Time series analysis of real-time music perception: approaches to the assessment of individual and expertise differences in perception of expressed affect

Roger T. Dean*, Freya Bailesa,c, and William T.M. Dunsmuirb

aMARCS Institute, University of Western Sydney, Sydney, Australia; bDepartment of Statistics, University of New South Wales, Sydney, Australia; cSchool of Drama, Music and Screen, University of Hull, Hull, UK

(Received 12 March 2013; accepted 23 May 2014)

We use time series analysis methods to detect differences between individuals and expertise groups in continuous perceptions of the arousal expressed by Wishart’s electroacoustic piece Red Bird. The study is part of a project in which we characterise dynamic perception of the structure and affective expression of music. We find that individual series of perceptions of expressed arousal often show considerable periods of stasis. This may challenge conventional time series methodologies, so we test their validity by application of a general linear autoregressive moving average (GLARMA) approach, which supports it. Acoustic intensity is a dominant predictor of perceived arousal in this piece. We show that responses are time-variant and that animate sounds influence the conditional variance of perceived arousal. Using vector autoregression and cross-sectional time series analysis (which preserves the integrity of each individual response series), we find differences between musical expertise groups (non-musicians, musicians, and electroacoustic musicians). Individual differences within each group are greater than those between expertise groups. The companion paper applies the developed methods to all four pieces in our overall project (Dean, R.T., F. Bailes, and W.T.M. Dunsmuir. 2014. “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). An Online Supplement is available at http://dx.doi.org/10.1080/17459737.2014.928752.

Keywords: music; time series analysis; continuous response analysis

1. Introduction

Music is the quintessential temporal art: unfolding in time, it often uses the time position and duration of events to establish structure. The perception of musical events often depends on how long they take to be generated or are maintained (Huron 2006). Yet relatively little work has addressed real-time (continuous) perceptions of music, see for example (McAdams et al. 2004; Dubnov, McAdams, and Reynolds 2006). Even less work considers its time series properties (Gregory 1989; Schubert 2004; Bailes and Dean 2009a). Time series analysis is a well-developed body of techniques (and will be introduced shortly). It has been used to detect musical pulse (Brown 1993). We use it to relate continuous acoustic features to listener perceptual features, and to inter-relate different perceptual features. For example, how do non-musicians perceive musical structure, since they do not necessarily understand it as conceived by composers? And

*Corresponding author. Email: Roger.Dean@uws.edu.au

© 2014 Taylor & Francis
do intensity (loosely, loudness) profiles (Ferguson, Schubert, and Dean 2011) or timbral features predict perceived structure and affect? Time series analysis investigates such continuous real-time correlations, whereas most discontinuous retrospective perceptions concern isolated chunks of music, 1 to 60 seconds long. Time series analysis takes account of the autocorrelation of the (non-independent) values comprising acoustic and perceptual time series. Standard statistical assumptions are violated by musical series, and time series analysis is one of the appropriate approaches: functional data analysis is a related alternative (Ramsay and Silverman 2005).

Focusing on Trevor Wishart’s piece Red Bird, we here define methods for analysing the responses of individuals and identifying differences amongst them; and for comparing responses of different musical expertise groups amongst the participants. In the companion paper (Dean, Bailes, and Dunsmuir 2014), we use the methods assessed here to undertake studies comparing four works. The studies belong to a larger project from which several papers have appeared. One paper introduced time series analysis methods for analyses of group average time series (Dean and Bailes 2010), and may be helpful to readers unfamiliar with these statistical methods.

Our purpose in this work is to understand listeners’ perceptions of the affect expressed by music more fully. Our approach is based on Russell’s classic “circumplex model” of affect (see Russell 1980, 2003). This model locates arousal/valence on two axes at right angles upon a 2D surface. Specific affective states are consequently represented in particular spots on the surface (on which only the axes are labelled). Consistent with arguments from nineteenth-century psychologist William James onwards, a listener may thus recognise features of music as indicating different levels of activity (“arousal”), and different positions on a continuum between positive and negative (“valence”), which is also sometimes construed as an axis between like and dislike. The expressed or perceived affective features need not be “felt,” that is, translated into the emotional or psychophysiological state of the listener.

The core hypotheses within this project are that changes in intensity are a dominant influence on perception of structure and expressed arousal; and that acoustic spectral properties are subsidiary influences on arousal, possibly more strongly influencing valence (Dean and Bailes 2010, 2011). We also hypothesised at the outset that what is perceived as structure influences what is perceived as expressing arousal or valence. If no structural change is perceived for a modest period, there may be no change in perception of expressed affect during much of that period. Thus, our overall statistical task is to relate the time series of acoustic features (such as intensity and timbre) to those of perception; specifically to develop and apply time series analysis methods to discern differences between individual and expertise group responses; and between musical pieces (see the companion paper).

For non-specialist readers, it may be useful to briefly explain how time series analysis operates and why it is needed. In virtually all musical time series (acoustic, structural, or perceptual), successive events are autocorrelated. This means that earlier events predict later ones to an important extent. Successive measures are therefore not independent measures of a common item; furthermore, linear models of such time-dependent processes may have residuals (unexplained portions of each value), which are themselves autocorrelated. Such residual autocorrelation needs to be removed before the model can be properly evaluated. Thus, consideration of the autoregressive nature of these time series is essential. Such endogenous control of the response series may act in conjunction with exogenous factors, both continuous (such as the acoustic intensity profile of a piece being heard) and discontinuous, such as a sudden change in instrumentation (and hence acoustic spectrum). The influences of exogenous factors can be modelled in conjunction with the endogenous autoregression; in fact, they cannot be determined reliably without consideration of the autoregression.

At any point in a piece of music, if a musical feature (such as a rise in acoustic intensity, or a particularly striking harmony) occurs, its impact on perception is moulded with the ongoing autoregressive process (which reflects the influence of both earlier perceptions and earlier
events). Thus, it is dangerous to isolate a few events from within a longer series for analysis without taking account of the surrounding autocorrelation structure, since within the chosen chunk, both local and more global (but no longer included) events are at work. So rather than taking chunks of series, models of the whole system are required. This also necessitates the process of “stationarisation” before analysis. Stationarisation involves removing trends that operate across the whole series, such that throughout the stationarised series, mean, variance and autocorrelations are constant, with respect to probabilistic limits. A variety of mathematical transformations of a series may be useful to achieve stationarity, but commonly, and in the cases described here, “differencing” suffices for this. Differencing is creating a new series by measuring the change between successive values of the original series.

The process of time series analysis then generates a principled mathematical model of the response variable in terms of autoregression and the influence of external variables (such as the acoustic profiles of the piece). The influence of the external variables is also known as the transfer function, indicating the transfer of influence from the exogenous variables to the response (the endogenous autoregressive variable). The models can be subjected to a range of tests of quality (such as the removal of autocorrelation in the model residuals), likelihood, and degree of fit to the data. Besides autoregression, moving average terms may sometimes be used to develop the endogenous component of the model: but this approach is generally less mechanistically interpretable. Moving average components can be re-parameterised as a purely autoregressive approach, and so we do not use them here. The resultant models of univariate responses are called ARMAX models, where the abbreviation indicates the consideration of autoregression (AR), moving averages (MA), and exogenous predictors (X), and this may also appear as ARX in some literature (omitting the moving average components, as we do). If appropriate, a statistical definition of intermittent temporal variabilities in the form of the model as the response series continues, may still follow. We provide examples of this in the work here. Vector autoregression with exogenous variables (VARX) is a multivariate counterpart to ARX. In previous papers we have provided extensive introductions to these techniques in the context of their application to music perception and structure (Bailes and Dean 2012; Dean and Bailes 2010), and we hope these articles will be helpful to interested readers.

Some notation will help formulate our models, analysis approach and hypotheses. Let \( Y_{r,k,t} \) denote a response \( r \) for \( r = 1, \ldots, R \) responses measured at time \( t \) for individual \( j = 1, \ldots, J_k \) in the \( k \)th expertise group where \( k = 1, \ldots, K \). Let \( X_{i,j} \) for \( i = 1, \ldots, I \) represent acoustic features (e.g. intensity, spectral flatness) or structural features observed at time \( t \) that respondents hear. For this paper \( K = 3 \), and \( J_1 = 8, J_2 = 8, \) and \( J_3 = 16 \) for the electroacoustic musicians (EA), musicians (M), and non-musicians (NM) respectively (described further below). The number of responses is \( R = 3 \), comprising perceptions of “change” in the music; and of expressed affect (arousal and valence). These responses are described more fully later; we concentrate on arousal here. All response time series here are of length \( T_{i,j,k} = 393 \), since we only discuss one musical extract. An alternative is to study the multiple response series \( (Y_{1,j,k,t}, \ldots, Y_{R,j,k,t}) \) as a vector series (Bailes and Dean 2012).

Typically, the change in each individual’s response series is analysed in relation to past changes in musical features through a transfer function of the following type

\[
\nabla Y_{j,k,t}^{r} = f_{r}(\nabla X_{1,t}, \nabla X_{1,t-1}, \ldots, \nabla X_{I,t}, \nabla X_{I,t-1}, \ldots, \delta_{j,k}^{r}) + U_{j,k,t}^{r} \tag{1}
\]

where \( U_{j,k,t}^{r} \) is a zero mean, stationary residual time series (Box, Jenkins, and Reinsel 2008; Greene 2008) modelled in terms of parameters \( \psi_{j,k}^{r} \) and \( \delta_{j,k}^{r} \), the unknown parameters specifying the transfer function relationship between inputs and outputs. These parameters of the residual series are not shown in the equation above. The symbol \( \nabla \) signifies taking first differences in time: \( \nabla Y_t = Y_t - Y_{t-1} \), termed “differencing,” as described above. We also refer in the body of
the text to the first differenced form of variable \( x \) as \( \Delta x \). We consider the situation in which

\[
f(\cdot)
\]

is a linear function of past musical feature series. We ask: do the \( \delta_{jk} \) vary across individuals \( j \) in group \( k \)? Do they vary with group \( k \)? The parameters of the residual process \( U_{jk,t} \) can also vary with individual or group but these are secondary to our objectives.

Strong statistical associations suggest hypotheses that can drive experiments from which causality can be inferred. Our first substantial outcome from such experiments, predicts and confirms a strong effect of manipulating the intensity profile of several pieces (Dean, Bailes, and Schubert 2011) on the perception of their expressed arousal.

Here, we assess differences in perception between groups of participants of different musical expertise, and between the individuals in those groups. We largely avoid familiar Western popular (or classical) music, and deal with a style unfamiliar to most participants (Bailes and Dean 2012). Specifically, we study here an electroacoustic composition, *Red Bird*, which we term a “sound-piece.” Unfamiliarity allows us to address perceptions of non-musicians grappling continuously with a piece and style for the first time. One group of listeners was specialised in electroacoustic music and so generally aware of the idioms of “sound-pieces.” However, even the members of this group were mostly unfamiliar with *Red Bird* (Bailes and Dean 2012). The music lacks instrumental sound and rhythmic repetition, but uses sculpting of sounds (only some originating from identifiable physical sources), and the compositional organisation in time of the resulting sounds. Like virtually all other music, the piece emphasises changes in intensity, timbre, and their temporal organisation. It involves motives and segmental structure, again features shared with most musics. The cue redundancy model (Balkwill and Thompson 1999) suggests that, in the absence of familiarity with a piece or genre, listeners use those features of music that are universal to sound: again pointing to intensity fluxes, timbral changes, and temporal organisation – hence our analysis of perceptual responses in relation to these acoustic features.

We can characterise the perceptual responses requested of participants briefly as follows. Participants express their continuous impression of “change in the music,” by mouse-scrubbing (see Section 2). We give no clues as to what musical change might be, making it clear that we are interested in participants’ views. We refer to this measure of perceived structural transition as “structure,” distinguishing it from affect, although we consider “change” in both parameters. Turning to affect, the participants used a computerised 2D-emotion space (Schubert 2004) to display their continuous impressions of the arousal and valence expressed by the music. A few studies have directly compared felt and expressed/perceived affect (Salimpoor et al. 2009), suggesting they are related but distinct. We and most others focus on expressed affect.

While non-musicians’ perception of structure does not necessarily coincide with compositional analysis, multiresolution analyses of musicians’ piano performances of music by Schumann (Beran and Mazzola 1999) show that performance timing and its deviations from notation can be modelled on the basis of musicological interpretation of the score. The authors took account of the autocorrelated streams. Thus, musicians at least may perceive similar structures to those conceived by musicologists.

Previously, we hypothesised that musical expertise groups might vary in their “subtlety of discrimination” of change and affect (Bailes and Dean 2012). As detailed in Section 2 our participants comprised a group with experience of electroacoustic music (later abbreviated EA), a group of musicians with no such specific experience (M), and a group of non-musicians (NM). Thus, we anticipated that discrimination might be ordered \( \text{EA} > \text{M} > \text{NM} \). Evidence from means and coefficients of variation for perceptual responses was broadly supportive. We also showed that both acoustic intensity and spectral flatness were predictors of inter-individual differences in perceptual responses to the Wishart and Webern pieces (the other two pieces addressed in our companion paper here were not analysed in this respect). Extending this, we show here that individual listeners vary substantially in their responses to the Wishart, and yet expertise-group specific differences remain. Initial time series analysis studies identify considerable periods of
stasis (no change) in individuals’ arousal response series. To assess whether this might invalidate the time series analysis analyses, by violating assumptions about the distribution of events across the response space, we apply methods of GLARMA (generalised linear autoregressive moving average modelling), with supportive outcomes. Thus, the conventional time series analysis methods are used in the companion paper for a systematic comparison of all four different musical works within our project. We also find here that responses to the Wishart work are time-varying. We observe influences of some specific timbral events, human and animate sounds (developing earlier analyses), but confirm that acoustic intensity change is a major influence on perceived arousal, predicting both the mean response and its conditional variance.

2. Materials, participants, acoustic analyses, perceptual methods, and statistical approach

Additional details of experimental methods have been given in our previous papers.

Pieces. This paper extends our previous studies on some electroacoustic music by Trevor Wishart, the section from c. 22’30” to 25’46” of his 45 min piece Red Bird (1977; available at http://www.ubu.com/sound/wishart.html) (Dean and Bailes 2010, 2011; Bailes and Dean 2012). A practice trial with another short piece was provided.

Acoustic analyses. We assess two continuous acoustic properties of the music, intensity (Sound Pressure Level, dB, on a log scale) and spectral flatness (Wiener’s entropy), a global parameter of timbre. Methods for analysing these have been detailed previously (Bailes and Dean 2009b). Note that Wishart’s abrupt and dramatic sonic events (Figure 1) can sometimes be interpreted in terms of recognisable physical objects (human or animate).

Figure 1. Acoustic and Unweighted Grand average Perceptual Wishart profiles. Intensity (db SPL) and Spectral Flatness (Wiener’s Entropy) are shown together with the grand average perceived structural change and perceived arousal profiles. For ease of presentation the original spectral flatness values are multiplied by five, (and so are around −50 to −25) and the perceived structural change values are multiplied by 100 (and so are between 0 and about 50).
Participants. As previously noted, there were 32 participants, aged 18 to 58 years (median 25.5 years), 16 non-musicians (NM), and 16 musicians. Amongst the musicians eight had electro-acoustic expertise (EA), while the others were generalists (M). The Ollen musical sophistication index (Ollen 2006) scores (abbreviated OMSI) were respectively: 16 non-musicians had a mean of 136 (“not sophisticated”); and 16 musicians had a mean of 789 (“sophisticated”). We refer to individual participants by a number $j = 1, \ldots, 32$ prefixing their group code; thus “17NM” refers to participant 17, NM group. Section 3 provides a summary of the statistical features of individual arousal responses.

Perceptual methods. At the outset of the experiment, participants learned to use the 2D-emotion space, and performed a practice trial in each of the “structure” and “affect” tasks. The perceived structure and perceived affect tasks both used a computer interface, and the methods are detailed in (Bailes and Dean 2009b). Briefly, in the structure task participants move the mouse if they perceive “change” in the sound, and rates of movement are averaged at 2 Hz. The second task provides continuous perceived affect measures, using the 2D-emotion space. Participants heard the piece twice, once for each of the two counterbalanced tasks. The piece was presented together with the other three works discussed in the companion paper, and the order of pieces within each task was randomised. The arousal response to Red Bird, averaged over all 32 participants, is shown in Figure 1.

Statistical approach. A variety of analyses were outlined in the introduction. Foremost we use time series analysis methods to relate the changes in musical features to changes in arousal response at the individual level, the group average level, and the grand average level as described previously (Bailes and Dean 2012; Dean and Bailes 2010). The methods we use for data exploration and model development include cross-correlation and Granger causality assessment of the relationships between changes in intensity and arousal. Since the piece was unfamiliar, we expected only preceding events to influence the next response: these preceding events are termed the lags in the series. Thus, lag 1 indicates the event immediately preceding the current one. We also considered possible influences of exogenous events still to come (termed leads), because in principle familiarity can lead to anticipation based on such leads. Leads were not influential in the present work. Based on the data exploration and rigorous model evaluation, transfer function models are developed and residual autoregressive properties determined and modelled. The resulting overall model is then applied to the grand average arousal response series and the group average series (Section 4) as well as to individual series (Section 5). These analyses suggest that differences exist between groups and individuals. To assess this we use vector autoregressive models for multiple series and random effects analysis. In these models it is again assumed that the transfer function relationship and autoregressive structure stay constant over the entire observation record and that the model obtained provides an overall assessment of time-lag relationships between musical features and listener responses. We later investigate the question of variation of model parameters across a response in Section 7 using rolling window model fitting techniques. In that section, we also use generalised autoregressive conditional heteroskedasticity (GARCH) models to find that the residual variation is conditionally time-varying. GARCH is somewhat like ARX in that it determines autoregressive influences on variability, while ARX focuses on predicting the mean response. We have introduced GARCH previously (Bailes and Dean 2012; Dean and Bailes 2010).

Later in the present papers we also seek sequentially to provide comprehensible introductions to the more specialised techniques we apply (such as cross-sectional time series analysis). Details on all these techniques are given in (Box, Jenkins, and Reinsel 2008; Greene 2008) and in the
supporting documentation for the two main statistical packages we use: STATA (StataCorp 2009) and R (R Development Core Team 2011).

Next (Section 4), we analyse the overall arousal response series obtained as the unweighted average

$$\bar{Y}_t = \frac{1}{J_1 + J_2 + J_3} \sum_{k=1}^{J_k} \sum_{j=1}^{J_j} Y_{j,k}$$

(2)

at each time of observation of the individual responses. Section 4 compares the average arousal response series of the three expertise groups

$$\bar{Y}_{k,t} = \frac{1}{J_k} \sum_{j=1}^{J_j} Y_{j,k}$$

(3)

for \(k = 1, 2, 3\). In Section 5 we develop standard time series models for all 32 individual arousal responses and assess whether we can explain differences between individual arousal responses in terms of expertise group or sophistication (OMSI) score.

**Modelling relationships between musical features and single response series.** The models used for our analysis require the time series to be stationary (that is, of essentially constant mean and variance, and with constant autocorrelations). Many series are not stationary, as for example they vary around long term trends or other smooth underlying patterns in the mean value. To assess stationarity, we used standard unit root tests. When a transformation to stationarity is required most commonly differencing is adequate, as was the case here. When this occurred, other variables were differenced similarly: in all cases the first difference was appropriate.

Once the response and predictor are stationary, we can assess whether they bear meaningful relationships to each other and which series lead or lag which others. Cross-correlation analysis is a data exploration tool that may indicate such influences, but it needs to be done with proper assessment of the significance limits of cross-correlations. The autocorrelations in the system influence these significance limits, and hence cross-correlation should be done after “pre-whitening,” which consists of removing autocorrelations from the putative predictor variable, then filtering the putative response variable with the same function, and finally determining the cross-correlations. Cross-correlation functions are not in themselves adequate definitions of inter-relationships between time series. What is needed is the transfer function model, which quantifies the regressive influences at various time lags of both endogenous (autoregressive) and exogenous variables on the endogenous response. Note that a perceptual response may actually mediate changes in another perceptual response (such that neither are pure “dependent” variables in psychological parlance), and this possibility needs to be considered. Clearly, acoustic variables are exogenous/independent variables, since they cannot be influenced by perceptual ones. Besides cross-correlation, we also determined Granger-causalities between changes in musical features and average or individual responses with VAR (vector autoregression). As mentioned, vector autoregression is, as its name implies, autoregressive analysis of a multivariate system that consists of a vector of values at each time point. Granger causality can give a probabilistic assessment of whether any variable influences any other; variables may be modelled explicitly as exogenous (that is, not influenced by others, as appropriate for the acoustic variables), or endogenous.

The next step was to formulate an ARMAX (autoregressive moving average analysis with exogenous variables) for the response series, where \(X\) is a series of relevant acoustic variables.
For a single exogenous variable this had the general form

\[ \nabla Y_t = W_t + U_t \]  
(4)

in which

\[ W_t = \sum_{l=1}^{L} \omega_l \nabla X_{t-l} \]  
(5)

and

\[ U_t = \sum_{l=1}^{P} \phi_l U_{t-l} + e_t, \quad e_t \sim \text{i.i.d.}(0, \sigma^2). \]  
(6)

Such ARMAX models are applied to the overall and group average series, and to each individual’s series in later sections. Note with reference to the final equation modelling the residuals that normality of the \( e_t \) are typically assumed for obtaining maximum likelihood estimates of model parameters in the R and STATA packages that we use. However, for our discussion of model properties the normality assumption is not required but only the assumption that these \( e_t \) be independent, identically distributed zero mean and constant variance. Indeed, the Gaussian likelihood (constructed assuming that they are normally distributed) can still be used to obtain what are sometimes called quasi-maximum likelihood estimates even when the Gaussian distribution is not the correct one. The associated inference for these quasi-maximum likelihoods is valid for large sample sizes. When there is substantial deviation from normality (as is the case when there is substantial stasis) then the usual large sample statistical inferential theory would be challenged and hence that is why we test the robustness of our findings using the binomial models in the Online Supplement.

In the usual regression setting the proportion of response variability explained by the regression component (here \( W_t \)) is expressed as an \( R^2 \) value. However, this does not incorporate the impact of lagged responses and so we reformulate the model (4)–(6) as

\[ \nabla Y_t = W_t + \sum_{l=1}^{P} \phi_l (\nabla Y_{t-l} - W_{t-l}) + e_t \]  
(7)

where \( e_t \sim \text{i.i.d.}(0, \sigma^2) \) as in equation (6). This form includes in the regression part the impact of changes in intensity \( \nabla X_{t-l} \) from the transfer function and past deviations in responses \( \nabla Y_{t-l} \) from the transfer function component \( W_{t-l} \).

In order to summarise the extent to which the model explains variation in the response, we let the total sum of squares be \( SST = \sum_t (\nabla Y_t)^2 \), the residual sums of squares be \( SSE = \sum_t e_t^2 \), and the sum of squares due to the model be \( SSM = \sum_t (\sum_{l=1}^{P} \phi_l (\nabla Y_{t-l} - W_{t-l}) + W_t)^2 \). Then \( P_M = SSM / SST \) is the proportion of total variation in response series \( Y_t \) “explained” by the combined effects of variation in previous changes in response around \( W_{t-l} \) and changes in intensity directly through the transfer function \( W_t \). Note that \( P_e = SSE / SST \) is the proportion of total variation in \( SST \) owing to unpredictable randomness in equation (7). We will additionally introduce the ratio \( P_W = SSW / SSM \) to describe the impact due to the direct action of \( X_t \) through the transfer function relative to the overall model contribution.

We wish to compare the shapes and impact of the transfer functions fit to individual series and relate these to other features of the listeners such as group membership or OMSI score. Use of the full transfer function (4) is not so convenient for this. To simplify comparison we define two
summaries of the impact of past changes in intensity on average level, \( E(\nabla Y_t) \), of changes in arousal:

**Impact of a sustained change:** If \( \nabla X_t \) increases in level by a unit amount from time \( t = \tau \) onwards (sustained change) then \( E(\nabla Y_t) \) will eventually (for \( t \geq \tau + L \)) change to a new level by the amount

\[
\omega^* = \sum_{l=1}^{L} \omega_l. \tag{8}
\]

**Impact of a pulse change:** If \( \nabla X_t \) has a unit pulse at time point \( t = \tau \) then the maximum impact on \( E(\nabla Y_t) \) will occur at time \( \tau + l_{\text{max}} \) where \( l_{\text{max}} \) is the lag at which the transfer function is maximized

\[
\omega_{\text{max}} = \max(\omega_1, \ldots, \omega_L). \tag{9}
\]

Significance assessments of model coefficients, parsimony, and information criteria guided the selection of the best models, which were thus not usually simply those with the smallest residuals. Parsimony means that we preferred models with fewer variables, and our use of the Bayesian information criterion (BIC) also enforces this, since it penalises strongly as the number of variables increases. Model precision is normally improved by increasing the number of variables, and so the BIC provides a consistent tempering of this temptation. The maximum lags required on the input series was 10, corresponding to 5 seconds, in keeping with data from the literature discussed above. Besides ARMAX and VAR analyses, cross-sectional time series analyses (including multilevel mixed effect analyses) were also used to assess the distinctions between participants: see the following material, including the companion paper (Dean, Bailes, and Dunsmuir 2014). Models were accepted when they gave rise to white noise residuals lacking significant autocorrelation, and uncorrelated with the observed data. Other statistical issues are detailed below.

3. Characteristics of individual perceptions of arousal in Wishart’s *Red Bird*

Figure 2 shows the diversity of the time series of original, unscaled, perceived arousal responses of individuals from each expertise group. The series for the non-musicians are shown in two groups of eight.

All three expertise groups comprise individuals who are quite variable in their responses. Some series in each group show very fine responses, with rapid change most of the time, while others show prolonged plateaux (stasis). When the time series are standardised (to zero mean and unit variance), patterns are more apparent, particularly for the musician group, but substantial variability remains. Figure 3 shows the lag 1 differenced standardised arousal responses for the individual participants 5EA, 13M, 22NM, and 23NM. The percentage of points for which there is no change in perceived arousal is noted, revealing large differences between individuals. Note that some participants vary their response during the piece, from or to fine-grained. Since participants heard each of the four pieces twice (once for each perceptual task) in a counter-balanced design, and were pre-trained on the interface, this is unlikely to be due to progressive familiarisation with the interface, but represents changing responses to the piece.

To assess this further, we measured the periods in which there was no change in perceived arousal \( (\nabla Y_{j,k,t} = 0) \) for each participant in three ways: the percentage of time for which the value of \( \nabla Y_{j,k,t} \) is zero, the average length of runs of zeros in \( \nabla Y_{j,k,t} \), and the maximum zero run length (a run length of 40 occupies 20 seconds). The substantial differences amongst participants (Figures 2 and 3) were confirmed using these measures. The three measures of response statistics
are independent of expertise group (using one way analysis of variance) or OMSI (using regression). Yet mean OMSI scores differ significantly between the NM and the other two expertise groups separately or combined (analysis of variance p-value less than $10^{-10}$). Hence, an explanation for the variety of response types (flat versus reactive) cannot be found in expertise grouping nor OMSI.

We assessed whether frequency of individuals’ perceptions of stasis in arousal (mean 206.8 points) was simply predicted by their perception of stasis in structure (mean 146.5). These values differed ($p < 0.02$ by t-test). There was no significant Pearson correlation between individuals’ values of the two parameters. Thus, individuals show distinctive relationships between perceptions of structure and arousal. Again, OMSI scores could not predict the frequency of perception of structural transition.

In Section 5 we present time series analyses of all individual series. In the next section we present linear transfer function analyses of group average responses using ARMAX.
4. Overall and group average responses to acoustic features

Figure 4 presents the unweighted group average responses $\bar{Y}_{k,t}$ for $k = 1, 2, 3$ given by equation (3) (corresponding to the EA, M, and NM groups) and the overall unweighted average response $\bar{Y}_t$ given by equation (2). This figure shows that, in spite of significant individual differences, the group average responses are very similar to each other, and clearly related to the intensity profile of the piece, as demonstrated previously. We also found (not shown) that when the participants were split (at the median) into two groups based on their proportion of stasis, the resultant groups were very similar to each other in perceived responses of change, arousal and valence, again suggesting some shared cohesion in responses across all participants.
4.1. Analysis of averaged responses to acoustic features of the Wishart piece

The acoustic and average perceived structure and arousal series $\bar{Y}_t$ of the Wishart piece are shown in Figure 1. The acoustic measures (Intensity, Spectral Flatness) and the perceived structural change track one another closely. We previously developed a strong model of these kinetics, based on autoregression and the exogenous influence of intensity (Bailes and Dean 2012; Dean and Bailes 2010), and we analyse it further here.

Based on the principles we introduced in Section 2 our modelling process was as follows. First, we tested the average arousal response $\bar{Y}_t$ and intensity $X_t$ for stationarity. We rejected stationarity of both by the Augmented Dickey Fuller, KPSS (Kwiatkowski, Phillips, Schmidt & Shin), and PP (Phillips, Perron) tests. Additionally the best fitting ARMA model to both series individually was AR(4) of which there was a polynomial operator root close to unity in confirmation of difference non-stationarity. Therefore we lag 1 differenced intensity $X_t$ and arousal $\bar{Y}_t$. To arrive at our model we prewhitened the series $\nabla X_t$ using the automatically selected AR($P$) model with $P = 3$, filtered the response series $\nabla \bar{Y}_t$, and computed the cross-correlation function. This showed clearly that the significant cross-correlations were between $\nabla X_{t-l}$ and $\nabla \bar{Y}_t$ for lags $L = 1, 2, \ldots$ and that there was no evidence that $\nabla \bar{Y}_t$ was leading $\nabla X_{t-l}$. Transfer function (ARMAX) models were then given a model for the noise term $U_t$ as below. Here, we prioritised BIC to reduce overfitting, retaining an extremely good model fit and white noise residuals.

The resultant model is specified by equation (4) with $L = 10$ and (6) with $P = 3$. The coefficients $\omega_l$ are shown in Figure 5 along with error bars as $1.96 \times s.e.(\hat{\omega}_l)$. The autoregressive coefficients were estimated (with standard errors) as $\hat{\phi}_1 = 0.323(0.051)$, $\hat{\phi}_2 = 0.103(0.054)$, $\hat{\phi}_3 = 0.113(0.053)$.

The values of the percentage of total variation in the arousal response series due to combined impacts ($P_M$) of transfer function, past responses, and random error ($P_e$) in equations (4) and (6) re-expressed in the form of equations (7) and (5) are (see Table 1) $P_e = 50\%$, $P_M = 50\%$. That is, 50% of total variation is accounted for by the model (7). Transfer function variability as a percentage of model variability is 71%: so around 35% of total variability is attributable to the differenced intensity transfer function.
Figure 5. Panels 2 to 4: Transfer function coefficients for three groups (EA, M, NM) (with $1.96 \times$ individual standard errors in green). Panel 1: Transfer for average of all 32 participant responses.

Table 1. Percentage of total variation in arousal response series due to the model and to unpredictable random error. As described in the text, we use the alternate model form (7). $P_M = \frac{SSM}{SST}$ is the proportion of total variation in response series $Y_t$ associated with the combined effects of variation in previous changes in response around $W_{t-l}$ and changes in intensity directly through the transfer function $W_t$, and $P_e = \frac{SSE}{SST}$ is the proportion of total variation $SST$ due to unpredictable randomness. The ratio $P_W = \frac{SSW}{SSM}$ describes the impact due to the direct action of $X_t$ through the transfer function relative to the overall model contribution (and not to the total $SST$).

| Group | $P_e$ | $P_M$ | $P_W$ |
|-------|-------|-------|-------|
| All   | 50%   | 50%   | 71%   |
| EA    | 57%   | 43%   | 39%   |
| M     | 74%   | 26%   | 82%   |
| NM    | 66%   | 34%   | 65%   |
| Individual | Mean  | 71%   | 29%   | 29%   |
|        | Minimum | 55%   | 6%    | 1%    |
|        | Maximum | 94%   | 45%   | 90%   |
4.2. Analyses of differences between groups in Wishart arousal perception

The individual ARMAX model (4)–(6) was developed for the average arousal response for the groups separately, each with the same structure (i.e. \( L = 10 \) and \( P = 3 \)) for \( k = 1, 2, 3 \). The transfer function coefficients are shown in Figure 5 with error bars of \( 1.96 \times \text{s.e.}(\hat{\omega}_l) \). The maximum significant autoregressive degrees were \( P = 2 \) for EA, \( P = 1 \) for M, and \( P = 1 \) for NM. Hence, the models for individual groups could be simplified by either reducing \( P \) or reducing \( L \) or both. However, we also wish to compare the models across groups and for that purpose we retained the (overfit) structure with \( L = 10 \) and \( P = 3 \) for all groups. The proportions of variation \( P_e, P_M \), and \( P_W \) based on the re-expressed models (7) and (5) are shown in Table 1. The variability in changes in intensity explained by the combined transfer function and autoregressive lags ranges from 43% for EA to 26% for M.

The similarities and differences of transfer function parameter values that can be observed in Figure 5 are of interest:

1. The longest significant transfer function lag is \( L = 7 \) for EA, \( L = 6 \) for M, and \( L = 10 \) for NM. Values of \( L > 10 \) were not significant.
2. The overall impact of a step change in intensity on long run expected change in arousal defined by equation (8) is largest for M (\( \omega^* = 0.93 \)), followed by EA (\( \omega^* = 0.80 \)), and NM (\( \omega^* = 0.67 \)).
3. The lag at which maximum response to a pulse change in intensity impacts expected change in arousal is defined by equation (9) and both EA and M groups have peak response at lag \( L_{\text{max}} = 2 \) and taper off to zero as lag increases, except for the strong secondary peak at lag \( L = 5 \) for the M group, whereas the NM group has a flatter transfer function shape with rather weak peaks at lags 2 and 7.

We next assess whether these differences in transfer functions are significant between group average perceived arousal profiles. To do this we retain the same model specification for all three models (that is \( L = 10 \) and \( P = 3 \)) and test the null hypothesis that \( \omega_k,l \equiv \omega_l \) for \( k = 1, 2, 3 \) versus the alternative that they differed between groups. For the first method, vector autoregression analyses (VAR) were done in which the group average series were used as the endogenous variables and lags of intensity as the exogenous variables. The group average series were constrained in the VAR model so as not to influence the other group series by using a diagonal VAR structure. The likelihood ratio test \( (G^2 = 52.7) \) rejected the null hypothesis \((p < 0.0001)\), and hence the three participant groups were not identical in their response to intensity. Thus, we made similar intensity coefficient-constrained vs. unconstrained comparisons between the three pairs of participant groups. EA and M differed (likelihood ratio test \( p = 0.04 \)), as did NM and M \((p < 0.0001)\), whereas EA and NM were not distinct \((p = 0.08)\). In contrast, the shape of the transfer functions (Figure 5) suggests greater similarity between EA and M. It may be that the transfer function for the NM group is that pertaining to intensity profiles in environmental or everyday sound (largely “bottom up”: see discussion), while that for the two musician groups is influenced more by musical knowledge (somewhat more “top down”).

5. Analysis of individual responses

To account for inter-participant differences, an analysis of individual series as an ensemble (as in panel or longitudinal data) was undertaken. Two approaches to analysing the panel of 32 time series were considered: the first, presented in more detail, uses fixed effects analysis where each individual series is fit with its own parameter values for model terms treated as fixed effects; the
second is a random effects specification (model terms use random effects to explain differences between individuals).

5.1. Fixed effects type analysis

An initial VAR included all individual time series as endogenous variables, constrained so as not to influence each other (as in Section 4), retaining all the significant lags of intensity. For each group, the null hypothesis that all individuals share the same coefficients, was rejected ($p < 0.0001$). So there is overwhelming evidence that the individuals differ significantly in their model parameters and hence in their arousal responses to changes in intensity; in some analyses these might overwhelm group differences. Thus, possible clear cut differences between each pair of groups were investigated further here by the use of individual ARMAX fits to each of the 32 series and in Section 5.3 by cross-sectional time series analyses (CSTSA).

The model specified by equations (4)–(6) was fit to each series yielding 32 sets of individual model parameter estimates $\hat{\omega}_{i,j,l}$, $\hat{\phi}_{i,j,l}$ (Figure 6). These were then compared with parameter estimates (or corresponding log-likelihoods) for the single model fit to all series with the parameters constrained to be the same across all series. There is no ready-made software for this purpose.

![Figure 6](image_url)

Figure 6. Parameters of transfer function model (4)–(6) fit to 32 individual listener arousal responses. Left panel (a): Transfer function coefficients $L = 10$. Right panel (b): Autoregressive coefficients $P = 3$. Red solid lines: transfer function significant (Wald test) at 5% level; black dashed lines: not significant; the thick solid turquoise line is the constrained fit to 32 series; the thick solid blue line gives parameters estimates for the model for overall averaged series from Figure 5.
We utilized the ARMAX function from the time series analysis library in R to compute the individual model fits and the constrained fit in which the parameters were the same for all series. The constrained optimization required programming using the built-in `optim` function. The optimization was initialised with the parameters from the model for the overall average series reported above (Figure 6, thick solid blue line). The converged result for the fit is the thick solid turquoise line; the main difference between average series and constrained fits is the lag 2 autoregressive coefficient in panel (b); the similarity of transfer function coefficient estimates is reassuring. The likelihood ratio statistic for testing that all 32 series have identical values for the transfer function and autoregressive components of the common model was also significant ($p < 0.0001$) confirming the VAR analysis above.

We next re-examine the question of intergroup differences (cf. Section 4). The individual parameter estimates $\hat{\omega}_{ij} = (\hat{\omega}_{i,j,1}, \ldots, \hat{\omega}_{i,j,L})^T$, $\hat{\phi}_{ij} = (\hat{\phi}_{i,j,1}, \ldots, \hat{\phi}_{i,j,P})^T$ are independent multivariate observations. Multivariate analyses of variance with respect to group did not reject the null hypotheses that the vector of transfer function and autoregressive parameters $(\omega_{ij}, \phi_{ij})$ are the same for $j = 1, 2, 3$ (denoting groups EA, M, and NM) ($p = 0.85$) or the hypothesis that compared a common mean for the EA and NM groups with that of the M group ($p = 0.35$). When focusing on the autoregressive coefficient vectors alone the same conclusion holds. The difference between the transfer function mean coefficient vector for the EA and NM groups combined compared with that of the M group was not significant ($p = 0.19$) and hence this analysis did not distinguish the mean response transfer function for the M group relative to the EA and NM combined group – unlike that obtained in Section 4.

Thus, when inter-respondent variability is accounted for in this way, there are no longer significant differences between the average transfer function responses for the three groups. We next characterize inter-individual differences in arousal response.

### 5.2. Characterising the differences between individuals in their arousal responses to the Wishart piece

A direct measure of responsiveness (not dependent on a fitted model) is the percentage of times of stasis in response. As noted above (Section 3), this is not associated with group or OMSI, but it is related to measures of model fit and form (below).

Considering fit, using a Wald test, the ensemble of $L = 10$ transfer function coefficients is significant. Using the alternative representation, equations (7) and (5), we also assessed the percentage of variability in response due to noise $P_e$, the fitted model given by $P_M$, and the proportion of that associated with the immediate action of the transfer function $P_W$ – see Table 1. We also considered the degree of autoregression required (which ranged over $P = 0, 1, 2, 3$ for 1, 11, 5, 15 series respectively) and the estimated variance of $e_t$. Considering form, we assessed the lag $L_{\text{max}}$ given by equation (9) at which the maximum of the transfer function was attained, and the sum of the transfer function coefficients, $\omega^*$, as given by equation (8). None of these measures of fit and form was significantly associated with group or OMSI.

However, these summary measures are typically highly correlated with the measure of stasis. Stasis is: strongly positively correlated with the logit of the Wald $p$-value for testing significance of the transfer function ($r = 0.56, p = 0.001$); negatively correlated with $\omega^*$ ($r = -0.56, p = 0.021$) suggesting increasing stasis weakens the overall transfer function response; strongly positively correlated with $L_{\text{max}}$ ($r = 0.60, p = 0.000$) suggesting the more stasis the higher the lag at which the transfer function is maximised, and a slower response, not associated with proportion of variability $P_e$ that is unexplained ($r = -0.25, p = 0.160$). These results suggest that listeners who respond frequently have a more significant and stronger transfer function structure (as might be expected given that stasis decreases the quantity of data available).
Figure 7 gives boxplots of the two measures of transfer function features $\omega^*$ and $l_{\text{max}}$ split by significance of the overall transfer function (at the 5% level of the Wald statistic) or by whether stasis is above 50% (19 cases) or below 50% (13 cases). For instance, for the 17 individuals whose transfer function was significant, the lag at which it peaked tended to be shorter (median value of $L_{\text{max}} = 2$) than that of the 15 individuals whose transfer function was not significant (median value of $L_{\text{max}} = 5$). Thus, those with statistically significant responses responded more rapidly than those without. The correlation between $l_{\text{max}}$ and the logit of the Wald $p$-value is significantly positive ($r = 0.44, p = 0.01$). Peak response lags for the 17 significant responders were distributed as follows.

| $l_{\text{max}}$ | 1 | 2 | 5 | 6 | 7 |
|------------------|---|---|---|---|---|
| Number of Listeners | 2 | 8 | 3 | 1 | 3 |

Peak lags for the 15 non-significant responders were distributed more diffusely, as follows.

| $l_{\text{max}}$ | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 10 |
|------------------|---|---|---|---|---|---|---|----|
| Number of Listeners | 2 | 1 | 3 | 2 | 3 | 2 | 1 | 1 |

5.3. Cross sectional time series analyses of differences between groups and individuals

In the previous two sections, each of the 32 individual response series was modelled using its own set of transfer function and autoregressive coefficients. An alternative approach, which allows a direct estimate of the relative contribution of inter-group versus inter-individual differences in perception of arousal to be obtained, is pooled cross-sectional time series analysis (CSTSA) with random effects. In this approach, multiple individual time series, which describe different “units” (groups or individual participants) responding to the same conditions (i.e. to the Wishart piece) are analysed together, as a panel of data. We further introduce and fully define this method in the accompanying paper.

We briefly summarise here its implications for the Wishart piece. The approach showed that between group effects (i.e. between the averaged series) are significant, but only contribute 0.05–0.08 of the modelled variance, while overall models have $R^2$ values of between 0.53 and 0.72. These numbers come from STATA's xtreg, xtregar, xtggee and related procedures, which generally allow either no or low order autoregression. These $R^2$ values are not identical in interpretation to standard $R^2$s, but they are an index of variance explained by the models. A range of related random effects models was computed, with generalized least squares (GLS), maximum likelihood, generalised method of moments (GMM), and generalized estimating equations (GEE) approaches, generating quite similar coefficients for the impacts of intensity lags, as expected. These data again suggest small inter-group differences, and larger inter-individual differences. They also support the interpretations of the transfer function in our grand average model, that acoustic intensity is a powerful predictor. In the accompanying paper we show how another STATA procedure, xtmixed, can be used to bring lags of the exogenous predictors and chosen orders of autoregression into play, and can deal simultaneously with all individual series identified to individuals and to groups; with confirmatory and far more nuanced results. Overall, we conclude for the Wishart piece that inter-individual differences in perception of arousal are dominant, but there remain subtle inter-group differences.
6. Binary analyses of individual response time series: a solution to the “perceived arousal stasis” conundrum

In Section 5 we presented results from modelling individual responses using the standard ARMAX time series modelling approach. However, Figure 3 shows that listeners vary substantially in the amount of time that their responses are in stasis: from 15% to 90%. Such a skew distribution of responses may challenge conventional time series analysis.

In the Online Supplement to this article, we assess this question. We develop an approach based on component series and binary time series modelling. This binary model considers solely whether an arousal response is positive (the value increases), or alternatively is static or negative. It is thus highly conservative, and preserves only a small part of the information contained in the original continuous data series. The resultant alternative model explicitly accounts for the highly discrete nature of the individual response series during periods of stasis. We consider the
results from this alternative approach with the four selected series just discussed, and the results from applying it to all 32 individual response series. We use the binary model to re-examine the question of differences between groups and individuals and whether any differences can be explained by other factors, as we did in Section 5. All the methodological details and model formulations are provided in the Online Supplement, and here we very briefly summarise the key conclusion.

We find significant differences between individual listeners’ propensities to respond with a positive change in arousal to changes in acoustic intensity, but that this is not itself due to differences in average or overall group responsiveness. Consequently, we go on to compare the participant groups in more detail. It is only in these comparisons between listener groups that the discrete binomial analyses are more equivocal than the continuous ARMAX analyses, and the strong possibility remains that this is because of the reduction in data usage in the binomial approach. The Online Supplement binomial analyses thus support the ARMAX analyses, in spite of the caveats raised, strongly confirming the variation between individuals in responsiveness to acoustic intensity changes and the robustness of ARMAX under the circumstances studied here. Consequently, ARMAX and CSTSA methods are also adopted in the companion paper.

7. Time-varying features of the model for average arousal in Red Bird

Here we investigate the stationarity of the transfer function model (4)–(6) fit to the average perceived arousal response series $\bar{Y}_t$ defined by (2). We find strong evidence for time-varying transfer function model parameters and residual conditional variance.

7.1. Evidence for time-varying transfer functions

The influence of differenced intensity $\nabla X_t$ on arousal across the piece was assessed by rolling ARMAX determinations of the standard model of equations (4)–(6) (with AR degree $p = 3$) and $l = 1, \ldots, L, L = 10$ lags of differenced intensity) for the grand average arousal time series, using a window of 60 (corresponding to a musically plausible period of 30 seconds) and advancing in unit time steps (Figure 8). Panel (a) shows that participants vary the lag at which the transfer function is maximised. The transfer function weights at typical maximal lags are shown in sub figures (b) to (d). Associated with the varying lag response is the change in autoregressive dependence. This suggests that the extent to which listeners emphasise change in intensity relative to their inertia from previous responses varies throughout, as does the estimated variance of the residual noise process $e_t$ in the autoregression Figure 8(f), consistent with the conditional variance analysis below.

We also divided the series of 393 values into four roughly equal segments and fit the parent model (4)–(6) to each segment. The likelihood ratio test that the model is the same in the four segments was ($G^2 = 127.6$ on 39 degrees of freedom), which is highly significant. Even if the dependence between non-overlapping segment fits were to be allowed for, the test would likely remain significant.

We conclude that the average response of participants to Wishart’s piece varies significantly through time, with both slow and abrupt changes. The change around time point 275 appears to relate to the ceasing of animate sounds there (see next section), but other relationships are not yet apparent.
Figure 8. Time-varying estimates of key model features using a rolling analysis with a step of 1 time unit and window length of 60. (a) \( l_{\text{max}} \), lag at which maximum of the transfer function weights occurs. (b),(c),(d) \( \omega_l \), transfer function at lag \( l = 1, 2, 8 \) respectively. (e) \( \phi_1 \) lag 1 autoregressive coefficient. (f) Estimate of noise standard deviation \( \hat{\sigma} \). Time count is at the sampling rate of 2 Hz.

### 7.2. Modelling the conditional variance of perceived arousal

We hypothesised that structural changes in *Red Bird* signalled by the presence or absence of human and animate sounds would impact on variability through time. In previous work we showed that spectral flatness modestly predicts ongoing perceived valence in the piece, and that this predictor could be replaced by animate sounds (Bailes and Dean 2012). Here we constructed two indicator covariates, \( A_t \) and \( H_t \), representing the absence or presence of animate and human sounds respectively (shown in Figure 9 with the fitted conditional variance of the innovation process, defined below). We modelled the residuals \( e_t \), in the transfer function model specified by equations (4)–(6) using a variety of multiplicative ARCH (autoregressive conditional heteroskedasticity) models including covariates \( A_t \) and \( H_t \), and intensity). ARCH is a process in which the variance is itself autoregressive and conditional. Thus, the variance might be conditioned by the input acoustical series. This is assessed by testing for the impact of the exogenous variable upon the overall model fit, both with and without a generalized ARCH model.

In all models tested, \( A_t \) and \( H_t \) were not significant in the transfer function but, together with lag 2 of differenced intensity, they were significant in the GARCH specification. BIC selected the best model for the mean average arousal response series of the form of equation (4), as before with \( L = 10 \), and equation (6), with a reduced \( P = 1 \) order, with the additional specification that the residuals follow the multiplicative GARCH(1,1) model where \( e_t \) are independent normal
random variables with variance $\sigma_t^2$ conditional on the past given by

$$\sigma_t^2 = \exp(\lambda_0 + \lambda_1 \nabla X_{t-2} + \lambda_2 H_t + \lambda_3 A_t) + \alpha_1 e_{t-1}^2 + \beta_1 \sigma_{t-1}^2.$$  \hspace{2cm} (10)

Higher orders than $(1, 1)$ were not significant in the GARCH or EGARCH (Exponential GARCH) specifications. The fitted transfer function coefficients were all significant and gave a similar shape to that in Figure 5 but with values for the first six lags reduced by around 20% relative to the previous fit without GARCH. The lag with maximum transfer function remained at $l_{\text{max}} = 2$. Estimated values of the GARCH model (s.e. in parentheses) are: $\hat{\lambda}_0 = -2.25(0.39)$, $\hat{\alpha} = 0.169(0.065)$, $\hat{\beta} = 0.613(0.075)$, $\hat{\lambda}_1 = 0.184(0.024)$, $\hat{\lambda}_2 = -0.641(0.224)$, and $\hat{\lambda}_3 = 0.670(0.234)$. Thus, human sounds decrease conditional variance while animate sounds increase it.

8. Conclusions on the analysis of Red Bird and on the statistical methodology

Because we proceed in the companion paper to apply the methods of this paper to a full comparison of the four pieces studied in our project (Dean, Bailes, and Dunsmuir 2014), we restrict ourselves here to brief comments in conclusion.

Overall, there were statistical differences between the participant groups, but they were less coherent and graded than might have been expected. It would be valuable to increase participant
number for the EA group, but obtaining the in-laboratory participation of more expert EA musicians even in a large city is not feasible. We went on to more detailed consideration of individual differences, since the bigger these are within each group, the more likely it is that they will submerge group differences. There was significant variation in performance within each group, and these differences were larger than the inter-group differences. The NM group was generally more homogeneous, and the M group the least. These differences were expressed in the lag structure of the intensity transfer function, and also in the frequency of stasis in response, although the latter was not predicted by OMSI. Nevertheless, the OMSI was more powerful than group membership in some predictions. It seems that familiarity may in some contexts influence perceptual performance as strongly as the development of explicit musical expertise. Exposure, for example, may well mould bottom-up performance alongside the acquisition of explicit knowledge, which can operate top down. Indeed, novices and experts alike rapidly adjust their responses as they hear a piece.

8.1. Statistical issues

This paper has used conventional time series analysis with some success in revealing the predictive power of the acoustic variable intensity, particularly in analyses at the expertise group and overall participants' level. However, these analyses also revealed a potentially troubling dominance of periods of stasis. A large probability density at the zero response measure may violate assumptions of continuous or even Gaussian distributions typically used in time series methods. Thus, we have developed new generalised linear ARMA methods (see Online Supplement to this paper) extended to binomial analysis to simplify and overcome this problem. By treating the series as probabilistic determinations of positive, or zero/negative change, we have been able to largely confirm the interpretations of the conventional ARMAX/VAR analyses at the level of individual participants. The methods we have developed may be useful in other cases and situations of response series in psychology or physiology that share this “zero mass” issue.

The statistical armoury at our disposal seems adequate to allow us to understand real-time music perception (and performance). But there remain vast unanswered questions. In the companion paper, we apply it to the full comparison of the four pieces under study in our project. The statistical tools evaluated here can also be useful in a range of psychophysiological studies, such as those with skin conductance, or EEG and fMRI, particularly in those cases where it is not possible to determine in advance the kinetics of the relevant stimuli that might constitute exogenous variables. For example, we are interested in monitoring psychophysiological features of musical performers, notably improvisers, while they perform: we cannot dictate the precise timing of musical events in these circumstances. It is then not possible to use, for example, impulse convolution of the intended experimental stimulus with a putative archetypal psychophysiological response (such as the BOLD response in fMRI, or the “event related” skin conductance response kinetic). Statistical techniques such as those discussed here will allow us to understand mutual interactions of cognition, performance and affect.

Acknowledgements

The authors would like to thank editor Thomas Fiore for his assiduous and perceptive work on our manuscripts.

Supplemental online material

The online supplement for this article can be accessed at http://dx.doi.org/10.1080/17459737.2014.928752.
References

Bailes, F., and R. T. Dean. 2009a. “Empirical Studies of Computer Sound.” In The Oxford Handbook of Computer Music, edited by Roger T. Dean, 473–492. New York: Oxford University Press.

Bailes, Freya, and Roger T. Dean. 2009b. “Listeners Discriminate Affective Variation in Computer-Generated Musical Sounds.” Perception 38 (9): 1386–1404. http://www.perceptionweb.com/abstract.cgi?id=p6063.

Bailes, Freya, and Roger T. Dean. 2012. “Comparative Time Series Analysis of Perceptual Responses to Electroacoustic Music.” Music Perception: An Interdisciplinary Journal 29 (4): 359–375. http://www.jstor.org/stable/10.1525/mp.2012.29.4.359.

Balkwill, Laura-Lee, and William Forde Thompson. 1999. “A Cross-Cultural Investigation of the Perception of Emotion in Music: Psychophysical and Cultural Cues.” Music Perception: An Interdisciplinary Journal 17 (1): 43–64. http://www.jstor.org/stable/40285811.

Beran, J., and G. Mazzola. 1999. “Analyzing Musical Structure and Performance—A Statistical Approach.” Statistical Science 14 (1): 47–79.

Box, G. E. P., G. M. Jenkins, and G. C. Reinsel. 2008. Time Series Analysis: Forecasting and Control, 4th Edition. Wiley, New York.

Brown, J. C. 1993. “Determination of the Meter of Musical Scores by Autocorrelation.” The Journal of the Acoustical Society of America 94 (4): 1953–1957.

Dean, Roger T., and Freya Bailes. 2010. “Time Series Analysis as a Method to Examine Acoustical Influences on Real-Time Perception in Music.” Empirical Musicology Review 5 (4): 152–175. http://hdl.handle.net/1811/48550.

Dean, Roger T., and Freya Bailes. 2011. “Modelling Perception of Structure and Affect in Music: Spectral Centroid and Wishart’s ‘Red Bird’.” Empirical Musicology Review 6 (2): 131–137. http://hdl.handle.net/1811/51217.

Dean, Roger T., Freya Bailes, and William T. M. Dunsmuir. 2014. “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. http://dx.doi.org/10.1080/17459737.2014.928753.

Dean, Roger T., Freya Bailes, and Emery Schubert. 2011. “Acoustic Intensity Causes Perceived Changes in Arousal Levels in Music: An Experimental Investigation.” PLoS ONE 6 (4): e18591. http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0018591.

Dubnov, Shlomo, Stephen McAdams, and Roger Reynolds. 2006. “Structural and Affective Aspects of Music from Statistical Audio Signal Analysis.” Journal of the American Society for Information Science and Technology 57 (11): 1526–1536.

Ferguson, Sam, Emery Schubert, and Roger T. Dean. 2011. “Continuous Subjective Loudness Responses to Reversals and Inversions of a Sound Recording of an Orchestral Excerpt.” Musicae Scientiae 15 (3): 387–401. http://msx.sagepub.com/content/15/3/387.abstract.

Greene, W. H. 2008. Econometric Analysis. 6th ed. Prentice Hall, New Jersey.

Gregory, D. 1989. “Using Computers to Measure Continuous Music Responses.” Psychomusicology: A Journal of Research in Music Cognition 8 (2): 127–134.

Huron, D. B. 2006. Sweet anticipation: Music and the psychology of expectation. Cambridge, MA: The MIT Press.

McAdams, Stephen, Bradley W. Vines, Sandrine Vieillard, Bennett K. Smith, and Roger Reynolds. 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. E. 2006. “A Criterion-Related Validity Test of Selected Indicators of Musical Sophistication Using Expert Ratings.” PhD thesis. The Ohio State University.

R Development Core Team. 2011. R: A Language and Environment for Statistical Computing, Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0. http://www.R-project.org/.

Ramsay, J. O., and B. W. Silverman. 2005. Functional Data Analysis, 2nd Edition. Springer Verlag.

Russell, James A. 1980. “A Circumplex Model of Affect.” Journal of Personality and Social Psychology 39 (6): 1161–1178.

Russell, James A. 2003. “Core Affect and the Psychological Construction of Emotion.” Psychological Review 110 (1): 145–172.

Salimpoor, V. N., M. Benovoy, G. Longo, J. R. Cooperstock, and R. J. Zatorre. 2009. “The Rewarding Aspects of Music Listening are Related to Degree of Emotional Arousal.” PloS ONE 4 (10): e7487.

Schubert, Emery. 2004. “Modeling Perceived Emotion With Continuous Musical Features.” Music Perception: An Interdisciplinary Journal 21 (4): 561–585. http://www.jstor.org/stable/10.1525/mp.2004.21.4.561.

StataCorp. 2009. Stata: Release 11 Statistical Software. College Station, TX: StataCorp.