'non-musicians'. We do not have specific information on their listening habits, but in previous work we have found that such a group of students is unfamiliar with related genres of computer and electro-acoustic music, and late 20th Century Western instrumental and orchestral music. We would expect them to be unfamiliar with the improvised music styles engaged here, and none indicated any familiarity during the experiments.

2.2 The Improvisation Tasks

The improvisers played a Yamaha Disklavier grand piano (capturing MIDI data). They each performed an opening free improvisation, eight referents, and a closing free improvisation. For each performer one task from each row of Table 1 below was undertaken, and they were systematically cycled through the columns of tasks, such that each individual undertook a roughly balanced section of the fully-specified and the two types of partially-specified referents. Across all improvisers, there was close to balanced coverage of all tasks. The performers were asked to make each piece <= 3 min, and to play only on the keys. They were to use free pedalling, but with their right foot only. The left ankle was the site of skin conductance and pressure sensors (see below), and was to be kept as still as possible, since movement, especially local flexure at sites of skin conductance measurement drastically influences the measure (see below for our means of handling this). Performers gave a single clap before and after each performance, for file time alignment checking.

2.2.1 Measurements during the improvisations

During performance we recorded continuous measurements of skin conductance (SC), "pressure" (a measure of the performers' ankle surface movement), sound amplitude[1], and MIDI, as defined previously. Because all performances were solo piano, and played solely on the keyboard, spectral flatness change throughout the pieces was limited, and we did not investigate it: in previous work on pieces with complex and substantial spectral flatness change we have found modest predictive roles for it in models of continous perceptions of arousal and valence, but almost no role with solo piano music.

2.3 The Perception Tasks.

After they had completed their improvisations, performers listened back to five of them, in each case their first free improvisation plus four more items spread systematically across the range of referents in relation to the different performers. The referent characters auditioned in this part of the experiment (see Table 1 for details) were, apart from the free improvisation, Dense, Pulse, Staccato and Tonality[2]. An individual performer listened only to their own performances. The recordings were presented over headphones in a random order, with the participants now seated at a computer. They were told that their task was to detect whether the music changed,

[1] Sound amplitude was converted to relative logarithmic intensities by taking log10 of the squared amplitude, instead of measuring logarithmic acoustic intensity in dB SPL from the recordings.

[2] In each case, here and later we use the single term to refer to a polarity e.g. Dense vs. Sparse; Pulsed vs. Unpulsed. As shown in Table 1, individual referent structures may require transitions across this polarity, or maintenance of one or other of the two polarities.

and to indicate this while listening by moving the mouse during any perceived change.

> The greater the change, the faster you should move the mouse. For example, it may be that you wish to make a scrubbing motion with the mouse to indicate a strong and sudden change in the music.
>
> The smaller the change, the slower you should move the mouse. For example, it may be that you wish to move the mouse only slightly to indicate a subtle change in the music.
>
> Please move the mouse for the **duration** of any change.
>
> If you DON'T think the music changes, keep the cursor still.
>
> Please try to maintain your CONCENTRATION throughout each piece.

2.3.1 Non-musician listening tasks

In a later listening study, 14 non-musician participants also performed this ‘change in sound’ task as well as an ‘affect’ task while listening to a subset of nine of the recorded improvisations. The recordings comprised the following referents: Dense-Sparse-Dense, Staccato-Sustain-Staccato, Tonal-Atonal-Tonal (two performances), Dense with optional change, Pulsed with a required change, Pulsed with optional change, Staccato with required change, Free improvisation (the first of the set).

In the ‘affect’ task, respondents were asked to continuously rate the affect conveyed by the music along two dimensions simultaneously, arousal and valence, by moving a cursor around the screen, with arousal on the vertical axis (ranging from very active to very passive) and valence on the horizontal axis (ranging from very negative to very positive).

The 'change in sound' and the 'affect' tasks were performed in a counterbalanced order, and each was preceded by a practice trial. We have discussed these measures in detail in previous papers (e.g. Bailes & Dean, 2012; Dean & Bailes, 2010; Dean et al., 2011).

2.4 Analyses

All data were sampled at 2Hz (or when necessary down-sampled to this frequency) so all models discussed are conducted at that resolution (one time series lag is 0.5 s). The analyses here do not concern the free improvisations, since we have already shown that each of these performances can be segmented on the basis of several individual musical features (Dean, Bailes & Drummond, 2014), while here we aimed to study the relationship of perceptions of a piece to its single pre-specified referent musical feature.

2.4.1 Performers' skin conductance: Segmentation

Skin conductance (SC), and the music perception time series defined above, are strongly autoregressive (e.g. Dean & Bailes, 2015 for SC). This means that several events in the recent past are statistical predictors of the next; this also indicates that the data are not independent of each other, and have to be treated by non-standard statistical analyses which do not assume such independence, notably autoregressive Time Series Analysis (TSA). We have previously provided a tutorial introduction to the basic techniques of TSA (Dean & Bailes, 2010) and discussed its applications to continuous musical data extensively (Bailes & Dean, 2012; Dean & Bailes, 2010, 2015; Dean, Bailes & Drummond, 2014). We try to provide pointers to the key features here and later in the paper. Besides taking account of autocorrelation, TSA normally also requires that the data series under study be stationarised, that is trends removed such that the autocorrelation between events n samples apart remains unchanged across the series, for every n, within statistical limits. Commonly, stationarity is achieved by differencing, that is constructing a new series which is the difference between pairs of successive values of the original (and hence one member shorter), that necessarily still contains all the original information providing the starting value is known. Differencing is used here, and *dseriesname* designates the differenced form of *seriesname*.

In the previous work we showed that SC time series from improvising pianists often require 'cleaning', to obtain an SC trace from which the sometimes major influence of foot and leg movement has been removed. We developed there a method for doing this, by determining the 'transfer function' of such movement onto SC for each individual performance. The transfer function here is the predictive relationship between the time series of movement and that of the SC, when the autoregressive properties of each are properly taken into account (see detailed discussion in Dean & Dunsmuir, 2015). The SC analyses in Results section 3.1 use appropriately cleaned SC data obtained using this method. In later Results sections, where we undertake time series models, the movement parameter is retained as a possible predictor to be considered in the analyses, and results on it are presented, thus in these cases SC is used without cleaning.

We needed to obtain principled estimates of the times in both perceptual and SC time series which delineate statistical segments; in other words provide measures of the times of start and finish of the ABA segments in computational musical terms; or provide estimates of how many statistically distinct segments a continuous SC or perceptual time series contains and where they too start and finish. For this purpose, the 'changepoint' package in R (Killick & Eckley, 2012) was used (abbreviated cpt in the method names), and specifically the cpt.meanvar method, which as the latter part of its name is intended to indicate uses information both about changes in mean value and in the variance across the series. This is appropriate in our perceptual data, and in the SC studies, where we are particularly concerned with changes in the phasic response patterns and not simply any canonical SCR (skin conductance response, over a short period) which might occur; there were very few such SCR in the traces, as expected given the absence of defined external triggers: see below. Note that changepoints detected by this package primarily indicate where segments begin and end: they distinguish segments, rather than individual points of change. The changepoint algorithms can be used to identify all segments at a given probability level (as below), or required to determine a user-set number of most likely segments (using maximum likelihood methods). The latter method has been used in the prior work on computational segmentation of the performed musical streams, showing that in virtually all cases the performers successfully generated the requested ABA referent with respect to the relevant pre-specified musical referent feature (Dean, Bailes & Drummond, 2014).

2.4.2 Performers' skin conductance: measurement of SC events

We refer to an SC changepoint as a skin conductance event (SCE), to try to make clear that it is not a canonical SCR (skin conductance response), rather a reflection of a point at which SC series characteristics segment. It is particularly important to realise that the SCE is merely the point at which one segment ends and the next begins, and what distinguishes them are the features throughout the segment length, not just at the SCE. With known triggers eliciting canonical SCR, the time delay between trigger and response is variable, and might well average 10 s and range up to 20 s, as discussed previously (Dean & Bailes, 2015). Given that the SC changes we are interested in here might commence before a musical transition point (perhaps because they relate to the planning of this transition), and continue thereafter, the SC changepoint event (SCE) might be correlated but not necessarily absolutely coincident with the musical one. Thus we chose in advance of the analyses to focus on SC windows of 20 s, within 10 s either side of musical transition points.

Given that SCE are not triggered SCR to experimentally defined external stimuli, we analyse them on the basis that in common with most repeating and related events they are distributed independently in time under a Poisson law. The Poisson distribution indicates the probability of observing given numbers of events in a unit time period (or space), given the average number of events in such a unit time (or space): it is a discrete probability distribution based on a power law. For an individual participant, we expect the SC behaviour (i.e. the baseline event rate) to be relatively consistent. Thus for an individual, we assess the Poisson mean SCE/time for the 20 s segments around the musical transitions (the 'transition zones'), and compare this with the Poisson mean SCE/time for the remaining (control) period. This can be done for each performance, though there may be too few events for this to be statistically meaningful. It can also be done for all the performances of an individual taken together, and the mean difference between events/time in the transition and control periods is informative. We also measure the rate ratio (the ratio between SCE/time for the two conditions, musical transition zones vs. control period, which effectively scales the different performers' responses) between the two Poisson distributions for each piece. We can again aggregate across all the performances the number of SCE/time occurring in the musical transition zones, in comparison with the number of SCE/time occurring in the control periods. This permits a Poisson statistical rate ratio comparison.

2.4.3 Detection of musical segmentation in performances

A group of algorithms was coded in R for the purpose of measuring musical features and then identifying changepoint segmentation based on their changing characteristics as requested in some of the referents of Table 1, using solely the MIDI data acquired during performance. The algorithms are detailed in a previous paper (Dean, Bailes & Drummond, 2014) and summarized here. Most involve windowed analyses, where means and standard deviation over a chosen window length describe the potentially changing musical parameter (such as the event density), and can reveal transitions. Only in the case of the register referent (pitch range) was simple inspection (or measurement) of individual pitches sufficient to define the transition points; in all other cases, the windowing approach determines that the performed transition will be analysed as a relatively smooth one and the changepoint package is effective to determine the segmentation.

2.4.4 Time series analyses of perceptions of change and affect

With the non-musicians we chose to measure both perceived change and perceived affect. We considered the latter task inappropriate for performers who had already generated and experienced their own arousal in making the music, and we were also limited by experimental time available per session with the performers. Thus the performers only undertook the change task. We describe the analysis of perceived change next, and then the analyses of affect.

The perceived change series were analysed first to assess whether increases in the temporal patterns of perceived change coincided with the computational musical segments of the performances. An approach similar to that described above for SCE was used, but rather than simply quantitating the frequency at which changepoints occurred, it was necessary also to assess the summed absolute magnitude of perceived change across the time series zones being compared. The analyses were done on the perceived change differenced once to stationarity, and prewhitened to avoid spurious cross-correlations which otherwise often obtain between pairs of highly autocorrelated series (Dean & Dunsmuir, 2015). Pre-whitening is removing the autocorrelations from one of the pair of series in question (by filtering with an autoregressive function obtained by modelling the series purely autoregressively); this allows a reliable estimation of a cross-correlation function and its significance, when required, and a secure assessment of relationships between pairs of series. This is elaborated in the Results section. These analyses were done for responses from both the performers and the non-musicians.

The analyses mentioned so far involve univariate time series models (and in the last case, their comparison). But in the specific case of performers' perceived change, we also optimised multivariate VARX (vector autoregression with eXogenous predictors) models. In VAR, a joint model of several variables is constructed, and it can allow not only for their being autoregressive, but also possibly showing uni- or bi-directional relationships with each other (such variables are commonly termed 'endogenous'). The Granger Causality test is a means of determining the statistical likelihood that a given endogenous variable influences another, and equally importantly, in which direction(s) the influence flows. In contrast to endogenous variables, an eXternal predictor is 'exogenous', in the sense that it cannot be influenced by the endogenous variables, though it may influence them. For example, the acoustic intensity of a segment of music cannot be influenced by listeners' perceptions of the music, but it may influence those perceptions. Candidate eXternal predictors here were the referent musical feature series (e.g. density-sparsity), and acoustic intensity, while SC was initially treated as an endogenous variable (potentially both influencing and being influenced by perceived change), and autoregression of the endogenous variables was included as appropriate. Note the limitation in this approach that performers' in-performance SC cannot be measured at the same time as their perceptions of change. Depending on the results, SC can if appropriate be reframed as an eXogenous variable, simplifying the model.

Turning to the non-musicians' perceptions of affect, we then sought to obtain models, by means of time series analysis techniques, of their perceived change, arousal and valence time series, considering all the relevant predictors within our data. These were autoregression; acoustic intensity (a dominant factor in previous work, and the driver of perceived loudness); the computational musical parameters; the perceptions of change by both the performers and the non-musicians themselves; and the performers' SC. Since the on-going SC series itself embodies the data which allow the detection of SCE, it is not expected that SCE themselves (being infrequent in any case) would be significant predictors of perceived change. Perceived change was considered as a possible predictor of perceived arousal and valence, but arousal and valence themselves were not considered as predictors, but solely as autoregressive influences upon themselves. Such possible inter-relationships have been studied in depth previously (Dean & Bailes, 2010), and are at most small.

3. RESULTS

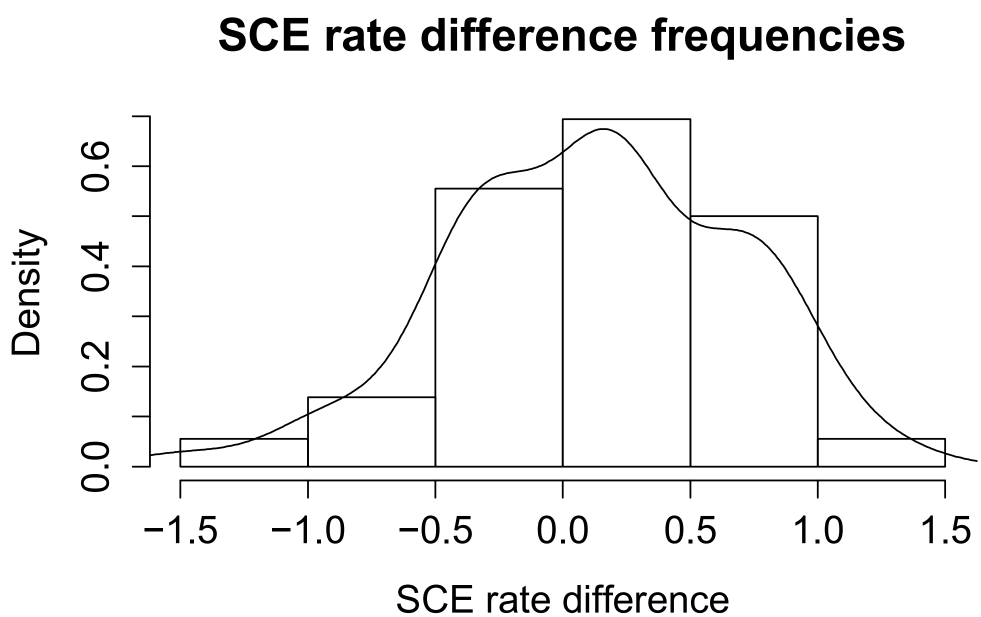
3.1 Segmentation of Performers' Behaviour and Retrospective Perceptions

3.1.1 Skin conductance (SC) segmentation points during performances: Do these relate to the generation and perception of musical structure?

In the first analyses, the number of changepoints to be detected was not user-set, but determined by an asymptotic probability limit (one to which the data are determined to converge as the number of data points goes towards infinity). At an asymptotic probability limit of $p < .05$, the changepoint analyses of SC showed two changepoints (i.e. two SCE and hence three segments of skin conductance) in every improvisation. A Poisson analysis compared the rate at which SCE occur in the music segmentation transition times (taken as ±10 s from the musical changepoints) with the rate in the other periods of time (the 'control' periods). None of the event rate differences for an individual performance were significant (as judged by the Poisson rate ratio test), which is to be expected given the small number of data points involved. The mean SCE rates (events/20 s) across all performances were: 0.38 for musical transition periods, and 0.26 for control periods. The mean difference, calculated from separate difference values for each performance was correspondingly 0.12. Forty-five (out of 72) performances showed higher rates in the musical transition periods than in the remainder. Testing whether this proportion is greater than 50% showed $p = .02$. Figure 1 shows a histogram of such rate differences across all performances.

Given that SC varies considerably between individuals, a ratio by individual between the SCE rate in transition versus control periods effectively scales the data, and facilitates comparison. Thus aggregating all the performance SCE and the corresponding musical transition and control time windows showed that overall the rate ratio between the occurrence of SCE in the musical transitions vs. control periods was 1.71 (95% confidence intervals (CI) [1.21,2.40], $p = 0.02$). We conclude that skin conductance change in the performers did occur preferentially in the periods in which musical change was being improvised, as hypothesised.

Figure 1. Histogram of SCE rate differences (all performances)



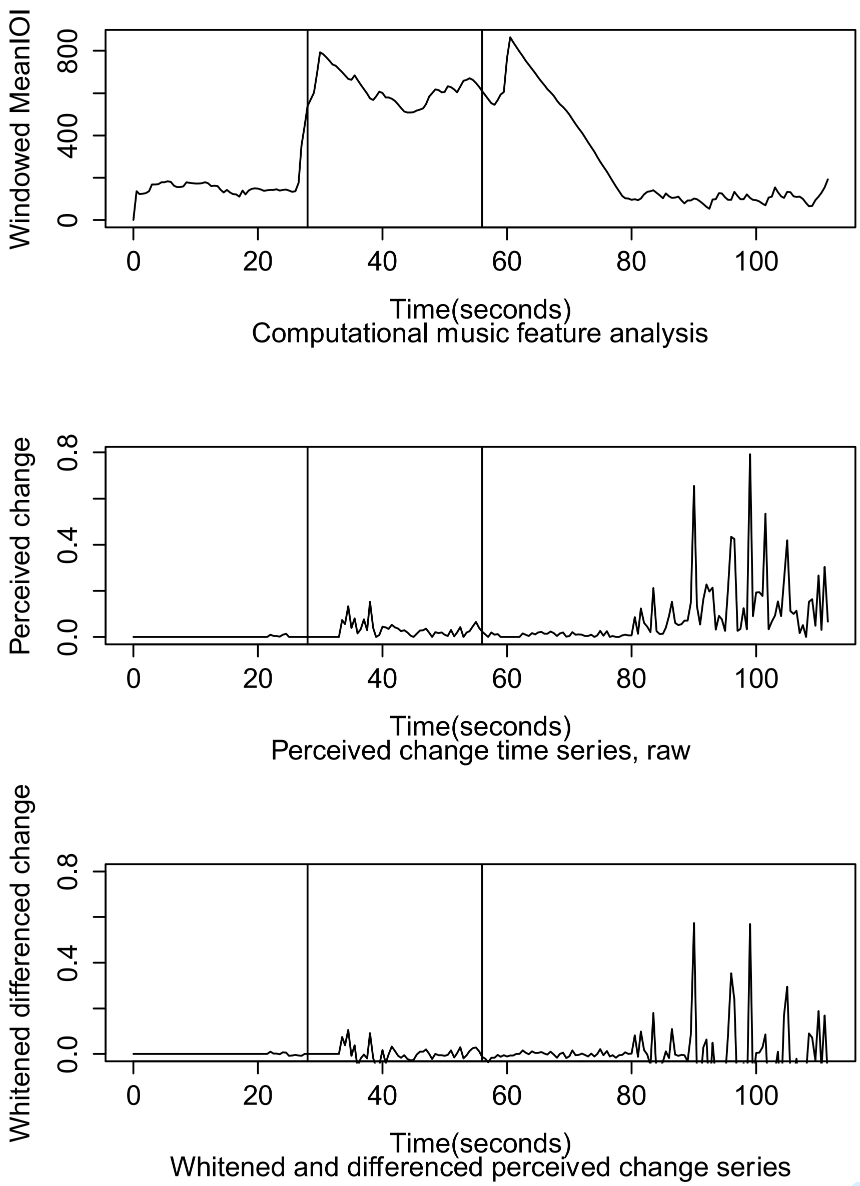
SCE rate differences for all performers are shown as frequency densities. The line shows a kernel estimate of the density distribution. It is clear that differences greater than zero exceed those lower.

3.1.2 Were the segments in performers' perceived change in the improvisations related to the computational musical segments?

We attempted to use exactly the approach just described to address this analogous question, with the hypothesis that performers' perceptions would be clearly related to computational changepoints. Amongst the 12 performer-auditioned referents, three performers heard each of the fully-specified referents Dense, Pulse, Staccato and Tonality (those with both A and B as specified musical features), and we studied these first. We found the perceptual changepoints (again using asymptotic $p < .05$) were difficult to detect, because the perceived change data were very sparse; mostly the algorithms just detected the first point at which the performer-listener registered change, and the end of their responses, with little or no delineation between (using cpt.meanvar or cpt.mean). Inspection of the time series data showed that this was comprehensible.

Thus an alternative approach was used, developed from one we used in Dean and Bailes (2015) for the analysis of skin conductance data. We compared the absolute differenced perceived change series across the music computation segments, with the hypothesis that the sum of this series per 20 s in the musical transition zones would be greater than the sum per 20 s immediately preceding, because the listeners perceive change as it happens. It is not reliable to work on isolated time chunks of an autocorrelated time series (Dean & Dunsmuir, 2015), as spurious relationships can be detected, rather for such purposes the series needs to be 'pre-whitened', which as mentioned above means removing its autocorrelation, and further analysis should be conducted on the resultant 'residual' series (that is, the part of the data which are not explicable simply by autocorrelation). We thus pre-whitened our differenced series using auto.arima, from the R 'forecast' package. This is an algorithm for automatically optimising an autoregressive model of a series; we allowed a maximum permitted autoregressive-order of 12, and used the Bayesian Information Criterion, which penalises strongly for the addition of predictors, as the selection criterion. The BIC tends towards parsimony and helps to avoid 'overfit' models, those that lack general applicability to the kind of series being studied, but are instead unique to the individual series instantiation in question. By inspecting the graphed trajectories, we observed that the changepoints detected in the music computational analyses were at the beginning of transitions, and thus we compared the data for the perceived change series for the 20 s after the music computational changepoints (perceptual transition zones), with that for the 20 s before them (perceptual control zones). Figure 2 shows a relatively clear-cut exemplar of the analysis, with respect to the Dense-Sparse-Dense referent (analysed as windowed mean musical inter-onset interval). There are three panels: top the musical referent feature data stream, middle the raw perceived change data and bottom the pre-whitened perceived change series (i.e the residuals from the purely autoregressive model), in each case with the musical segment positions indicated. Perceived change is limited in the opening and middle (B: sparse) section, and more extensive in the final dense section. It is also apparent, particularly in relation to the start of the B section and also clearly for the opening of the C section, that change is concentrated after the transition compared with before the changepoint (notably at 80 s).

Figure 2.



Analysing the Performers' perceived change. The panels illustrate the results from a single performance of Dense-Sparse-Dense (Referent #5). The vertical lines on each panel indicate the measured changepoints in the music computational data stream (windowed mean inter-onset interval: i.e. note or chord duration).

For the 12 listen-backs by performers to fully specified ABA referents, the mean of the ratios between the musical transition and control zone change sums (measured by individual performance) was 2.99(2.75) (geometric mean, dimensionless (*SD*)). Eleven out of 12 perceptual series showed ratios > 1, and for a one-sided proportion test showed $p = .005$, 95% CI [0.65,1.0] for this proportion. The difference between paired transition and control sums was significantly greater than 0 ($p = .028$), and showed CIs of [0.25, Infinity]. Thus the musicians seemed to be influenced by the referent parameters.

This analysis was extended to consider all the eight partially-specified referents (four with a required change, four with an optional change) auditioned by the performers (three performers undertook and listened to each of the referent character types, for a total of 24 listenings). In each of these partially-specified cases the 'A' feature was maintained throughout the performance. As we showed previously, these partially-specified referents were virtually always realised successfully: any change in 'A' was slight in comparison with that occurring in the fully specified ABA referents. Nevertheless, given that we had forced the changepoint algorithm to find three segments in the musical data in the earlier analyses, we could detect changepoints even in the partially specified referents, but these may or may not be perceptually relevant, as we investigated further.

Given this, our expectation was that performers' perceptions for the additional 24 cases would not necessarily align with musical segmentation based on the 'A' feature, since it would not have been the feature they used to create their segmentation. Taking all the performers' perceptions of the 36 referent pieces they auditioned, 25 of 36 showed a ratio > 1 (proportion test: $p = .015$, CI for proportion [0.54, 1.0]). Since 11/12 of the fully specified referents showed a ratio >1, this result also reveals that only 14/24 of the incompletely specified referents showed such a ratio. The overall geometric mean (*SD*) ratio values for all 36 listen-backs were 1.43(3.09), and correspondingly this was not significantly greater than 1 (p = .07). Clearly, the performers' perceptions seemed much less driven by the primary referent in the partially-specified cases than in the fully-specified ABA referents, as we predicted.

3.2 Segmentation of Non-Musicians' Perceptions

3.2.1 Were the segments in the non-musician participants' perceived change related to the computational segments?

In view of these results on the performers themselves, we expected little or no relationship between music computational segmentation and perceived change by our non-musician listeners, excepting perhaps for the fully-specified ABA referents, where the referent basis of change might still be apparent. The 14 non-musician listeners heard four cases of the fully specified ABA referents (Dense-Sparse-Dense, Staccato-Sustain-Staccato, and two versions of Tonal-Atonal-Tonal), and four of the partially-specified referents (Dense with optional change, Pulsed with a required change, Pulsed with optional change, Staccato with required change: see Table 1 for details). We assessed the mean perceived change for all audio files. Non-musicians generally indicated more perceived change to a given piece than did the performers themselves, and the non-musicians' change series could be easily segmented into three or sometimes more parts using the asymptotic p = .05 cpt.meanvar changepoint approach.

We undertook the analysis of the relationship of perceived change to computational musical structure for the non-musicians in the same way as for the performers, by comparing summed changes in the control and transition zones. Only for one item, one version of the referent Tonal-Atonal-Tonal, was the transition period change sum not greater than the control rate. Thus the proportion test showed the CIs for the estimated participant proportion having higher transition sums as [0.52,1.00], and $p = .038$, suggesting greater perceived change in the transition zones. The geometric mean transition: control ratio values (*M*(*SD*)) were 2.50 (3.04) but this was not significantly greater than 1 ($p = .29$). The non-performer participants were thus clearly but not strongly driven in their ratings of perceived change by the specified referent musical parameter, taking all pieces together.

We wondered whether the mildly positive proportion test might be based in greater perception of change in the referent musical parameter in the fully-specified ABA cases, even by these non-musicians, so we also assessed the differenced changes individual by individual for those cases (14 x 4 responses). 33/56 responses showed ratios >1, but this was not a proportion significantly higher than 0.5. The geometric mean(*SD*) of the rate ratios for these fully-specified referents was 1.35(2.77). The individual differences between rates for the 56 perceived series were overall positive but the CIs of the difference value were [0.398, Inf], ($p < .0004$), indicating their considerable variability. The results support the earlier conclusion that non-musicians' perceptions of change were partially but not strongly driven by the specified musical parameter, even in these fully specified ABA cases.

3.2.2 Comparing the performers and non-musicians' perceptions of change.

We can now reasonably ask whether the non-musicians' and performers' segmentations were related, as would be the case if they all perceived the same combination of musical features as those creating the computational structural segments. This was assessed using the same method of summing pre-whitened and differenced perceived change series as just described, but now taking the performers' segmentation as the point of comparison. Note that for the non-musicians we have no strong grounds for predicting what the musical features predicting such segmentation may be. For example, we identified many parameters that can segment the free improvisations in particular (Dean, Bailes & Drummond, 2014) and expect that there are several ways of segmenting any performance, just as we noted above that note-event density or note-length referents may sometimes be better segmented on the basis of acoustic intensity. Rather what is being assessed first here is the relation or otherwise of performers' and non-musicians' perceptions of the location of transition zones in the performances.

The prediction we tested was again that the non-musicians' summed absolute differenced perceived change over the 20 s following the relevant performer's change series changepoints (transitions) would be greater than the corresponding sum for the preceding 20 s (control), as this would indicate significant similarity in perceptions. Initially, we used the mean perceived change series from all non-musicians taken together for each performance (excluding the free improvisation), and the geometric mean ratio (*SD*) between them was 1.1(1.2), but this was not statistically significant. Similarly there was no difference when all the individual non-musicians' change perception series (pre-whitened) were analysed. However, when the analysis was focused solely on the auditioned fully specified referents, an increased change in the transition zones was found: geometric mean (*SD*) of the ratio was 1.2(2.4), the difference between change rates had a mean of 4.2 ($p < .0001$), and the difference normalised in units of mean change rate by individual was 0.19 ($p < .015$). Apparently, the non-musicians' and performers' perceptions of the major transition zones are quite closely related in the cases of the fully specified referents, where the relevant features are likely most apparent, but not in the other cases. We return to the nature of the influences on non-musician perceptions of change in Section 4.

We conclude that the non-musicians perceived enhanced rates of change shortly after the performers' transition point in pieces with fully-defined referents, which apparently make the referent feature more transparent, and may drive the perceptions fairly uniformly across both performers and non-musicians, at least in those transition zones. Section 4 includes an assessment of whether there is any generality to this: does the performer's perception of change predict the non-musicians' perceptions throughout a piece i.e. between the transition zones? We expect that the non-musicians commonly perceive the on-going musical features in different ways from the performers, which would be revealed by a common failure of performers' perceptions of change to contribute to overall models of those of non-musicians.

3.3 On the Possible Relationships in Performers Between SC, Musical Structure Generation and Perceived Change.

It is well known that acoustic intensity changes, particularly large and abrupt ones, can influence SC (Bach et al., 2009). In contrast, our primary interest here was investigating whether performer SC is related to cognitive processes involved in the fulfilment of the specified ABA referents, and production of the necessary musical contrasts. Following our observation above that SC change is intensified around the music structure transition zones, we considered more specifically that SC changes would precede the musical transition, and be predictive of it. Given the variable time delays in SC in relation to a stimulus, this could not be tested directly. If, but only if, these particularly SC changes are in general large in comparison with other SC changes, then such a feature of the transition zones may be reflected in an overall time series relationship between SC and the changing musical referent parameter. If the SC change at such points is limited (yet influential), then such a relationship can only be detected in the transition zones themselves, and is absent elsewhere. So as a weak test of the specific proposal that SC changes predict transitions we tested these possibilities sequentially, and using VARX could also assess the complementary possibility that the musical change process itself might initiate SC change. We also investigated the possible influences of the other measured parameters, the musical features, acoustic intensity and the movement parameter. For these analyses we were necessarily restricted to the 12 fully-specified ABA referents which the performers cumulatively auditioned (because in the others we do not know which musical features to include in such a 'competitive' analysis). While there were for example three improvisations around the Dense-Sparse-Dense referent auditioned by three different performers, the three improvisations are necessarily very different and need to be treated separately.

In order to give the reader an awareness of the detail of the analytical method and interpretation, Table 2 shows a complete example of the joint VARX models of SC and the performer’s own perceptions of change in an improvisation around the referent Dense-Sparse-Dense. Here L1 and L2 indicate lags 1 and 2 respectively, and the data required differencing to stationarity (as mentioned, *dseriesname* indicates the first differenced form of *seriesname*). This VARX modelled jointly the performer's skin conductance during performance and perceived change in the music during retrospective listening. These two variables were treated as endogenous (that is potentially both autoregressive and with mutual influences). The Granger causality analysis (bottom of table) showed they were indeed mutually influential in this case. The Table 2 output first describes the parameters, overall fit (R^2) and probability (p-values) of the components of the model. Secondly, it presents the coefficient for each predictor and its standard error, significance and 95% confidence intervals. Note that the VARX model design requires that the two dependent variables are modelled together, and hence a predictor may be essential for one of the dependent variables and have a highly significant coefficient, but yet be unimportant with a non-significant coefficient for the others (e.g. here lag 2 of dchange is not significant for predicting dsc, but is required for the autoregression of dchange).

Table 2 shows that the computational musical variable series (windowed mean inter-onset interval) was not a predictor of perceived change, but acoustic intensity paralleled it (short note sections have high intensity) and seemingly replaced it as a predictor. The exogenous predictors retained in the optimised model were lag 1 of the differenced intensity series, lags 1-3 of the differenced movement series (the measurement of the performer's left leg movement during performance), and second order autoregression was selected. The BIC (Bayesian Information Criterion, used to select the preferred model) was -588.56.

Table 3 summarises the information from all the VARX analyses taken together that is pertinent to our purposes (the summary of the data from Table 2 is included). Only relationships whose individual coefficients are significant at $p < .05$ are listed; some models were purely autoregressive and hence do not appear in the table. All the selected VARX models shown had some predictor variables, and were highly autoregressive, with lags 1-4, and all concern first differenced variables. In some cases component models are very poor ($R^2 \sim 0.1$) particularly for SC.

Table 3 shows that overall SC was statistically influential on perceived change only in two cases, as judged by Granger Causality. In two cases, there was evidence that perceived change could aid modelling SC. So the earlier analyses demonstrate clearly the coincidence of SC change and transitions, but this does not generally translate into an overall SC/perceived change relationship across the whole time series. In five of the 12 cases the referent musical parameter influenced change and/or SC, and additionally in the Dense-Sparse-Dense referent (for participant ID 3) intensity change was coincident with the referent segmentation (not shown) and influenced perceived change. Evidence for the relationship between the musical referent and perceived change for performers was thus strengthened by these data. There were 4 of 12 cases in which acoustic intensity was included in the selected model, agreeing with earlier work that it is an influential factor on perceived change in music but showing that with demand-driven focus on other musical parameters, its relative influence may be diminished.

We showed above that performers' SC changes particularly in regions of musical transition. However, we conclude from the VARX analyses that this effect is not strong enough to contribute often to models of perceived change, most probably because of a lack of relation between the SC and perceived change in the longer time zones outside the transition periods.

3.4 Time Series Modelling of Non-Musicians' Perceptions of Change and of Affect

Fourteen non-musicians undertook both the change and the affect tasks, successively and in counterbalanced manner (whereas performers undertook only the change task). The non-musicians auditioned one free improvisation; four fully-specified referents (Dense-Sparse-Dense, Staccato-Sustain-Staccato, two performances of Tonal-Atonal-Tonal), and four partially-specified (Dense with optional change, Pulsed with a required change, Pulsed with optional change, Staccato with required change). We present analyses of all except for the free improvisation here, so that the computational measure of the core-referent musical feature can be considered as a candidate predictor throughout.

Somewhat more data (31,915 observations) were available for analysis here than for the performers' perceptions. Hence as a preliminary assessment and given the relatively extensive data we treated the eight referents as separate items but within a communal cross-sectional time series analysis (CSTSA). Detailed introductions to CSTSA are provided by Dean, Bailes & Dunsmuir (2014a, 2014b). Unlike any conventional averaging procedure across time series of responses, this preserves the integrity of all the individual data series during the analysis, and permits multilevel mixed effects analyses. The musical parameters were uniformly scaled, but used as if all representative of one parameter, whereas of course they were actually different features such as pulsedness, tonalness, mean note length (defined in Dean & Bailes, 2014). Only autoregression (and random effects on participants and/or items involving it) could be detected in CSTSA of all items taken together for models of either differenced perceived change, arousal or valence series, which were in turn poor. This was the case whether we used individual response series, permitting random effects by participant and by item, or mean response series with random effects only by item. Only poor models were obtained with CSTSA.

It seems from this that the improvisations are too disparate for such a CSTSA treatment. So a more refined analysis considered each referent performance individually, and used the mean time series for non-musicians' perceptions of change, arousal and valence as a basis for each model, in conjunction with the other available predictors. We noted that intensity was found above to replace some of the musical features (such as sparsity) and so included it as potential predictor together with the computational musical features (now of course treated as the unique features they are), both the performers' and non-performers' perceived change series, and autoregression of the chosen dependent variable. In this study, we did not assess possible relationships between the perception of arousal and valence, but for perceived arousal specifically we also considered SC as a predictor, as prefigured above in the analyses of SC and performer perceptions of change.

During initial data exploration of the perceptions of the individual pieces we noted that the mean change series showed strong discrimination within each improvisation, while the mean arousal and valence series did not. This can be illustrated through their coefficients of variation (CV: the ratio of *SD*/mean)[3]. The mean and range of the CVs were: change 0.55(0.34, 0.81); arousal 0.14 (0.07, 0.28), valence 0.11 (0.05, 0.26). The CV distributions for arousal and valence were also skew, as indicated by the ranges around the means.

These data suggest that the non-musicians did not perceive much expressed affect in these improvisations, consequently it is not surprising that further analysis gave limited information. Analyses of non-musician perceived change series for the eight pieces showed poor to fair models (R^2 from 0.08 to 0.33) with autoregression

[3] Solely for the determination of CV, we adjust the arousal and valence scales to run from 0-200 instead of the original -100 to +100 (as in previous work; unless this is done the CVs are not meaningful and cannot be compared).

dominant. Only for the referent of Staccato with required change was any other predictor useful: in this case, acoustic intensity.

For perceived arousal, four of the eight cases gave models with multiple $R^2 >$ 0.15. We found that intensity was influential for Dense-Sparse-Dense, Pulsed with a required change, and Staccato-Sustain-Staccato (three of those four cases). For Dense-Sparse-Dense and Staccato-Sustain-Staccato this is readily comprehensible as intensity paralleled the musical parameter. Consistent with this, the musical parameter itself was influential for Pulsed with a required change only. Non-musicians' perception of change was influential for perceived arousal for the Dense-Sparse-Dense and the Staccato-Sustain-Staccato referents. Autoregression was strongly influential in all cases, as expected (and it was the only predictor for Pulsed with optional change, Staccato with a required change and the alternative performance of Tonal-Atonal-Tonal). The performer's SC could not contribute to any of the models of non-musicians' perceptions of arousal, which is not surprising given earlier results on the overall SC series. The performer's perception of musical change was a predictor of non-musicians' perceptions of arousal for Dense with optional change, Staccato-Sustain-Staccato (presumably reflecting some additional musical feature(s) beyond the core referent in this particular case), and for one of the performances of Tonal-Atonal-Tonal. Thus there was not a uniform failure of performers' perceived change as a predictor, as we had wondered earlier; rather the data again indicate some commonality between the performers and the non-musician listeners.

As noted, valence responses were remarkably unchanging (mean CV = 0.11), indicating that the listeners hardly discriminated changing positivity/negativity during a piece, even though the mean values for the series for individual pieces were distinct. Correspondingly, the autoregressive models for the valence series were all extremely poor (multiple $R^2 <= 0.10$). Across the pieces, the same range of predictors as for arousal was found to have detectable (if very small) roles: intensity, the computational musical structure, the performers' perceptions of change, the non-musicians' perception of change, and autoregression. Perhaps this is not surprising given that non-musician listeners' perception of valence varies little during these relatively simple referent structures. This is reasonable in spite of the fact that sophisticated and complex professional levels of improvising ability are required to fulfil the referents coherently, particularly when concepts such as tonality are involved.

4. Discussion

This study was concerned with the structures improvised by solo pianists, examining their perception by the performers themselves as well as by non-musician listeners. In a bid to explore the ‘online’ cognitive processes involved in generating the musical material, we measured the SC of the performers, analysing the relationship between SC data and the observed moments of musical transition. Specifically, we hypothesized that SC would be enhanced during transitional periods, reflecting changes in the performers’ attention and/or arousal when generating contrasting musical material. For such an effect to indicate cognitive processes, it should precede the performed musical changes, with increased arousal predicting the musical transition. We found evidence that SC did change significantly around the times of observed musical change, but we could not show whether it was strongly predictive of the perceived changes (in a statistical sense). This can probably be best understood by considering the extreme case of a canonical SCR occurring in association with an imminent musical change, even though such canonical SCR were not common. With a relatively symmetrical rise and fall of SC during such an SCR, predictive information corresponding to an intervention spike or an intervention step would be neutralised as the fall counterbalances the rise. Thus it is perhaps unsurprising that SC was not predictive of the time course of perceived change. This was also supported by additional analyses in which we coded SC as zero in control periods of the response series but retained the measured SC values across the transitions, and tested whether such a modified SC series was predictive of perceived change. The results (not shown) were negative. As shown above, in some cases change was itself predictive of SC change, and both complementary and simultaneously bidirectional relationships of this kind (SC <-> Change) are quite common.

We further predicted that changes in SC while performing would relate to the performers' perceptions of change in the music when they retrospectively listened back to their work. In some instances, this was the case, with changes in SC predicting perceptions of change when listening. The direction of this relationship suggests that SC changes foreshadow audible transitions in the music, and thus signify increased attentional and cognitive focus rather than a perceptual reaction to the performed changes.

Asking the performers to later provide continuous ratings of their improvisations is susceptible to response biases. For example, the performers might be influenced by their desire to rate perceived change in accordance with the referent they had been instructed to perform. Indeed, the results suggest that the performers were influenced by the primary referents when rating their perceptions of change in the fully-specified improvisations, but less so when the referent structure was only partially-specified. So performers' perceived change in the fully specified referents is in line with the musical changes they introduced. Although this perception could in part have been driven by awareness of the earlier referent demand of the performance, it seems most reasonable to conclude that performers can probably perceive their own segmentation mechanisms in other circumstances too. This might be tested further by obtaining data on what parameters a musician-listener considered they were using to segment change in improvisations like the partially-specified referents here, and comparing these with both their continuous perceptions of change and the specified musical parameter(s). Clearly such parameters might well vary at different stages of a single listening. Currently we do not have such data.

Although the performers listened back to the recordings following a break, their memory for the improvisations is likely to have still been relatively fresh. Moreover, their involvement in the generation of the music surely provided them with a heightened knowledge and understanding of its features. For these reasons, non-musicians with no prior experience of the improvised material were invited to rate their perceptions of change in the music. In spite of their removal from the generation of the music, the non-musicians perceived structural changes in the music similarly to the performers, though to a lesser extent, and only in cases where the musical referent was fully specified, when the instruction to create a segmental contrast was explicit and evidently realised. Movement was commonly an influence on SC, as described previously (Dean & Bailes, 2015). In four cases, movement influenced perceived change, and possibly this can be understood in terms of the well-known relationships between performer movement and expression, or simply applied force (e.g. Todd, 1992).

Finally, it was of interest to examine the ways in which the non-musicians' perceptions of the affect expressed by the improvisations - along the dimensions of arousal and valence - related to the improvised structure, the performers' perceptions of this, and their SC during performance. The improvisations were not perceived to be particularly expressive with respect to arousal, though the continuous ratings of arousal varied sufficiently to be modelled in relation to a combination of musical features, acoustic intensity, and autoregression. It is noteworthy that the performers' physiological arousal (SC) did not contribute to any of the models of the non-musicians' perceived arousal. Listeners did not perceive continuous changes in the valence expressed in the improvisations, and consequently this affective dimension could not be modelled.

In conclusion, changes in the performers' SC occurred preferentially during the periods in which they improvised musical change, possibly reflecting enhanced attention and arousal at pivotal moments of heightened decision-making. As such, the continuous measurement of SC during improvisation is a promising approach to understanding performers' experiences as they occur, and aligning these to the musical decisions that they make. Further research is underway to measure the physiological arousal of keyboard duos as they improvise together. This will allow for an in-depth exploration of the mutual influence of cognitive, perceptual and physiological processes on the generation of musical structures in real-time. In this circumstance, future studies of inter-brain kinetic coordination, as well as intra-brain sequential and causal networks, will be important.

References

Bach, D. R., Flandin, G., Friston, K. J., & Dolan, R. J. (2009). Time-series analysis for rapid event-related skin conductance responses. *Journal of Neuroscience Methods, 184*, 224-234. doi: 10.1016/j.jneumeth.2009.08.005

Bailes, F., & Dean, R. T. (2012). Comparative time series analysis of perceptual responses to electroacoustic music. *Music Perception, 29*(4), 359-375.

Bailey, D. (1992 revised edition). *Improvisation, its nature and practice in music*. London: British Library (first published 1980).

Beaty, R. E. (2015). The neuroscience of musical improvisation. *Neuroscience & Biobehavioral Reviews, 51*, 108-117.

Bengtsson, S. L., Csíkszentmihályi, M., & Ullén, F.. (2007). Cortical regions involved in the generation of musical structures during improvisation in pianists. *Journal of Cognitive Neuroscience, 19*(5), 830-842.

Berkowitz, A. L. (2010). *The improvising mind: Cognition and creativity in the musical moment*. New York, NY: Oxford University Press.

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

Dean, R. T., & Bailes, F. (2010). Time series analysis as a method to examine acoustical influences on real-time perception in music. *Empirical Musicology Review, 5*(4), 152-175.

Dean, R. T., & Bailes, F. (2014). Influences of structure and agency on the perception of musical change. *Psychomusicology: Music, Mind, and Brain, 24*(1), 103-108.

Dean, R. T., & Bailes, F. (2015). Using time series analysis to evaluate skin conductance during movement in piano improvisation. *Psychology of Music, 43*(1), 3-23. doi: 10.1177/0305735613489917

Dean, R. T., Bailes, F., & Drummond, J. (2014). Generative structures in improvisation: computational segmentation of keyboard performances. *Journal of New Music Research, 43*(2), 224-236.

Dean, R. T., Bailes, F., & Dunsmuir, W. T. M. (2014a). Time series analysis of real-time music perception: Approaches to the assessment of individual and expertise differences in perception of expressed affect. *Journal of Mathematics and Music, 8*(3), 183-205.

Dean, R. T., Bailes, F., & Dunsmuir, W. T. M. (2014b). Shared and distinct mechanisms of individual and expertise-group perception of expressed arousal in four works. *Journal of Mathematics and Music, 8*(3), 207-223.

Dean, R.T., Bailes, F., & Schubert, E. (2011). Acoustic Intensity Causes Perceived Changes in Arousal Levels in Music: An Experimental Investigation. *PLoS One, 6*(4), e18591.

Dean, R. T, & Dunsmuir, W. T. M. (2015). Dangers and uses of cross-correlation in analyzing time series in perception, performance, movement, and neuroscience: The importance of constructing transfer function autoregressive models. *Behavior Research Methods*, 1-20.

Donnay, G. F., Rankin, S. K., Lopez-Gonzalez, M., Jiradejvong, P., & Limb, C., J. (2014). Neural Substrates of Interactive Musical Improvisation: An fMRI Study of 'Trading Fours' in Jazz. *PLoS One, 9*(2), 1-10. doi: 10.1371/journal.pone.0088665

Keller, P. E., Weber, A., & Engel, A. (2011). Practice makes too perfect: Fluctuations in loudness indicate spontaneity in musical improvisation. *Music Perception, 29*(1), 107-112. doi: 10.1525/MP.2011.59.1.107

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

Large, E. W., Palmer, C., & Pollack, J. B. (1995). Reduced Memory Representations for Music. *Cognitive Science, 19*, 53-96.

Lim, C. L., Rennie, C., Barry, R. J., Bahramali, H., Lazzaro, I., Manor, B., & Gordon, E.. (1997). Decomposing skin conductance into tonic and phasic components. *International Journal of Psychophysiology, 25*, 97-109.

Limb, C., J., & Braun, A. R. (2008). Neural Substrates of Spontaneous Musical Performance: An fMRI Study of Jazz Improvisation. *PLoS One, 3*(2). doi: e1679

Madsen, C. K., & Fredrickson, W. E. (1993). The experience of musical tension: A replication of Nielsen's research using the continuous response digital interface. *Journal Of Music Therapy, 30*, 46-63.

McAdams, Stephen, Vines, Bradley W, Vieillard, Sandrine, Smith, Bennett K, & Reynolds, Roger. (2004). Influences of large-scale form on continuous ratings in response to a contemporary piece in a live concert setting. *Music Perception, 22*(2), 297-350.

Ollen, J. (2006). *A criterion-related validity test of selected indicators of musical sophistication using expert ratings.* (Doctor of Philosophy), Ohio State University, Columbus.

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

Schubert, E. (1996). *Continuous response to music using a two dimensional emotion space.* Paper presented at the 4th International Conference of Music Perception and Cognition, Montreal

Schubert, E. (2006). Analysis of Emotional Dimensions in Music Using Time Series Techniques. *Context: Journal of Music Research, 31*, 65-80.

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

Todd, N. P. McAngus. (1992). The dynamics of dynamics: A model of musical expression. *Journal of the Acoustical Society of America, 91*(6), 3540-3550.

Table 1: Improviser Tasks

Referent Character	**Basis of change specified ('fully-specified' referents)**	**Basis of change not specified ('partially-specified' referents)**	**Change not required (i.e. is optional) ('partially-specified' referents)**	**Whether there is change is not known in advance, nor is a basis of change specified**
Free (item #1)				**Yes**
Sparse	**Sparse-> dense-> sparse (item #2)**	Sparse throughout (#3)	Sparse throughout (#4)	
Dense	Dense-> sparse-> dense (#5)	**Dense throughout (#6)**	Dense throughout (#7)	
Register of a single hand melody	low register -> high-> low (#8)	Register unchanged throughout (#9)	**Register unchanged throughout (#10)**	
Pulse	**pulse -> unpulsed-> pulsed (#11)**	pulsed throughout (#12)	pulsed throughout (#13)	
Quiet	Quiet-> loud-> quiet (#14	**Quiet throughout (#15)**	Quiet throughout (#16)	
Staccato	Stacc -> sustain -> staccato (#17)	Staccato throughout (#18)	**Staccato throughout (#19)**	

Tonality (performers interpret the term)	**tonal -> atonal -> tonal (#20)**	tonal throughout (#21)	tonal throughout (#22)	
Textural (rather than pitch motive based)	textural -> pitch based -> textural (#23)	**textural throughout (#24)**	textural throughout (#25)	
Free (#26)				**Yes**

Note. The ten improvisations done by the performers included one from each row (systematic choice pre-determined in a similar pattern to a Latin square), including the two free improvisations. Thus they did: Item 1 (Free); One of (Sparse 2-4); One of (Dense 5-7); One of (Register focus, single hand 8-10); One of Pulse (11-13); One of Quiet (14-16); One of Staccato (17-19); One of Tonality (20-22); One of Textural (rather than motivic) (23-25); and Item 26 (Free). Bold text gives the example of the improvisations performed by Participant 1 The cycle continues and then repeats with subsequent participants.

Table 2. A Complete Vector Autoregressive (VARX) Model Output of a Performer's Perceived Change: Dense-Sparse-Dense

Equation	Parameters	R^2	p-values	
Dsc	8	0.49	<.001	
Dchange	8	0.48	<.001	
Modeled variable Dsc	**Coef.**	***SE***	**p-values**	**95% CI**
Autoregression:				
L1.dsc	.80	.06	<.001	[.69, .92]
L2.dsc	-.47	.06	<.001	[-.59, -.36]
Endogenous Predictors:				
L1.dchange	-.24	.11	.029	[-.46, -.02]
L2.dchange	.02	.11	.834	[-.19, .24]
Exogenous Predictors:				
L1.dintensity	.22	.06	<.001	[.11, .33]
L1.dmovement	-.01	.32	.963	[-.64, .61]
L2.dmovement	.64	.37	.081	[-.08, 1.36]
L3.dmovement	.26	.33	.425	[-.38, .90]
Modeled variable Dchange	**Coef.**	***SE***	**p-values**	**95% CI**
Autoregression:				
L1.dsc	.16	.03	<.001	[.09, .22]
L2.dsc	-.17	.06	<.001	[-.23, -.10]
Endogenous Predictors:				
L1.dchange	-.73	.06	<.001	[-.85, -.61]
L2.dchange	-.26	.06	<.001	[-.38, -.14]
Exogenous Predictors:				
L1.dintensity	-.03	.03	.266	[-.09, .03]
L1.dmovement	.72	.17	<.001	[.38, 1.06]
L2.dmovement	1.37	.20	<.001	[.99, 1.76]

L3.dmovement	.94	.18	<.001	[.59, 1.29]

Granger causality Wald tests

Equation	**Excluded**	**chi2**	**df**	**Prob > chi2**
Dsc	Dchange	6.89	2	.032
	ALL	6.89	2	.032
Dchange	Dsc	34.32	2	<.001
	ALL	34.32	2	<.001

Note. dsc = differenced SC; dchange = differenced perceived change; dintensity = differenced intensity; dmovement = differenced movement (measured from the performer's left leg).

Table 3. Vector Autoregressive (VARX) Analyses of Improvisers' Perceived Change and Skin Conductance

Referent	Participant ID	Granger Causality between dSC and dChange?	Musical parameter/ influences	Intensity influences	Movement influences	R^2 dSC	R^2 dChange
Dense-Sparse-Dense	3	SC -> change, change -> SC	Mioi/no	Intens->SC	Move -> Change	0.49	0.48
	5	No	No	No	Move ->SC, Change	0.41	0.36
	7	No	Mioi ->change	No	Move ->SC	0.10	0.17
Pulsed-Unpulsed-Pulsed	6	No	Pulsedness/no	Intens -> Change	Move ->SC	0.09	0.23
	8	No	Pulsedness/no	No	Move ->SC	0.11	0.48
	9	No	Pulsedness -> Change	No	Move ->Change	0.23	0.48
Staccato-Sustain-Staccato	1	No	Notelength -> SC, Change	No	No	0.57	0.32
	2	SC -> Change	Notelength/no	No	No	0.04	0.08
	3	No	Notelength/no	No	Move -	0.45	0.29

					>SC		
Tonal-Atonal-Tonal	6	No	Tonalratio ->SC	Intens -> Change	No	0.02	0.18
	8	Change ->SC	Tonalratio -> Change	No	Move ->SC, Change	0.58	0.34
	9	No	No	Intens -> SC	Move ->SC	0.46	0.36

Note. Mioi = mean inter-onset interval (windowed); dSC = differenced SC; dChange = differenced perceived change; Intens = intensity; Move = movement. An arrow indicates a significant predictive relationship. Note that several participants appear twice in the table.

Figure captions

Figure 1. Histogram of SCE rate differences (all performances)

Figure 2. Analysing the Performers' perceived change. The panels illustrate the results from a single performance of Dense-Sparse-Dense (Referent #5). The vertical lines on each panel indicate the measured changepoints in the music computational data stream (windowed mean inter-onset interval: i.e. note or chord duration).