Time Series Analysis as a Method to Examine Acoustical Influences on Real-time Perception of Music

ROGER T. DEAN
MARCS Auditory Laboratories, University of Western Sydney

FREYA BAILES
MARCS Auditory Laboratories, University of Western Sydney

ABSTRACT: Multivariate analyses of dynamic correlations between continuous acoustic properties (intensity and spectral flatness) and real-time listener perceptions of change and expressed affect (arousal and valence) in music are developed, by an extensive application of autoregressive Time Series Analysis (TSA). TSA offers a large suite of techniques for modeling autocorrelated time series, such as constitute both music's acoustic properties and its perceptual impacts. A logical analysis sequence from autoregressive integrated moving average regression with exogenous variables (ARIMAX), to vector autoregression (VAR) is established. Information criteria discriminate amongst models, and Granger Causality indicates whether a correlation might be a causal one. A 3 min electroacoustic extract from Wishart's *Red Bird* is studied. It contains digitally generated and transformed sounds, and animate sounds, and our approach also permits an analysis of their impulse action on the temporal evolution and the variance in the perceptual time series. Intensity influences perceptions of change and expressed arousal substantially. Spectral flatness influences valence, while animate sounds influence the valence response and its variance. This TSA approach is applicable to a wide range of questions concerning acoustic-perceptual relationships in music.

Submitted 2010 June 16; accepted 2010 July 27.

KEYWORDS: *perception, affect, acoustics, autoregressive time series analysis*

HOW do acoustical properties of musical works influence their perception? This question has attracted research with a focus on acoustic/perceptual features that include: sound intensity/loudness (Chapin, Large, Jantzen, Kelso, & Steinberg, 2008; Olsen, Stevens, & Tardieu, 2007); frequency/pitch and tonality (Kessler, Hansen, & Shepard, 1984; Windsor, 1997); sound duration/meter and tempo (Boltz, 1998; Chapin, et al., 2008; Quinn & Watt, 2006; Todd, Cousins, & Lee, 2007; Vuust, Ostergaard, Pallesen, Bailey, & Roepstorff, 2009); and spectrum/timbre (Bailes & Dean, 2007; Caclin, McAdams, Smith, & Winsberg, 2005; Gordon & Grey, 1978), as these relate to perceived structure and emotional response (Bailes & Dean, 2009; Bigand, Vieillard, Madurell, Marozeau, & Dacquet, 2005; Dubnov, McAdams, & Reynolds, 2006; Gabrielsson & Lindstrom, 2001; Leman, Vermeulen, De Voogdt, Moelants, & Lesaffre, 2005; Schubert, 2004; Sloboda, 1991; Sloboda & Lehmann, 2001). An issue in this work is the inadequacy of treating music as a static entity, since perception is determined by the temporal organization of these acoustic properties (Brittin & Duke, 1997; Deliège, Mélen, Stammers, & Cross, 1996; Krumhansl & Kessler, 1982; Lalitte & Bigand, 2006; McAdams, Vines, Vieillard, Smith, & Reynolds, 2004; Schubert, 2004; Sloboda & Lehmann, 2001). This paper will detail an analytical method allowing for just such a dynamic examination of the relationship between acoustical properties of a composition with real-time listener perceptions of its structure and its affective content.

Previous studies of listeners' real-time perceptions of affect in music have attempted to map response through time to acoustic properties of the piece (e.g. Schubert, 2004). Missing are substantial attempts to assess which acoustic properties also drive listeners' perceptions of the structure of the same

music. Structure in this instance need not be a music-theoretic analysis of large-scale form (such as sonata form in classical music), but refers to the low-level assessment by a listener of change and continuity in the music. While either the relationship of acoustic properties to perceived affect or the relationship of acoustic properties to perceived structure would be informative in its own right, an amalgamation of both provides a more complete psychological account of ways in which acoustic properties implicitly and explicitly shape music perception. For instance, it may be that listener ratings of affective change more closely align with their ratings of perceived structure than with acoustic measurements of the music (see Bailes & Dean, 2009). This might most obviously be the case in tonal compositions where it is difficult to capture the hierarchical tonal relations through acoustic measures alone. Conversely, musical forms that do not rely on hierarchical structures such as tonality or meter might exhibit quite a close relationship between acoustic properties of the work, listener perceptions of structure (change in sound), and listener perceptions of affect. Electroacoustic music is one such form, and the subject of the current paper.

The electroacoustic composition *Red Bird*, by Trevor Wishart (1977), was selected for study (see T. Wishart, 1985; Trevor Wishart, 2009 for discussion of his compositional techniques) to demonstrate the widest possible utility of the analytical approach, both to instrumental and electroacoustic music, familiar and unfamiliar. The piece does not comprise distinct 'note' events like those in a piece of classical piano music. Rather, the work is based on transformations in timbre, texture, and loudness throughout its 45 min duration and it includes identifiable sounds from animate objects including birds, which also undergo transformation. The piece can be viewed as highly symbolic, and in diverse ways. Given that almost all research into the acoustic correlates of perceived emotion in music uses Western tonal music, an original motivation for this paper was to examine how listeners perceive the emotion expressed by non-tonal, alternative musical compositions (Bailes & Dean, 2009). We particularly chose a section of the piece which concentrates the animate (bird and other) sounds in juxtaposition with more widespread electroacoustic timbral gestures, such that we could illustrate the power of TSA to analyze the impact of specific (sometimes unusual or unique) timbral features, as well as universal ones such as intensity and spectral features. Although no note structure is present, other sound features, which have been suggested to relate well to the perception of affect in most music, are fundamental to this and much other electroacoustic music from a tradition that now spans more than 50 years (Dean, 2009). For instance, physical sound intensity can be measured, and past work points to a link between its perceptual counterpart of loudness and the real-time perception of tension (Krumhansl, 1996) and arousal (Schubert, 2001).

Past research also points to timbre as an influence on the perception of affect (Bailes & Dean, 2009; Leman, et al., 2005) but this has been investigated much less. Spectral flatness, which quantifies the distribution of partials in the signal, is a 'global' parameter of timbre: it is influenced by every spectral component, symmetric or otherwise. It is measured as geometric mean/arithmetic mean of the power spectrum. High values indicate noise, and low values suggest peaks in the spectrum (an infinitely narrow peak has a spectral flatness of minus infinity). It is one of the four 'basic spectral audio Descriptors' used in the MPEG-7 standard (for an overview: ("MPEG-7 Overview," 2004)). There it is termed the 'audioSpectrumFlatness Descriptor' and it is noted that low values are informative as they can 'signal the presence of tonal components'. Unlike spectral centroid, another key summary MPEG-7 Descriptor, which is an indicator of the central frequency range of the spectrum, spectral flatness is influenced even by symmetrical changes in the power spectrum. Our purpose here was to establish a method that can be applied both to electroacoustic music, which is often primarily timbral, and often without regular rhythmic or pitch structures, and equally to tonal Western music and to music of other cultures which emphasizes hierarchical pitch structures (Krumhansl, 1990). As an initial primary measure of frequency spectrum, spectral flatness was therefore our approach of choice. Measures specific to any individual music style could be added or substituted in subsequent work (e.g. equal tempered pitch measures for Western classical music). Spectral flatness has the added advantage that as a 'global' parameter of timbre, it is directly related mathematically to one measure of the information content and information rate of a sonic stream (Dubnov, 2006). Thus we use measurements of spectral flatness as our parameter of timbre, so that timbre can be studied from other specific points of view thereafter.

In addition to intensity and spectral flatness variations, *Red Bird* features many sounds reminiscent of everyday, environmental sounds. Indeed, the piece can be construed as highly narrative, with a combination of animate (human and other) and inanimate sound sources evoked. Animate sounds can be readily interpreted as possessing agency (as discussed previously by Maus (1997) with reference to music). So it also makes sense that sounds perceived as animate are associated with greater affective expression than inanimate sound sources (Bailes & Dean, 2009; Bradley & Lang, 2000). This study will

also examine whether the alternation of animate and inanimate sounds is a significant determinant of listener perceptions of structure and affect. In order to compare the structure of animate agency with physical measures of intensity and spectral flatness, and listener perceptions of structure and affect through time, an analytical method able to handle multiple time series is needed. This paper will develop and apply such a method based on TSA in detail, using *Red Bird* as a test case.

The techniques of autoregressive Time Series Analysis constitute a large and highly developed battery of methods specific to data which are not independent, and thus contravene the assumptions of most statistical approaches. Such data are often highly autocorrelated, such as series of real-time data, so called 'time series' (Enders, 2004; Hamilton, 1994). Few analyses of real-time data about perception of sonic features or affective expression in music have applied these techniques (e.g. Brown, 1993; Vos, Van Dijk, & Schomaker, 1994, in both cases concerned with perception of meter in music) and they are not used in most of the interesting studies of such musical time series such as those by De Vries (1991), Madsen and Fredrickson (1993), Krumhansl (1996) and Dubnov et al. (2006). It is important to realize that if two quite independent time series datasets are each highly autocorrelated, they may well seem to be significantly cross-correlated. Thus to avoid identifying spurious relationships between series it is necessary to create 'stationarity', and deal with other issues of autocorrelation. What is required is so-called 'Weak Stationarity', which means in essence removing trends in the data, such that mean, variance and covariance are all 'unaffected by a change of time origin' (Enders, 2004, p. 53), i.e., they are constant within appropriate statistical limits. Throughout the rest of this paper we simply use 'stationary' to refer to such Weak Stationarity. Stationary series may still be autocorrelated. But a key criterion of subsequent satisfactory models of an individual series or of models for relating multiple stationary series is that the residual errors (the time series constituting each successive residue left when the model estimate of a point is subtracted from the corresponding data point) are white noise, and thus no longer autocorrelated. It is only when this criterion is fulfilled that many of the statistical tests of significance which are routinely applied are meaningful. In addition, if there were still autocorrelation in the residuals from the model, this would represent unmodelled information. Note that there are other techniques applicable to the analysis of time series, both in the time and spectral domains, but in this paper we will refer specifically to autoregressive Time Series Analysis simply as TSA. There is some further consideration of other techniques in the Discussion section of this paper.

Amongst the few significant earlier applications of some aspects of TSA to this area is the pioneering work of Schubert (Schubert, 1999, 2001, 2004; Schubert & Dunsmuir, 1999), which indicated that with several pieces of classical music, acoustic intensity might be a major influence on perceived arousal. These studies have used a common feature of reliable TSA, the process of differencing, which means calculating the difference between successive values, thus creating a new series with one fewer point than the original. A differenced linear time series without error has a constant value, and differencing a time series which has a trend and substantial variability produces short runs of positive then negative values. As a result, such differenced series are commonly stationary even if the original series is not, and Schubert undertook some tests to establish the quality of his differenced series. There are alternative ways of achieving stationarity, but they were not required in the present work. In the present paper we apply a complete range of quality tests, and in particular, tests of stationarity prior to modeling. After differencing, Schubert then conducted simple forms of multivariate analysis by regressive/autoregressive techniques. We extend this approach considerably, using a wide range of the available plethora of techniques, and developing a logical and reproducible sequence for analysis. Here we develop the method and apply it to one electroacoustic work – in a detailed case study. Our choice of electroacoustic music complements the earlier work on Western classical music, and our approach is equally applicable to other forms of music.

MATERIALS

This paper focuses on a 3 min. sound extract of electroacoustic music by Trevor Wishart, from his *Red Bird* (1977). The section used is from c. 22'30 to 25'46" of this 45 min piece, and is taken from UbuWeb http:// www.ubu.com/sound/wishart.html (44.1kHz, 16bit, aiff, stereo). The piece was originally made by analogue techniques.

METHODS

Measuring Acoustic Intensity and Spectral Flatness

Acoustic intensity is measured across 500 ms windows, corresponding to those chosen for the subsequent perceptual analyses. Intensity is measured with respect to the frequency range 20-22050 Hz, and corresponds to Sound Pressure Level (SPL) in unweighted dB. Spectral flatness is measured as Wiener Entropy, using a slightly modified version of a script by Gabriel J.L.Beckers (2004; available online). Wiener Entropy is the ratio of a power spectrum's geometric mean and its arithmetic mean, as described above, but expressed on a log scale, which ranges from 0 (where the power spectrum is broad and relatively flat: 'noisy') to minus infinity (where the power spectrum is infinitely narrow: 'pure'). Frames of 500 ms are again used, with hop size 62.5 ms and a Gaussian window, and with the frequency range 0-22050 Hz. Frequency bins are c. 0.005 Hz. Note that changes in all regions of the frequency spectrum impact on this parameter.

Assessing Possible Statistical Cross-correlation between Spectral Flatness and Intensity

If intensity and spectral flatness are to be separable potential acoustic correlates of perceptual responses to a particular piece, then they should not themselves be closely collinear in that piece. A sound may certainly have its intensity adjusted without change to its spectral flatness, and vice versa. But when music develops acoustic intensity by increasing the number of individual pitches sounding at a particular moment (for example, the number of notes played simultaneously on an instrument such as the piano) or by increasing the spectral range within its timbres (in electroacoustic music), then a positive correlation between intensity and spectral flatness could result. This is assessed by determining the cross-correlation between the two time series. As mentioned already, cross-correlation between two time series that are themselves autocorrelated may be misleading. This issue is avoided by the process of 'pre-whitening': a standard TSA technique, a classic example of which is described in detail in Chapter 11 of Box et al. (1994). Pre-whitening comprises establishing a purely autoregressive statistical time series model of one series, in this case the spectral flatness series, so that the residuals from the model are white noise, i.e. free of autocorrelation (tested with Bartlett's periodogram-based test (Bartlett, 1966)). This standard process of modeling is detailed below in section 3.4. The resultant autoregressive model of spectral flatness, in terms only of its autoregressive lag structure and coefficients, is then used to model the second time series, in this case that of intensity, generating a further time series of residuals from the intensity profile. This second residual time series may or may not be free of autocorrelation, depending on how similar the intensity autocorrelation structure is to that of spectral flatness. Then the cross-correlation between the pair of residual series can be assessed meaningfully. By removing the autocorrelation within the first series, and any similar components in the second, pre-whitening allows a valid assessment of the cross-correlation between the two parent series. Technical details of the procedure are summarized clearly in McDowell (2002). Pre-whitening is not required for any of the further analyses because the autoregressive structure of the stationarized series under study is addressed directly in the modeling.

Perceptual Measurements of Change in Sound and Affect

The procedures for making perceptual measurements are described in detail in previous work in which participants responded continuously to short segments of constructed electroacoustic timbres (Bailes & Dean, 2009). In this study, our participants (N = 32, 16 female, median 25.5 years) were 16 'non-musicians', 8 classical musicians, and 8 experts in computer music. They were categorized on the basis of their Ollen Musical Sophistication Index (Ollen), together with information they provided about their experience with electroacoustic music. Non-musicians had OMSI scores <500, indicating 'less musically sophisticated'.

Listening over headphones, participants indicated their real-time perceptions of 'change' in the sound stream using a mouse scrubbing technique: they were instructed to move the mouse only if they perceived change in the sound, and move it faster for greater rates of change. Mouse movements were captured every 50 ms, and the rates of movement were averaged over 0.5 s windows. Such windows are appropriate given earlier studies (e.g. Schubert, 2004), which indicate that real-time perceptual responses

generally take at least 1 to 5 seconds for full registration. It is such delays (lags) that we study with TSA below. Real-time perception of expressed affect, that is explicitly the affect that listeners perceive to be associated with the sound (rather than induced), was measured in a separate experimental block using a '2D-emotion space' based on that established by Schubert and others, and written in Java. One axis displayed on a computer screen represents the perceived arousal of the sound (active – passive), and the other (at 90 degrees to it) represents the perceived valence (positive – negative). Cursor position within the 2D emotion space is recorded every 0.5 s.

Data series obtained from multiple participants were averaged to give a representative series for each perceptual parameter, which could be studied by time series analysis. This choice is commonly made, and was supported by the fact that when the time series were averaged by participant group subsets, the resultant series for each perceptual variable were similar across the three groups; the choice is considered further in the Discussion section.

Coefficient of variation (c.v. = standard deviation/mean) is used as an index of the variability of a time series. For the perceptual time series of arousal and valence, the measurement scale ranges from -100 to +100: for these series, c.v. is determined as a function of series constructed as the measured values plus 100, so that all series values are positive, and the c.v. then fairly reflects the degree of variability when taken in conjunction with the mean value. Any mean value displayed in the data below is of course that of the original series. When it is of interest to compare the relative impacts of different acoustic factors in a model of perceptual series, this is also done in such a way that the measurement is independent of the numerical ranges of the series under study (e.g. by fractional error variance distribution measurements in vector autoregression, see below).

Time Series Analysis of Correlations between Intensity, Spectral Flatness and Listener Perceptions: Methodological Approach

Since listeners cannot influence acoustic parameters, these are appropriately taken to be exogenous variables (independent, using the terms of empirical psychology), for the purpose of TSA, while the perceptual parameters are endogenous (dependent). The first step in the analysis (see Figure 1) is to remove outliers in the endogenous and the exogenous series, those values more than 2.5 standard deviations from the series mean, and to replace them by the nearest value within that range (i.e. the appropriate nearby value either + or – 2.5 s.d. from the mean). Generally no more than 5 values are adjusted (< 1.4% of the data points). Data under study in work such as this are often not normally distributed. Thus instead of removing outliers, one may alternatively use 'robust' statistics: the salient conclusions below have been confirmed by this alternative approach.

The next step in our approach is to obtain stationarized series (as summarized above). To this end, the autocorrelation and partial autocorrelation functions of the endogenous series are determined, and used to set the lag range for the Augmented Dickey-Fuller Generalized Least-Squares test for stationarity (Dickey & Fuller, 1979) using the Elliott, Rothenberg and Stock interpolated critical values (Elliott, Rothenberg, & Stock, 1996). Put simply, the test assesses whether the series value at a given time is a predictor of change to the next point: for a stationary series (constant mean), it should be a predictor with a negative coefficient, since larger than mean values tend to be followed by smaller ones, and vice versa so that in both cases the next value is closer to the mean. Conversely, for a non-stationary series, this expectation is not true. The 'augmented' part of the test allows for its autocorrelation structure, and the Elliott et al. critical values were empirically derived to enhance the power of the test. Endogenous series are differenced until stationarity is achieved. It should be noted that if a series is differenced, the result still bears a simple mathematical relation to its parent, thus a prediction from a model based on differenced variables can be converted back into a prediction of the parent variable, and the relationships will be qualitatively similar. Such conversions are not presented here, since they would not aid the interpretation of the results. After differencing, any further outliers are adjusted, but with the more stringent criterion of > 3 s.d. from the mean (generally, 0-2 values are adjusted). The resulting series may still show transient changes in variance, which are not sufficient to breach the overall criterion of weak stationarity. Such variance is described as heteroskedasticity (sometimes spelled heteroscedasticity). In the present work, additional variance stabilization techniques are not used since conditional heteroskedasticity, in which the variance of the ongoing data stream may be transiently affected by changes in the exogenous variable, is of interest (see below).

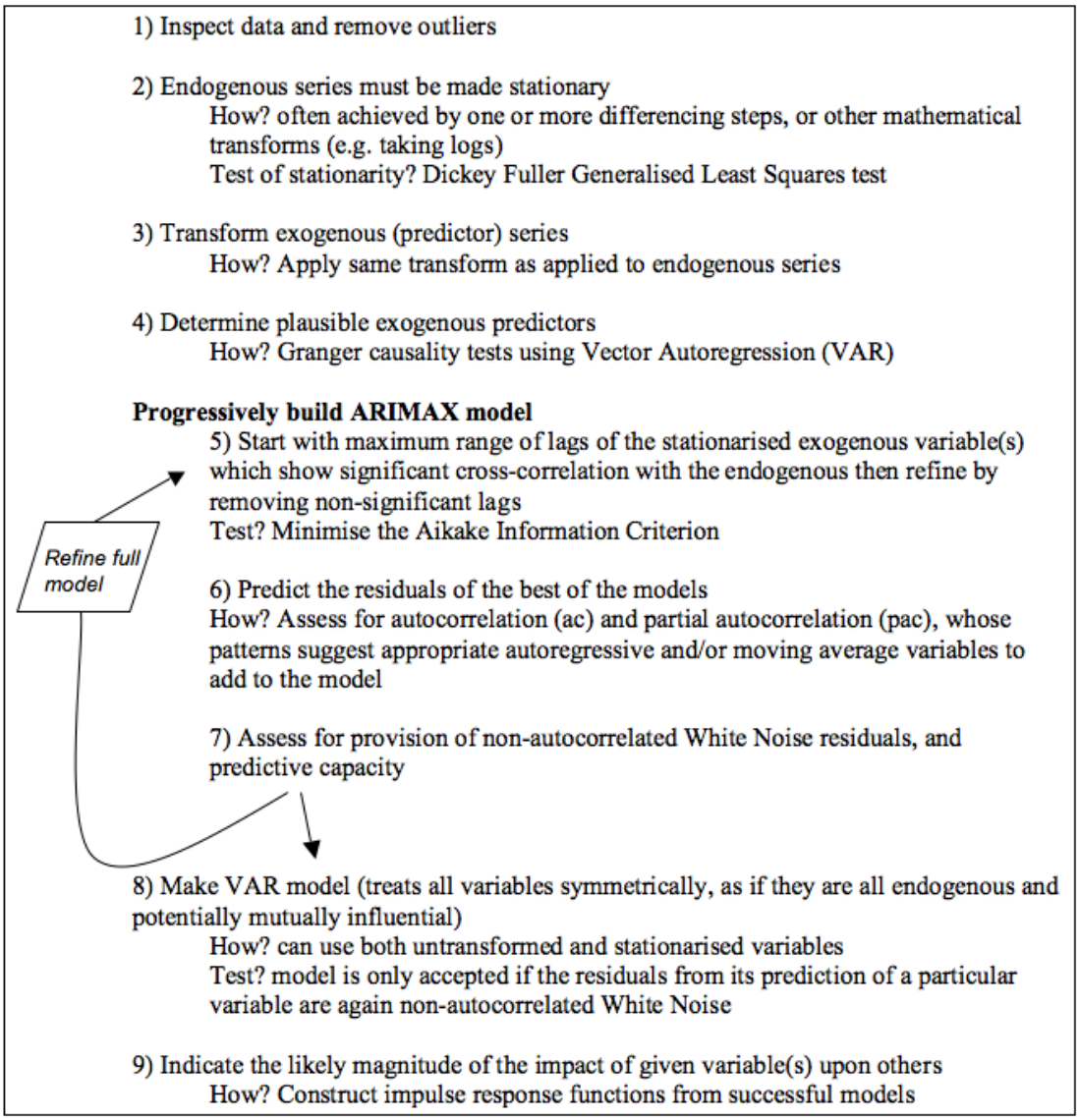

Fig. 1. Summary of the key stages of autoregressive Time Series Analysis (TSA), as discussed in the text.

For modeling the possible relationship between one endogenous and one exogenous variable, the exogenous time series is transformed by the same process as is used to make the endogenous series stationary (Enders, 2004). Granger Causality (an index of correlation between variables, which is used to assess the likelihood of a predictive relationship (Granger, 1969)) between the transformed variables is assessed with up to 8 lags (a lag being 0.5s) of the variables under consideration and using initially one acoustic variable together with the one perceptual variable. This is done using Vector Autoregression (see below), with the selection order criteria used to indicate the initial number of lags to study with the Akaike Information Criterion (AIC) (Akaike, 1974) as the determining factor, and treating variables temporarily and conservatively as if they are all endogenous, i.e. potentially mutually dependent; see below. The AIC is an estimate of the goodness of fit of a model, which uses the likelihood estimate for the model, together with the number of parameters it involves. Only relative values are informative, and lower values indicate a better model. We routinely determine the Bayesian Information Criterion also, and this penalizes for the number of parameters more strongly for large sample numbers than does the AIC. Since we have a large number of data points in every case, the BIC and AIC lead to the same interpretations with our data. The lag order may be increased if necessary to ensure that the residuals (errors) from the model for any

particular endogenous variable are White Noise (i.e. they have no remaining autocorrelation), as discussed further below.

Given Granger Causality of one or both acoustic time series upon the perceptual time series under study, we proceed to the corresponding bivariate modeling. The autoregressive integrated moving average procedure with an exogenous variable (ARIMAX) is used to model the relationship. The cross-correlation between the two series is used to determine the appropriate lags of the exogenous (acoustic) variable to model as predictors of the endogenous (perceptual) variable. The autocorrelation and partial autocorrelation of the now stationary endogenous variable is used to determine the autoregressive lags and moving average (MA) groupings to be included. First the model based on the exogenous component is optimized by the least squares fitting without modeling the autoregressive serial correlations in the residuals. Then an autoregressive and moving average model for those residuals is added (as appropriate), so that the overall model can be refit, often removing some of the exogenous variables, and now the remaining errors can be correctly calculated. Model refinement is throughout based on removing statistically insignificant terms, and testing for an improvement (minimization) in the Akaike Information Criterion (AIC) for the model. It is arbitrarily determined that not more than two lags of a variable which are not in themselves statistically significant may be included in the model if they improve the AIC and also provide a statistically significant likelihood ratio test in comparison with their nested parent model (i.e. these lags improve the predictive capacity of the model). Similarly, shorter lags are preferred over longer given that the AIC for the alternatives are similar. Note again that the AIC can only be meaningfully compared between ARIMA(X) models of the same series.

Given the best model from this process, residuals are predicted, and tested for lack of autocorrelation, and most importantly for the subsequent statistical significance tests, tested for White Noise character, as described above. A satisfactory model with White Noise residuals is then used to predict the time series of the endogenous (perceptual) variable, and its goodness of fit to the observed data is measured as the correlation between the two, as well as by testing for forecast bias and using Theil's U (model U2 of chapter 2 in Theil, 1966). Theil's U determines whether the model is better than a Naïve Method 1 forecast (which simply projects the last measured value forward, the so-called 'no change' approach). Theil's U is essentially the ratio of the root mean square error of the predicted series to that of the 'no-change' series, and hence values lower than one are indicative of a worthwhile model. Only models that meet these criteria are presented here. The ARIMAX models derived in this manner are compared with an ARIMA-only model of the same form (i.e. excluding the exogenous variable), and with ARIMA-only models that may include additional autoregressive lags, and improve the Akaike Information Criterion over the initial ARIMA model.

It is also sometimes of interest to determine whether the perceptual time series shows Autoregressive Conditional Heteroskedasticity (ARCH; where the variance is autoregressive and conditional) (Enders, 2004; Hamilton, 1994), and also whether the variance is conditioned by the input acoustical series (c.f. Enders, 2004, p. 141). This is done by testing for the impact of the exogenous variable upon the overall model fit, both with and without a Generalized ARCH model.

Following the ARIMAX analyses, Vector Autoregression (VAR) (Enders, 2004; Hamilton, 1994) analysis is undertaken (c.f. its earlier use simply to support a Granger Causality test to determine what further analyses should be made). Here the conservative choice is made that all variables are treated as potentially mutually influential (i.e. statistically endogenous). VAR with the acoustic time series treated as exogenous (VARX) is also undertaken for confirmatory purposes; these are not presented below, since they are uniformly in qualitative agreement with the interpretation from the VAR analyses with respect to the possible influence of the acoustic variables. Lag order selection statistics, information and likelihood criteria indicating the statistical adequacy of models of various lag number ('order'), are used to determine the order of the VAR to be performed, with the AIC again taken as primary.

VAR can be undertaken both with series that are and are not stationary, and statistical significance can be assigned providing the residuals from the model in question are white noise. This allows the use of the undifferenced series, and thus interpreting relationships between (for example), once differenced series can be avoided. Thus we first model VARs for combinations of variables transformed to stationarity during the ARIMAX analyses. However, after a VAR of an untransformed perceptual time series (but still with outliers removed) and with untransformed acoustic series, we also test for Granger Causality between the variables. We predict the VAR model for each individual transformed perceptual variable, but only with those acoustic variables that demonstrate Granger Causality upon it. The residuals from the predictions are tested both for White Noise character and lack of autocorrelation, and interpretation accepted if they meet

the criteria. VAR stability is assessed, and the Wald-lag exclusion criteria applied to determine the appropriateness of the inclusion of the modeled lags (Wald, 1955). The Wald-lag exclusion criterion tests whether the endogenous variables at a given lag are jointly zero for each equation and also jointly for all equations. Thus it indicates a judicious choice of lags to include in a model. Our approach is less conservative than one requiring the whole set of residuals from every component prediction of the VAR to be autocorrelation free and to comprise normally distributed disturbance. VAR results are displayed as Cholesky forecast-error variance decompositions (FEVD) (Lütkepohl, 2007), with standard errors obtained from bootstrapped residuals (so that no assumption need be made about their distribution). The FEVD is essentially an estimate of the proportional influence of each variable on the others, and it is projected through a series of lags (called steps in the Figures), as the mutual influences vary. The dominant influence on a highly autocorrelated variable at the earliest lags is of course the variable itself, while the effect of the other influences shows progressively thereafter. VARs are also undertaken with all acoustic and perceptual variables included together, since the perceptual variables may well influence each other.

At each relevant level of transformation of the variables (e.g. raw, once-differenced) a test for cointegration of each combination of variables is undertaken. Cointegration, simply described, occurs when two or more variables together form a linear combination that is itself stationary, even though the variables themselves are not. Cointegration can be modeled using Vector Error Correction (Engle & Granger, 1987), but it was not found in any of the series under study.

RESULTS

Acoustic Analyses

Figure 2 shows the spectral flatness and intensity temporal profiles of the Wishart extract. There are some clear parallels between the two series. Autocorrelation analysis of the intensity profile shows an autocorrelation close to 1 at lag 1 (corresponding to 0.5sec), declining smoothly by 13.5 seconds to a value indistinguishable from zero, using Bartlett's formula for 95% confidence bands. Correspondingly, the partial autocorrelation at lag 1 is close to 1, and lags 2-4 are significant, and just above the confidence band. The autocorrelation function for spectral flatness shows a slightly less smooth decline from .84 at lag 1 to insignificantly above zero by lag 24 (12 seconds), while the partial autocorrelations for lags 1-4 are again significant. Both series thus seem to be autoregressive with lag order 4 (which we abbreviate here and in the Tables as AR4).

An ARIMA model of spectral flatness reveals significant autoregressive lags 1-4 (confirming the AR4 characterization), and the residuals of this model are white noise (for a cumulative periodogram Bartlett's $B = 0.50$, Probability > $B = 0.967$). This model is used to filter the intensity time series so that the possible positive collinearity of the two acoustic series can be tested, as described in Methods. The raw series show a negative cross-correlation of -.83. Only limited significant and negative cross-correlation of the pre-whitened spectral flatness and filtered intensity residual time series remains at lags 0 and 1 (-.36 and -.38 respectively). Thus both intensity and spectral flatness remain separable potential influences upon listeners' perception of this music.

Perceptual Responses

Figure 3 shows the average perceptual response for sonic change, and Figure 4 shows those for perceived arousal and valence. The coefficients of variation of these perceptual response participant-averaged time series are 0.74, 0.33, and 0.19 for change, valence and arousal respectively. Both arousal and valence seem to be AR(2) processes, while change seems to be AR(4) (as judged again by the autocorrelation and partial autocorrelation functions: not shown). It can be seen that change and arousal correlate closely, while arousal and valence show a somewhat inverse relation. The relationships between these perceptual responses are assessed in greater depth as the analysis proceeds.

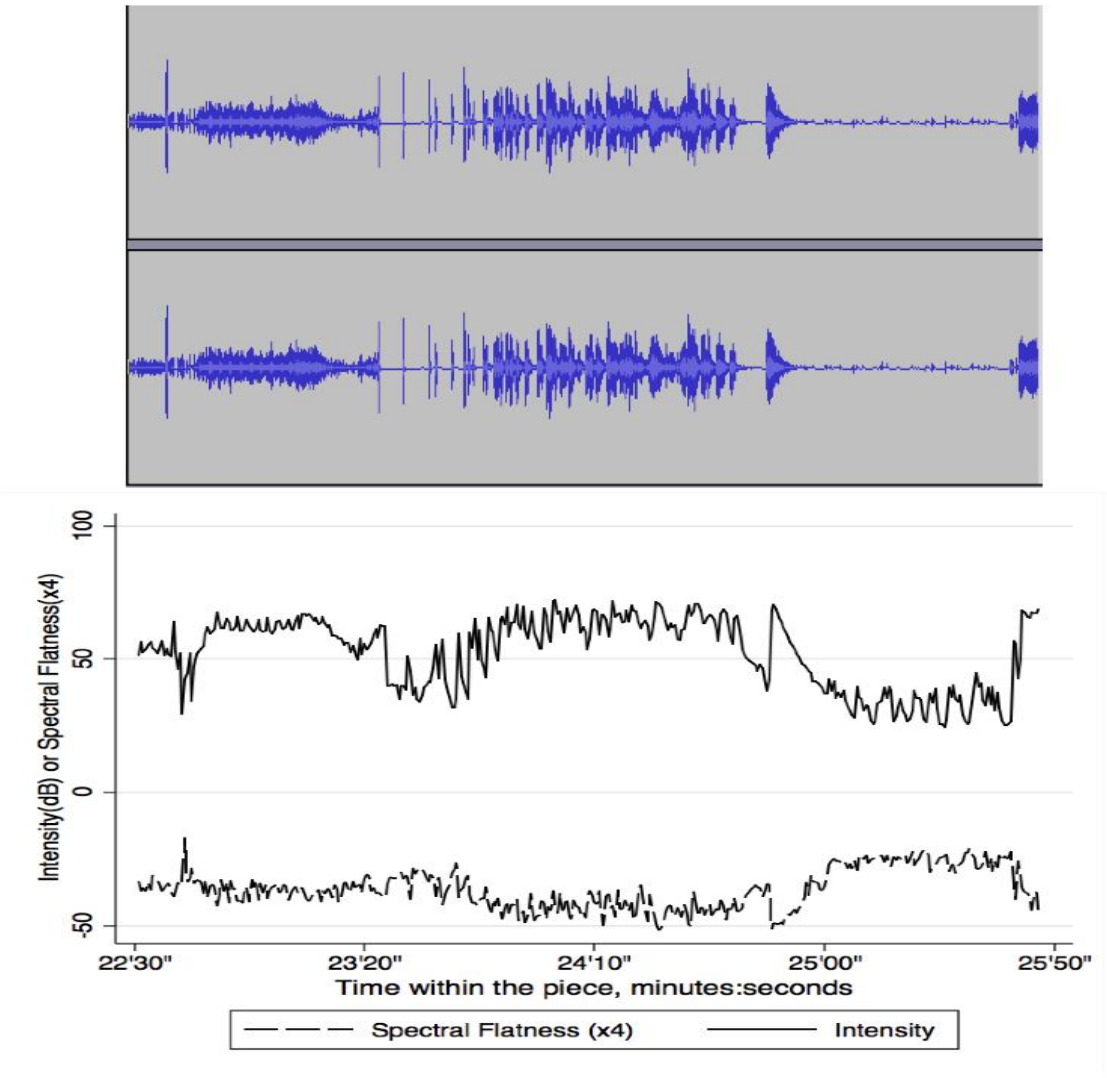

Fig. 2. Spectral flatness (Wiener Entropy) and intensity (dB) profiles of the Wishart extract. Spectral flatness is shown multiplied by 4, to make its features more apparent. The waveform, showing the two stereo channels, is above and aligned with the acoustic measures.

Relationships between Acoustic Intensity, Spectral Flatness and Perceived Change

The perceived change time series is only stationary after one differencing, and the resultant series is named dchange below (where d or d2 … dn indicate the number of differencing steps applied to achieve stationarity). Dintensity shows highly significant ($p < .001$) Granger causality upon the dchange series derived by differencing the perceived change series correspondingly, as judged by a preliminary VAR analysis (Table 1). There was no reciprocal causality, appropriate given that the acoustic series are literally exogenous. The differenced spectral flatness time series ('dspectralf') was not Granger causal of perceived change.

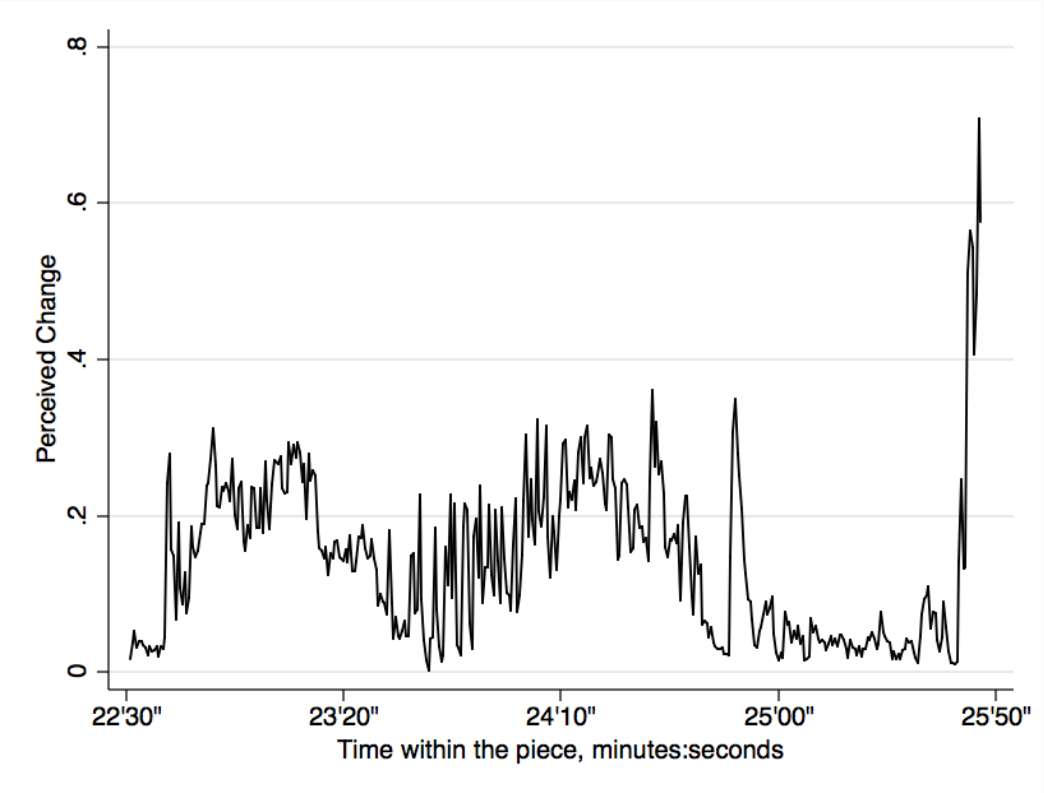

Fig. 3. Perceived 'change in sound' through time in the Wishart extract, averaged across participants.

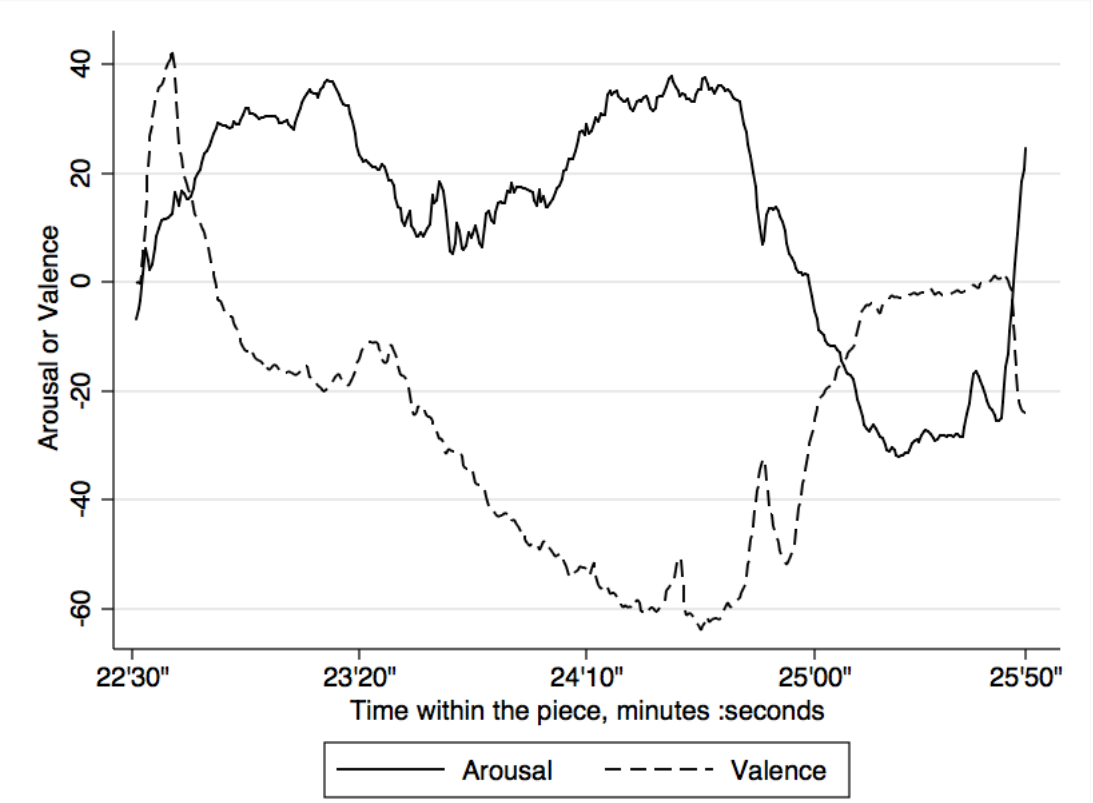

Fig. 4. Perceived arousal and valence through time in the Wishart extract, averaged across participants.
Table 1. Granger causality Wald tests for predictor series influencing the dchange time series

Predictor series	chi2	degrees of freedom (variables)	Prob > chi2
dintens	159.4	3	.000
dspecf	4.26	3	.234
ALL	178.5	6	.000

Note. The Table gives a probability that the chi-square value associated with the individual series dintensity and dspectralflatness as predictors of dchange could occur by chance, and the corresponding probability for the model as a whole. The degrees of freedom are those associated with the lags of the individual predictors and the model as a whole.

Thus detailed ARIMAX modeling solely of the relationship between intensity (dintensity) and perceived change (dchange) is undertaken, and the results (a model of the local mean value through time, as opposed to its variance: see below) are shown in Table 2 and Table 3. Cross-correlation analysis suggested that several lags of the dintensity series might impact on dchange, which would correspond to influences beyond 1 s, not unreasonable given previous literature (Schubert, 2001, 2004; Sloboda & Lehmann, 2001). Lags are referenced as L1 L2 … Ln. As shown by the Wald chi, the model overall is highly significant. L1 and L2 were also individually highly significant. The autoregressive part of the model of the dchange series is also shown. Thus AR lags 1-3 were highly significant. No constant was required. The results from the analysis are shown in detail in Table 2, to illustrate the source of the summarized results shown for other analyses in this paper.

Table 2. ARIMA(X) model of the influence of intensity on real-time perceived change in Wishart's *Red Bird* extract, using once-differenced series. ARIMA regression model: l(1,2).dintensity, ar(1,2,3) no constant. Model probability, $p < .0000$

	Coef.	Std. Err.	Coef. p <	[95% Conf.	Interval]
modeled series: dchange					
Predictor series:					
dintensity					
L1	.006	.000	.000	.005	.006
L2	.002	.000	.000	.001	.002
ARMA					
ar					
L1	-.391	.033	.000	-.455	-.327
L2	-.285	.043	.000	-.370	-.200
L3	-.145	.040	.000	-.224	-.065
/sigma	.045	.001	.000	.043	.047

Information Criteria:

Observations	Log likelihood(model)	degrees of freedom (model)	AIC
390	655.605	6	-1299.211

Note. The model summary means that dchange is predicted by lags 1 and 2 of dintensity (exogenous predictor series), together with autoregressive lags 1-3 of itself. No constant is needed for the model. The model predicts dchange as .006(l1.dintensity) + .002(l2.dintensity) - 0.391(l1.dchange) - 0.285(l2.dchange) - 0.145(l3.dchange). All the predictor series are individually highly significant. The number of observations reported in the model is reduced from that measured on account of the number of lags used in the model.

The residual series of this model as of all others presented is white noise, and the prediction from the model does not show forecast bias and has a Theil's U less than 1, indicating a useful predictive capacity. The Akaike Information Criterion (AIC) reaches its minimum value for ARIMAX models of -1299.2, which can be directly compared with those from the other ARIMA(X) models of this same time

series, discussed shortly. The correlation between actual and forecast is 0.62, and the sum of squared errors is 0.79. The median absolute percentage error is 16.1%.

Table 3 also shows the best ARCH model, which models the variance as well as the mean. Dintensity induced highly significant conditional heteroskedasticity, representing its influence on the variance change in the endogenous variable dchange, and both ARCH(1) and GARCH(1) were highly significant. This model was significantly improved (for example, AIC was -1376.0) in comparison with the ARIMAX only model.

Table 3. ARIMAX with ARCHX model of the influence of intensity on real-time perceived change in Wishart's *Red Bird* extract, using once-differenced series. ARCH family regression, ARMA disturbances and conditional heteroskedasticity: the model is l(1).dintensity, ar(1,2,3) noconst het(dintensity) arch(1) garch(1). Model probability, $p < .0000$.

	Coef.	Std. Err.	Coef. p <	[95% Conf.	Interval]
Modeled series: dchange					
Predictor series:					
dintensity					
L1	.004	.000	.000	.004	.004
ARMA					
ar					
L1	-.403	.066	.000	-.532	-.275
L2	-.197	.060	.001	-.314	-.079
L3	-.113	.060	.060	-.231	.005
HET					
dintensity	-.244	.030	.000	-.303	-.185
constant	-9.225	.294	.000	-9.802	-8.648
ARCH					
arch					
L1	.342	.056	.000	.232	.4523
garch					
L1	.638	.036	.000	.569	.708

Information Criteria:

Observations	Log likelihood(model)	degrees of freedom (model)	AIC
391	695.990	8	-1375.981

Note. Besides ARIMA components considered in earlier examples, this model includes autoregressive heteroskedasticity (represented by the ARCH and GARCH predictors) and the influence of an exogenous factor (dintensity) on the heteroskedasticity of dchange, the modeled series. This latter influence is commonly termed multiplicative heteroskedasticity.

In contrast to these ARIMAX/ARCH models, incorporating the several influences of the exogenous variable intensity, the best of the acceptable ARIMA models of the autoregressive dchange perceptual series alone (AR(1,2,3), with ARCH (1) and GARCH(1) components included), had an AIC of -1266, and showed a correlation between actual and forecast of only .31, had a sum of squared errors of 1.16 and median absolute percentage error of 49.5%, being substantially worse on all counts even than the simple ARIMAX model without ARCH components. These data reveal that even though an autoregressive ARIMA-only model progressively embodies the continuing impact of prior exogenous events, such as intensity changes, dintensity is an important predictor in the ARIMAX model, which enhances its performance. Such a positive correlation might be expected of a variable that is psychologically causal, but it is not necessarily always so strong. Furthermore, dintensity is also a significant predictor of conditional heteroskedasticity, as shown in the ARCH modeling, underscoring its potential perceptual influence.

As mentioned already, given the knowledge of the intensity time series, spectral flatness was not a significant predictor; though it is worth noting at this point that considered alone dspectralf showed Granger Causality upon dchange (and not vice versa). This possible modest influence of spectral flatness

on perceived change is assessed further in the VAR analyses below, which also provide estimates of the extent of the impact of the acoustic variables upon the perceptual outcomes.

Time-Series Modeling of Relationships between Acoustic Variables and the Perception of Affect

ACOUSTIC VARIABLES AND AROUSAL: ARIMAX/ARCH MODELS OF THE INFLUENCE OF INTENSITY ON AROUSAL

Only intensity is significant in Granger Causality tests based on VAR of the untransformed series; there is no reciprocal Granger causality. However the arousal time series is not stationary until once differenced ('darousal'), and thus consideration is given to darousal, dintensity and dspectralf. In this context again dintensity but not dspectralf was Granger-causal upon darousal, and there was no reciprocal influence. Thus modeling is undertaken for the relationships of dintensity to darousal.

The cross-correlation analyses indicate that lags of dintensity up to 20 might be influential, and accordingly the best ARIMAX alone model includes lags 1-10 and 17 (with coefficients from .13 at lag 2, to 0.02 at lag 17), together with AR(1,3) and MA(20) (where MA(n) is a moving average window of n lags in length). The AIC is 1112.8, and there is a correlation between actual and forecast of .70, with a median absolute percentage error of 12.3%. The ARCH assessment demonstrates that dintensity does not influence heteroskedasticity of darousal, but that AR(1), MA(20), ARCH(1) and GARCH(1) improves the AIC to 1040.8, without change to the ARIMA lags of intensity in the model, though their coefficients are now from .10 at lag 2 to .02 at lag 17 (ARCH(n) and GARCH(n) again indicate respective model components with n lags). The ARIMAX/ARCH modeling of this relationship suggests that dintensity is a positive influence upon darousal, and hence similarly intensity upon arousal.

ACOUSTIC VARIABLES AND VALENCE: ARIMAX/ARCH MODELS OF THE INFLUENCE OF SPECTRAL FLATNESS ON VALENCE

The VAR of untransformed perceptual valence, spectral flatness and intensity time series reveals neither acoustic series to be Granger-causal for the perceptual series; nor does the literature provide a previous strong association. However, valence is made stationary only by once differencing, and hence the VAR is also performed on the differenced series (dvalence, dintensity, dspectralf). This shows significant ($p = .021$) Granger Causality of dspectralf on dvalence (and not vice versa), and hence this is modeled by ARIMAX. A highly significant model with lags1-12 and 15 of dspectralf (coefficients between .55 and .04) and autoregressive lags 1 and 4 (coefficients .54 and .08) was obtained (no constant in the model). This model passed the quality tests, and showed an AIC of 1062.6, a correlation between actual and forecast of .68, median absolute percentage error of 4.4%, and had a sum of squared errors of 342. ARCH modeling revealed a significant conditional heteroskedasticity of dspectralf on dvalence and the best ARCH/GARCH model was dspectralf lags 1-11 and 15, together with ARCH(1), GARCH(1), AR(1,4) and the heteroskedastic input of the dspectralf series (coefficient -.57). It had an AIC of 1028.9, passed the quality tests, again showed a correlation between actual and forecast of .68 and had a sum of squared errors of 351, together with median absolute percentage error of 11.2%, thus being slightly worse than the ARIMAX only model. Figure 5 compares the ARIMAX forecast for 50'' to 1'40'' with the corresponding data series; these forecasts are extracted from the prediction of the whole series. The quality of precision shown is highly representative of that for the whole series (from which it is taken); a short section is chosen simply so that the detail is readily visible. As is generally the case in such good predictions, the forecast follows quite closely, but often in a damped-down fashion. These data suggest that as sonic complexity increases, valence becomes more positive.

In contrast with the best ARIMAX-only model, the best ARIMA only model is simply AR(1), and it had a significantly worse AIC of 1171.1. It just passed Theil's U, did not show forecast bias, and had an almost equally high correlation between forecast and actual (.65), with sum of squared errors at 460.5, and median absolute percentage error less than 0.1%.

Dspectralf thus seems to be potentially an influence on dvalence, but this view is slightly tempered by the relatively good fit achieved with the ARIMA-only model. On the other hand, as noted already the autoregressive components of an AR model include the representation of the earlier impacts of

any exogenous factors. The following VAR analyses consider further the possible influence of spectral flatness on valence.

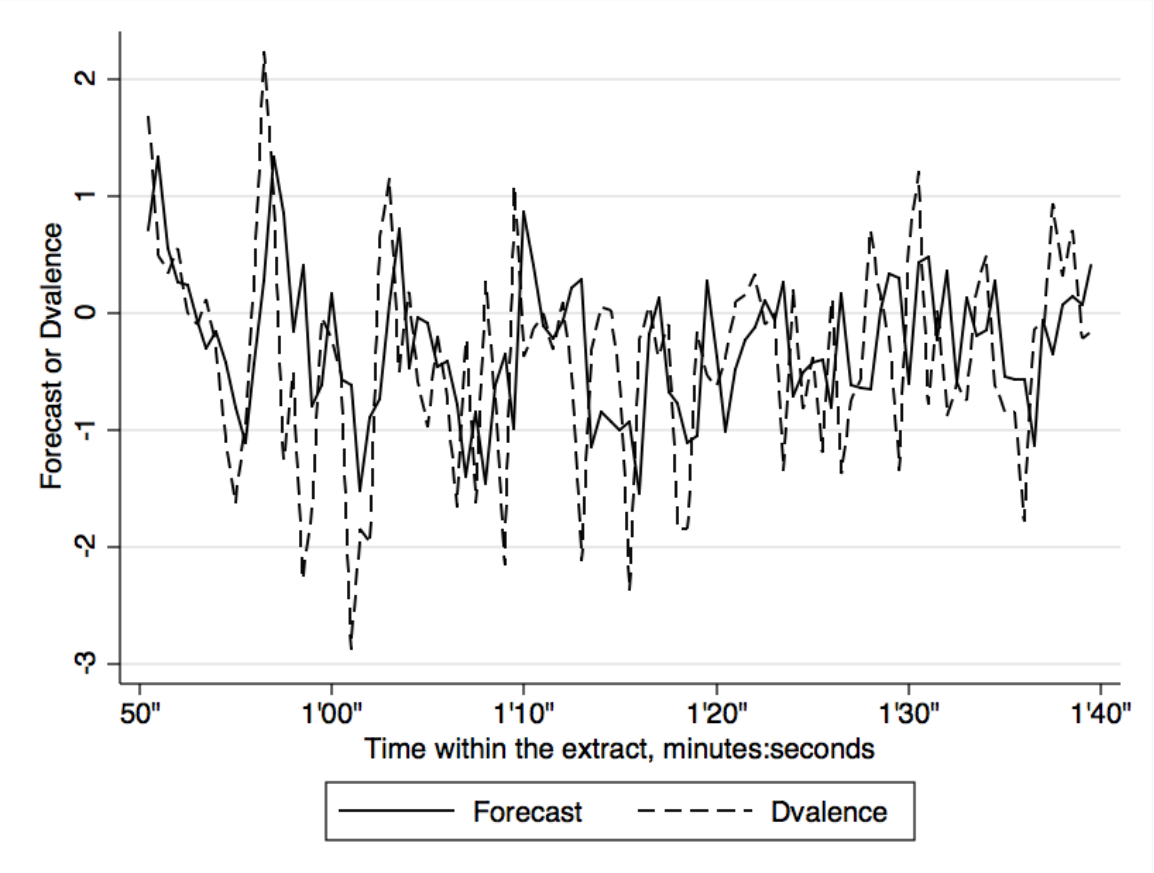

Fig. 5. The valence series (differenced) together with its ARIMAX model prediction, taken from 50-140 s into the extract. The forecast is an attenuated version of the measured valence series, but follows very closely.

VECTOR AUTOREGRESSIVE (VAR) ANALYSIS OF RELATIONS BETWEEN ACOUSTIC AND PERCEPTUAL VARIABLES

A key benefit of VAR is its capacity to co-relate multiple time series simultaneously, and to treat them either as potentially mutually influential ('endogenous'), or potentially independent, solely input ('exogenous') variables. We adopt primarily the statistically less restrictive approach in which variables are all treated as endogenous. In addition, most presented analyses concern stationary variables. Comparative VARX studies in which the acoustic variables are treated as exogenous (X) are made in each case; these simplify the vector decomposition. Both VAR on untransformed variables and VARX in the present papers produce results concordant with the VAR of stationarized variables (and hence VARX results are not shown).

In the case of dchange, as noted above, the VAR together with both dintensity and dspectralf shows only dintensity to be Granger causal, and not dspectralf. Thus Figure 6 shows the result of a (4-lag) VAR of dchange and dintensity in the form of the forecast-error variance decomposition (FEVD), which is an indication of the impact of variables on a given output, lag by lag, in this case on dchange. Such impacts are termed 'impulse response functions', as in the Figures. The 95% confidence intervals confirm the significant impact of dintensity on dchange, consistent with the ARIMAX correlation of the forecast from the dintensity/dchange model with actual, described above.

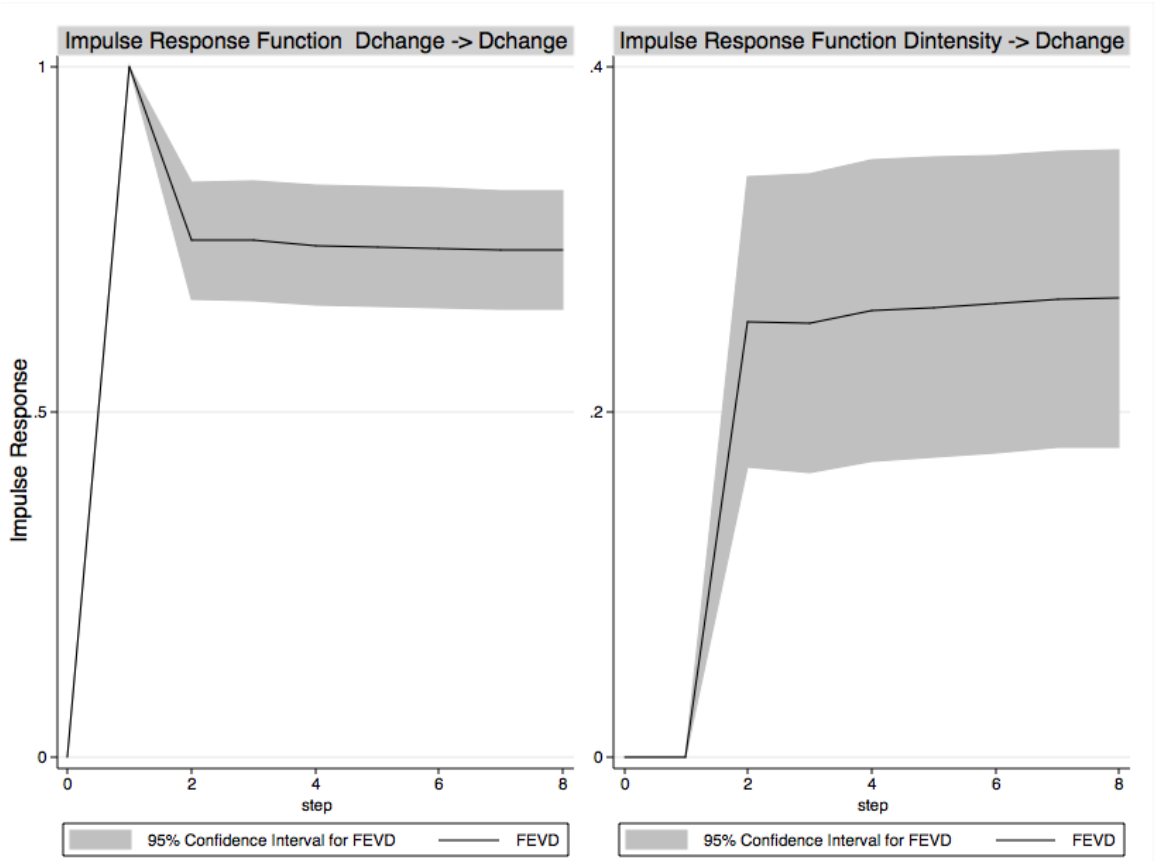

Fig. 6. Forecast-error variance decomposition (FEVD) from a 4-lag VAR, representing the lag-by-lag impact of dintensity on dchange, the 'impulse response function'. The 95% confidence intervals (shaded) confirm the significant impact of dintensity on dchange, since they do not breach zero. From lag 2 (1 sec) onwards the predictive power of change on itself (i.e. its autoregressive property) declines, while correspondingly that of dintensity increases and maintains statistically significant values.

Given the just significant ($p = .046$) Granger causality of spectral flatness on change judged by a VAR of the untransformed variables (with white noise residuals for the change model), Figure 7 shows the resultant FEVDs. This confirms the substantial impact of intensity (the FEVD reaches a maximum of almost 0.5), but shows that the confidence limits for the impact of spectral flatness render it not only very small (< 0.03) but also statistically insignificant.

Spectral flatness is not considered in VAR of arousal, because it is not Granger-causal with respect either to the untransformed or differenced series, as mentioned above. Figure 8 shows the significant impact of intensity on arousal as judged by the appropriately tested VAR of the untransformed time series.

VAR of valence with untransformed intensity and spectral flatness series shows neither as Granger-causal, within the 4 lag-analysis prescribed by the AIC criterion for order of VAR. After making the series stationary, dspectralf (but not dintensity) is Granger-causal on dvalence as noted above. However, its impact is very slight (FEVD maximum c. 0.02) and not statistically significant. The impact of dintensity is slightly bigger, but again, not statistically significant. The untransformed series is investigated further, and lags up to 7 (3.5s), predicated by the lag order selection criterion, again showing a significant Granger-causality for spectral flatness upon valence, while intensity remains just outside significance. However, it is only intensity whose FEVD reaches modest values (c. 0.17 at lag 8) to which confidence levels above zero attach.

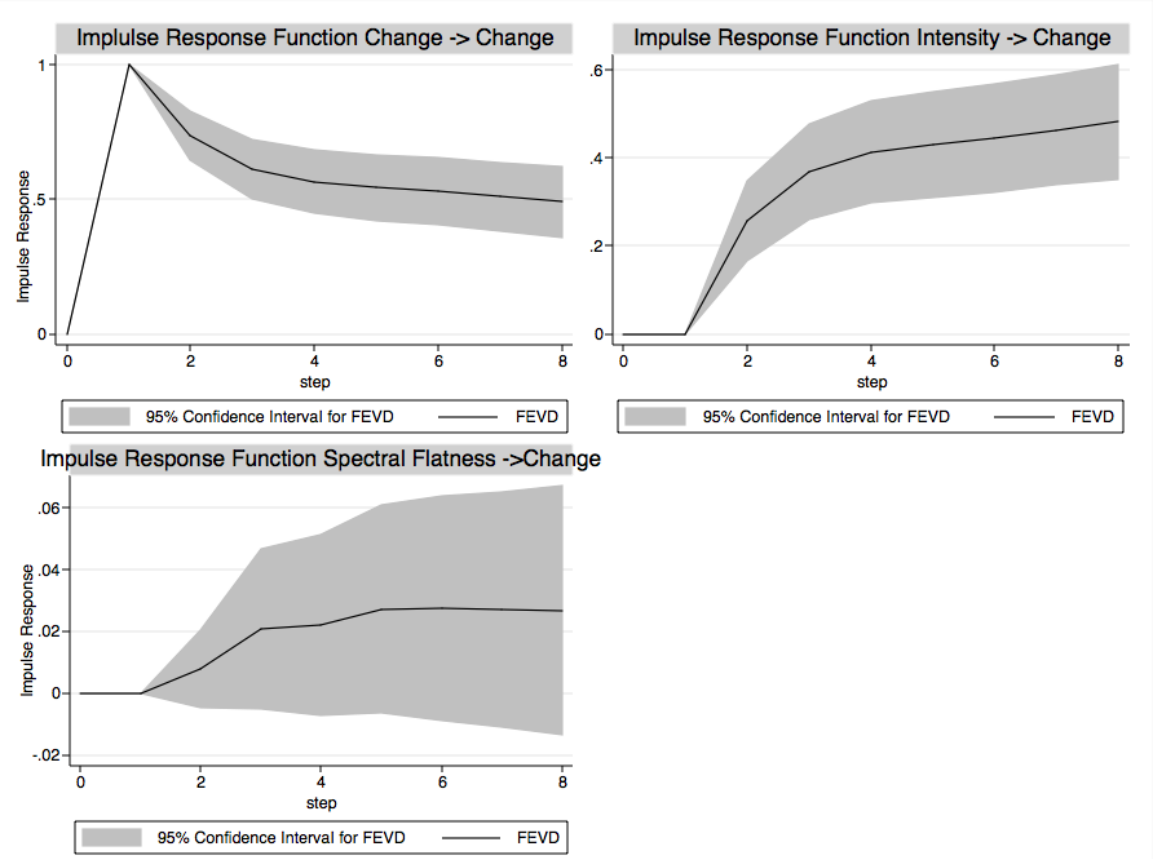

Fig. 7. FEVDs from a VAR of the untransformed variables (with White Noise residuals for the change model). The figure shows a strong impact of intensity on change, but not spectral flatness on change (small and statistically insignificant given the confidence limits).

Before leaving VAR analyses, advantage is taken of its capacity to model multiple series even-handedly, thus allowing the study of the possible mutual influences of the perceptual variables, notably mutual influences amongst change, arousal and valence. In other words, as mentioned already, variables entering a VAR model can be treated either as endogenous (potentially mutually influenced, such as our perceptual variables) or exogenous (solely a source of input influence, such as our acoustic variables). These two categories of statistical variable correspond to psychological dependent and independent variables, but with one difference. The difference is that statistical endogenous variables can be readily assessed for mutual influence, whereas to do this with 'dependent' variables in psychological experiments is more complex.

Using four lags, such a VAR model including all the perceptual and acoustic variables as endogenous passes all tests and suggests Granger-causality of arousal and intensity upon change; change and spectral flatness upon valence; and valence and intensity upon arousal. There are no causalities on intensity, but valence is Granger-causal of spectral flatness, which of course is only because spectral flatness is being treated solely for purposes of conservative analyses as an endogenous (dependent) variable though it is clearly not. After impulse response function analysis only intensity is significant for change (judged by the confidence intervals not breaching zero), reaching > 0.4 at lag 8, in agreement with earlier analyses. By the same assessment, only change influences valence (maximum FEVD c. 0.18 at lag 8); and only intensity influences arousal (maximum FEVD c. 0.38 at lag 8). The conclusions from the VAR with all series included are consonant with those reached earlier.

Thus, valence responses to the Wishart extract are only weakly influenced by the chosen acoustic variables. This encourages further consideration of the specific sonic features of the Wishart piece, which were amongst the bases for choosing it for study (see below); these might be of particular importance for valence responses.

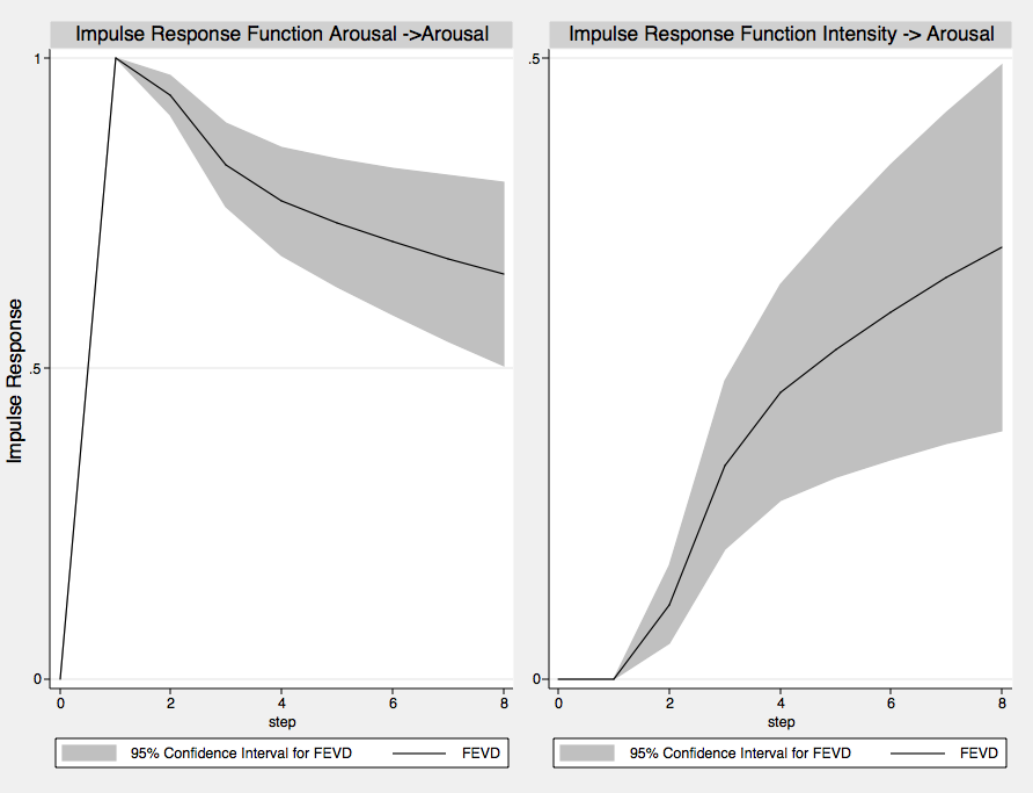

Fig. 8. FEVD from a VAR of the untransformed time series showing a significant impact of intensity on perceived arousal.

Animate Sounds and their Influence: Possible Impacts of 'Agency'

Current musicological theory is heavily concerned with issues of narrative and agency (Clarke, 2005; Maus, 1997; Overy & Molnar-Szakacs, 2009). 'Agency' is often a metaphor for perceiving or attributing 'anthropomorphic influence', and Wishart is a composer with an intense involvement with the human voice and its possible impacts (T. Wishart, 1985). So one reason for our choice of the section of his piece for study is that it contains many intermittent impressions of human vocal sounds and other animate but non-human sounds. The locations of the human vocal and the animate sounds in the piece, as judged by two musicologists experienced in making and listening to electroacoustic sounds, are shown graphically in Figure 9. The perceptible events are all short and discrete, thus the timing of these zones is clear-cut to a greater precision than the 2Hz sampling rate for the acoustic variables. The influence of these two potential agents upon arousal and valence was assessed in additional ARIMA and VAR analyses, entering them as impulse variables (i.e. of constant effect throughout their presence, and nil in their absence), and assessing whether such an effect is present and statistically significant (i.e. has a coefficient significantly different from zero).

The most interesting result is that the ARCH model of dvalence described above, which includes the conditional heteroskedasticity of the input dspectralf series, is improved by the heteroskedastic input of animate sounds ('animatesd') (AIC 1023.0), but dspectralf itself becomes statistically insignificant while animatesd is significant. When dspectralf is dropped from the model and animate sound retained amongst the potential conditional heteroskedastic variables, and amongst the mean model variables, the model is improved still further (AIC 1015.4), and animatesd remains highly significant ($p = .001$ for the mean model and $p = .005$ for heteroskedasticity). Given this model, the addition of the human voice parameter is not significant for the mean model, but it influences heteroskedasticity ($p = .008$) and enhances the AIC to 1010.6. These results suggest the importance of sonic and structural features that are specific to the perception of physical origins for sounds; amongst other things, in future studies these may supplement the acoustic variables of focus here.

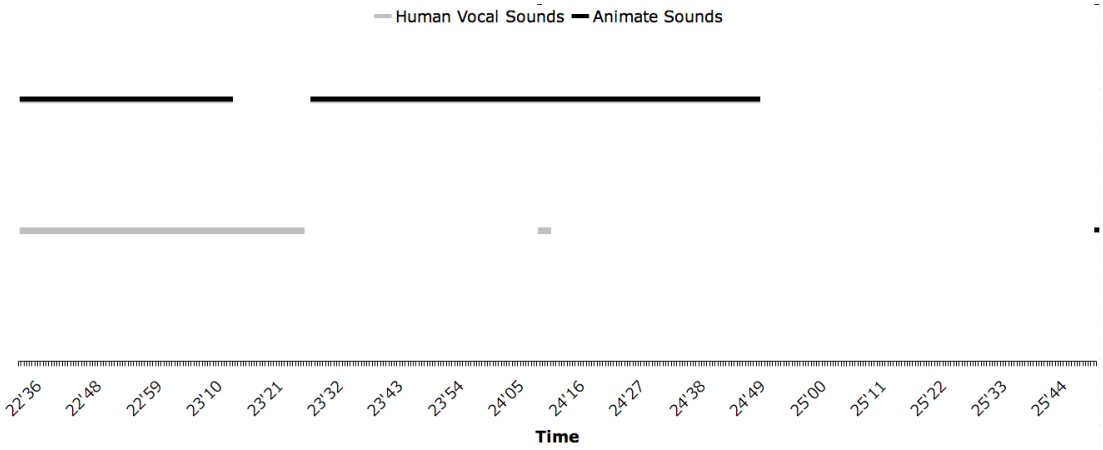

Fig. 9. A time line of the Wishart extract, indicating where there are animate and human voice sounds.

DISCUSSION

It is important to re-emphasize that statistical analyses such as these demonstrate significant correlations. These may reflect causal relationships, and they constitute a minimum criterion for considering such causality further. One of the important reasons for studying electroacoustic music is that it can be generated by entirely algorithmic means, such that experimenters possess a logical and quantitatively definable means of altering pieces, without having to make questionable assumptions about the nature and method of composition of a previous piece e.g. by Haydn; or about the potential perceptual impacts of the tuning or metrical systems it uses e.g. Western tonality (this is not to imply that Wishart himself composes in this way). Thus in future experiments, electroacoustic compositions can be entirely defined as algorithmic structure, and systematic perturbations of the algorithm (e.g. the range of spectral flatness over which it evolves, or the pattern of intensity expressing an otherwise unchanged sonic structure) can be used to assess the consequent changes in perceptual responses. This will allow further insight into causal relationships between acoustic properties and perception.

The most important purpose of this paper is to develop an autoregressive time series analysis platform for such further studies of acoustic-perceptual relationships in music. Besides this, the striking results are the demonstration of strong positive correlations between temporal patterns of intensity and perceived change and arousal. While there are few previous data on the perception of continuous change, the result for arousal is consistent with some earlier research, such as that by Schubert on Western classical music extracts (Schubert, 1999, 2001, 2004). By providing much fuller time series analysis, and by using unfamiliar music and a wide range of listeners, the current study strongly supports the conclusion that intensity profiles are important in perceived real-time arousal expressed in many kinds of music. The use of the 2-D emotion space is based on previous multidimensional scaling studies of relationships between the retrospectively perceived overall affect expressed in short extracts and their averaged acoustic properties: these suggest that arousal and valence dimensions of affect can at the least usefully be separated (Bigand, et al., 2005; Leman, et al., 2005). The results from the current VAR analyses overall confirm their statistical independence in the data assessed here.

It is also shown here that intensity can be a strong influence on perception of continuous change in music. The measure of perceived change through time has hardly been studied previously apart from our earlier study on short juxtaposed segments of computer generated sound (Bailes & Dean, 2009). It is difficult to test in a continuous response paradigm whether musically untrained listeners perceive 'structural' events in music in the sense that music analysts refer to structure, because listeners do not usually have an explicit awareness of the parameters of such analyses. Evidence on failure of perception of large-scale musical features such as tonal closure support this view (Cook, 1987). Evidence on the short time scale does suggest that listeners perceive segmentation in a manner largely influenced by surface events in the sound stream (Deliège, et al., 1996; Lalitte et al., 2004). The approach taken in the current study allows listeners to identify change for themselves, and deals directly with the sonic temporal patterns that they hear, and which therefore are likely to influence them affectively. This is not to suggest the

unrecognized or unattended features cannot be affective, but that it may be more feasible to demonstrate the influence of recognized features. Further analysis of the patterns of perceived change can allow the demonstration of larger scale 'structure' within it.

This paper illustrates an alternative approach to that of assessment of the importance of musical features defined by composer/musicologist, particularly in the investigation of the impact of animate and human voice sounds on listener perceptions and on their variance. Clear evidence is provided for an impact of these features, and the approach can readily be generalized. For example, in continuous perception of a sonata movement, large-scale analytical structures could be assessed as determinants of perceptual outputs. More limited versions of such an approach have suggested some coincidences between musical events defined analytically and perceptual changes (e.g. Krumhansl, 1996; Madsen & Fredrickson, 1993). The animate and human voice sounds in the present piece seem to possess agency, in that they can be metaphorically described as playing an 'active' role in a partially predictable series of events (the narrative). As noted above, such metaphors are generally anthropomorphic, but here some sound sources are literally human, while the animate sounds can be construed anthropomorphically. In future work it will be feasible to postulate agency in a concerto solo part, and empirically assess its perceptual influence; or to assess the impact of identifiable physical sources for environmental sounds of potential biological importance (Clarke, 2005; Gaver, 1993). In other recent work (Bailes & Dean, submitted), we have successfully used the current methodological approach to compare analyses of the Wishart piece with those of two other electroacoustic pieces, and with a piano piece from the same century and a classical orchestral piece, which are both note-centered rather than based on sonic texture. This work also compares the perceptions of highly trained musicians with others. The work confirms the general utility of our approach.

It is worth discussing two technical aspects of statistical procedures relevant to this work. The first is our approach of simply averaging the multiple time series from our 32 participants, as is commonly done. While this removes the variability between the individual series, they can be studied separately, as we have done in work under preparation. There are at least two alternative approaches that can be used, singly or in combination. The first is 'standardization' of each series, which means expressing the values in terms of their deviation from the series mean (which can be arbitrarily set at any appropriate value for this purpose) as a proportion of the overall standard deviation for that series. This routine technique reduces apparent variability between individuals, but does not reduce the multiple series to just one, and so it too is often followed by simple averaging. The second is 'registration', a feature of time warping. The principle here is either that every feature of each time series should be shared, and hence successive peaks and troughs should align, or at least that there are certain 'landmarks' which should align. Such an approach is obviously relevant when it is clear that the multiple replicate time series being studied do represent attempts to reproduce exactly the same sequence of events: for example, a physical movement of a computer mouse in straight lines between several successive fixed locations. It is not at all apparent that we should expect such consistent performance in perceptual time series, and hence this approach was not adopted, in favor of the less restricted approach of simple averaging. In other work we will analyze individual participant's time series, and present participant group comparisons.

It may be useful to note that another statistical technique, Functional Data Analysis (e.g. Ramsay & Silverman, 2002; Ramsay & Silverman, 2005) has been used in some studies of music performance and perception (Levitin, Nuzzo, Vines, & Ramsay, 2007; Vines, Nuzzo, & Levitin, 2005), again to deal with events patterned in time. Functional Data Analysis (FDA) is a newer technique than TSA, thus less extensively developed. It is more routine in FDA for standardization and registration steps to be used to create a reduced summary time series, and an iterative approach is normal. After this, discrete time series data points are converted to functions. One benefit of this is the possibility of differentiating the function, and thus looking at velocity and acceleration of the function, and their relationship in phase-plane plots, which can be highly informative with respect to some processes. The differencing process in TSA can provide similar velocity and acceleration information if required. This can be obtained either at the finest time resolution provided by the time series data, or by differentiating a locally smoothed polynomial line-fit, analogous to the smoothing step creating the FDA function. The velocity and acceleration of acoustic or perceptual parameters could be viewed as aspects of 'musical motion'. As might be expected given the comments above about physical movement, FDA has been particularly valuable for example in studies of repetitive movement patterns. It commonly also uses an initial smoothing step for each series under study, assuming that the time series are really representations of continuous and smooth processes, rather than discontinuous and abrupt ones. Music such as we study which is rapidly and irregularly changing may not securely fit this description; similarly in note-based instrumental music, we may have a discontinuous

event series, as much as an ongoing process. FDA does not generally focus so extensively on autocorrelation structures as does TSA, though it could, and it is not yet fully developed for multivariate analyses, as Ramsay and Silverman themselves note, though there are many variants of it already. FDA and TSA may be appropriate complementary approaches for work such as we describe.

The time series analysis methodology developed in this paper can support ongoing work to understand the perception of music in terms of dynamically changing information content (Abdallah & Plumbley, 2009; Pearce & Wiggins, 2006; Potter, Wiggins, & Pearce, 2007; reviewed by Wiggins, Pearce, & Müllensiefen, 2009). In an information dynamics approach it is assumed that the predictive capacity and decisions of an observer concerning future events in the musical stream are continuously evolving as new information is assimilated into the observer's (potentially individualistic) data probability structure. Such studies, though as yet lacking in empirical assessment, generally also implicitly assume that it is apparent what information can be extracted from each event (e.g. that pitches within equal-tempered tonal space can be readily categorized), and hence a probabilistic information theoretic value can be attached to features like a pitch event, such as the entropy or unexpectedness of the event. Information Dynamic approaches have so far treated events as successive and accretive, without taking regard for their relative timing, or directly addressing issues of temporal autocorrelation. Indeed, much of the Information Dynamics work has been done on monophonic isochronic music, such as Glass' Gradus (Abdallah & Plumbley, 2009; Potter, et al., 2007). The Information Dynamic approach can be enhanced further by superimposed time series analyses, as well as by awareness that notationally- or musicologically-defined features may or may not be perceptible and readily categorized, and hence may or may not contribute new information. For example, both pitch (Pollack, 1952) and intensity (Garner, 1953) can only be precisely and reliably identified in about seven steps across their whole audible ranges. Hence information dynamic scales rather different from those by which pitch is notated (with more than 100 steps) might ultimately be needed for them. While we can perceive fine-grained distinctions, arguably these cannot readily be placed into more than seven information categories.

One of our reasons for the choice of spectral flatness as our initial 'global' acoustic parameter related to timbre was that it has already been used as an alternative approach to determining a kind of 'information content' of continuously varying sound sources (Dubnov, 2006) and in future work this aspect of its perceptual influence can be studied further by our developed time series approach. Our approach also permits the use of vectorial data (such as the mel-frequency cepstral coefficients, MFCC) by VAR time series analysis, and these much more complex analyses will be important tools for both future information dynamic and acoustic-perceptual studies.

Time series analysis should be intrinsic to the analysis of neurophysiological data, such as those from EEG, and electrodermal activity studies. In some cases, techniques have been applied in depth (e.g. some EEG analyses) while in others they are largely still lacking (e.g. in studies of skin conductance responses reflecting aspects of the autonomic nervous system). Their utility in other fields, notably ecology, is undergoing rapid expansion at present (Zuur, Ieno, & Smith, 2007). This paper demonstrates several aspects of the utility of TSA in the study of acoustic/perceptual relationships. It also points to several future possibilities, particularly for multivariate analyses.[1]

NOTES

[1] We are very grateful for expert advice on time series statistics from Prof. William Dunsmuir, Dept of Mathematics and Statistics, University of New South Wales, Australia, and for his interest in our topic. We also benefit from ongoing collaboration and discussion with Dr Marcus Pearce, Prof. Geraint Wiggins and colleagues, Goldsmiths' College London. This research was supported by Australian Research Council Discovery grant DP0453179.

REFERENCES

Abdallah, S., & Plumbley, M. (2009). Information dynamics: Patterns of expectation and surprise in the perception of music. *Connection Science,* Vol. 21, No. 2-3, pp. 89-117.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, Vol. 19, No. 6, pp. 716-723.

Bailes, F., & Dean, R.T. (2007). Listener detection of segmentation in computer-generated sound: An exploratory experimental study. *Journal of New Music Research*, Vol. 36, No. 2, pp. 83-93.

Bailes, F., & Dean, R.T. (2009). Listeners discern affective variation in computer-generated musical sounds. *Perception*, Vol. 38, No. 9, pp. 1386-1404.

Bailes, F., & Dean, R. T. (submitted). Comparative time series analysis of perceptual responses to electroacoustic music.

Bartlett, M.S. (1966). *An Introduction to Stochastic Processes.* 2nd edition. Cambridge University Press.

Beckers, G.J.L. (2004). http://www.gbeckers.nl/pages/praatscripts.html

Bigand, E., Vieillard, S., Madurell, F., Marozeau, J., & Dacquet, A. (2005). Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts. *Cognition and Emotion*, Vol. 19, No. 8, pp. 1113-1139.

Boltz, M.G. (1998). Tempo discrimination of musical patterns: Effects due to pitch and rhythmic structure. *Perception & Psychophysics*, Vol. 60, pp. 1357-1373.

Box, G.G., Jenkins, G.M., & Reinsel, G.C. (1994). *Time Series Analysis - Forecasting and Control.* 3rd edition. Prentice Hall.

Bradley, M.M., & Lang, P.J. (2000). Affective reactions to acoustic stimuli. *Psychophysiology*, Vol. 37, pp. 204-215.

Brittin, R.V., & Duke, R.A. (1997). Continuous versus summative evaluations of musical intensity: A comparison of two methods for measuring overall effect. *Journal of Research in Music Education*, Vol. 45, pp. 245-258.

Brown, J.C. (1993). Determination of the meter of musical scores by the method of autocorrelation. *Journal of the Acoustical Society of America*, Vol. 94, No. 4, pp. 1953-1957.

Caclin, A., McAdams, S., Smith, B.K., & Winsberg, S. (2005). Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones. *Journal of the Acoustical Society of America*, Vol. 118, No. 1, pp. 471-482.

Chapin, H., Large, E., Jantzen, K., Kelso, J.A.S., & Steinberg, F. (2008). Dynamics of emotional communication in performed music. *Journal of the Acoustical Society of America*, Vol. 124, p. 2432.

Clarke, E.F. (2005). *Ways of Listening: An Ecological Approach to the Perception of Musical Meaning*. New York: Oxford University Press.

Cook, N. (1987). The perception of large-scale tonal closure. *Music Perception*, Vol. 5, No. 2, pp. 173-196.

De Vries, B. (1991). Assessment of the affective response to music with Clynes's sentograph. *Psychology of Music*, Vol. 19, pp. 46-64.

Dean, R.T. (Ed.). (2009). *The Oxford Handbook of Computer Music*. New York, USA: Oxford University Press.

Deliège, I., Mélen, M., Stammers, D., & Cross, I. (1996). Musical schemata in real-time listening to a piece of music. *Music Perception*, Vol. 14, No. 2, 117-160.

Dickey, D.A., & Fuller, W.A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, Vol. 74, No. 366, pp. 427-431.

Dubnov, S. (2006). Spectral anticipations. *Computer Music Journal*, Vol. 30, pp. 63-83.

Dubnov, S., McAdams, S., & Reynolds, R. (2006). Structural and affective aspects of music from statistical audio signal analysis. *Journal of the American Society for Information Science and Technology*, Vol. 57, No. 11, pp. 1526-1536.

Elliott, G., Rothenberg, T.J., & Stock, J.H. (1996). Efficient tests for an autoregressive unit root. *Econometrica*, Vol. 64, No. 4, pp. 813-836.

Enders, W. (2004). *Applied Econometric Time Series.* 2nd edition. Hoboken, NJ: Wiley.

Engle, R.F., & Granger, C.W.J. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, Vol. 55, No. 2, pp. 251-276.

Gabrielsson, A., & Lindstrom, E. (2001). The influence of musical structure on emotional expression. In: P.N. Juslin & J.A. Sloboda (Eds.), *Music and Emotion: Theory and Research*. London: Oxford University Press, pp. 223-248.

Garner, W.R. (1953). An informational analysis of absolute judgments of loudness. *Journal of Experimental Psychology*, Vol. 46, pp. 373-380.

Gaver, W.W. (1993). What in the world do we hear?: An ecological approach to auditory event perception. *Ecological Psychology*, Vol. 5, No. 1, pp. 1-29.

Gordon, J.W., & Grey, J.M. (1978). Perception of spectral modifications on orchestral instrument tones. *Computer Music Journal*, Vol. 2, No. 1, pp. 24-31.

Granger, C.W.J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, Vol. 37, No. 3, pp. 424-438.

Hamilton, J.D. (1994). *Time Series Analysis*. Princeton: Princeton University Press.

Kessler, E., J., Hansen, C., & Shepard, R.N. (1984). Tonal Schemata in the Perception of Music in Bali and in the West. *Music Perception*, Vol. 2, No. 2, pp. 131-165.

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

Krumhansl, C.L. (1996). A perceptual analysis of Mozart's Piano Sonata K. 282: Segmentation, tension, and musical ideas. *Music Perception*, Vol. 13, No. 3, pp. 401-432.

Krumhansl, C.L., & Kessler, E.J. (1982). Tracing the dynamic changes in perceived tonal organisation in a spatial representation of musical keys. *Psychological Review*, Vol. 89, No. 4, pp. 334-365.

Lalitte, P., & Bigand, E. (2006). Music in the moment? Revisiting the effect of large scale structures. *Perceptual and Motor Skills*, Vol. 103, No. 3, pp. 811-828.

Lalitte, P., Bigand, E., Poulin-Charronnat, B., McAdams, S., Delbé, C., & D'Adamo, D. (2004). The perceptual structure of thematic material in *The Angel of Death*. *Music Perception*, Vol. 22, No. 2, pp. 265-296.

Leman, M., Vermeulen, V., De Voogdt, L., Moelants, D., & Lesaffre, M. (2005). Prediction of musical affect using a combination of acoustic structural cues. *Journal of New Music Research*, Vol. 34, No. 1, pp. 39-67.

Levitin, D.J., Nuzzo, R.L., Vines, B.W., & Ramsay, J.O. (2007). Introduction to Functional Data Analysis. *Canadian Psychology*, Vol. 48, No. 3, pp. 135-155.

Lütkepohl, H. (2007). *New Introduction to Multiple Time Series Analysis*: Springer.

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*, Vol. 30, pp. 46-63.

Maus, F.E. (1997). Narrative, Drama, and Emotion in Instrumental Music. *The Journal of Aesthetics and Art Criticism*, Vol. 55, No. 3, pp. 293-303.

McAdams, S., Vines, B.W., Vieillard, S., Smith, B.K., & Reynolds, R. (2004). Influences of large-scale form on continuous ratings in response to a contemporary piece in a live concert setting. *Music Perception*, Vol. 22, No. 2, pp. 297-350.

McDowell, A. (2002). From the help desk: Transfer functions. *The Stata Journal*, Vol. 2, pp. 71-85.

MPEG-7 Overview § ISO/IEC JTC1/SC29/WG11N6828 (2004).

Ollen, J. Ollen Musical Sophistication Index. Retrieved October 28, 2008, from http://csml.som.ohio-state.edu:8080/OMSI/dispatcher

Olsen, K.N., Stevens, C., & Tardieu, J. (2007, 5-7 December 2007). *A Perceptual bias for increasing loudness: Loudness change and its role in music and mood.* Paper presented at the Inaugural International Conference on Music Communication Science, Sydney, Australia.

Overy, K., & Molnar-Szakacs, I. (2009). Being together in time: Musical experience and the mirror neuron system. *Music Perception*, Vol. 26, No. 5, pp. 489-504.

Pearce, M.T., & Wiggins, G.A. (2006). Expectation in melody: The influence of context and learning. *Music Perception*, Vol. 23, No. 5, pp. 377-405.

Pollack, I. (1952). The information of elementary auditory displays. *The Journal of the Acoustical Society of America*, Vol. 24, No. 6, pp. 745-749.

Potter, K., Wiggins, G.A., & Pearce, M.T. (2007). Towards greater objectivity in music theory: Information-dynamic analysis of minimalist music. *Musicae Scientiae*, Vol. 11, pp. 295-324.

Quinn, S., & Watt, R. (2006). The perception of tempo in music. *Perception*, Vol. 35, No. 2, pp. 267-280.

Ramsay, J.O., & Silverman, B.W. (2002). *Applied Functional Data Analysis: Methods and Case Studies*. New York: Springer.

Ramsay, J.O., & Silverman, B.W. (2005). *Functional Data Analysis.* 2nd edition. Springer.

Schubert, E. (1999). *Measurement and Time Series Analysis of Emotion in Music.* Ph.D. dissertation. University of New South Wales, Sydney.

Schubert, E. (2001). Continuous measurement of self-report emotional response to music. In: J.A. Sloboda & P.N. Juslin (Eds.), *Music and Emotion*. Oxford: Oxford University Press, pp. 393-414.

Schubert, E. (2004). Modeling perceived emotion with continuous musical features. *Music Perception*, Vol. 21, No. 4, pp. 561-585.

Schubert, E., & Dunsmuir, W. (1999). Regression modelling continuous data in music psychology. In: S.W. Yi (Ed.), *Music, Mind, and Science*. Seoul, Korea: Seoul National University Press, pp. 298-352.

Sloboda, J.A. (1991). Music structure and emotional response: Some empirical findings. *Psychology of Music*, Vol. 19, No. 2, pp. 110-120.

Sloboda, J.A., & Lehmann, A.C. (2001). Tracking performance correlates of change in perceived intensity of emotion during different interpretations of a Chopin Piano Prelude. *Music Perception*, Vol. 19, No. 1, pp. 87-120.

Theil, H. (1966). *Applied Economic Forecasting*. RandMcNally.

Todd, N.P.M., Cousins, R., & Lee, C.S. (2007). The contribution of anthropomorphic factors to individual differences in the perception of rhythm. *Empirical Musicology Review*, Vol. 2, No. 1, pp. 1-13.

Vines, B.W., Nuzzo, R.L., & Levitin, D.J. (2005). Analyzing temporal dynamics in music: Differential calculus, physics, and Functional Data Analysis techniques. *Music Perception*, Vol. 23, No. 2, pp. 137-152.

Vos, P.G., Van Dijk, A., & Schomaker, L. (1994). Melodic cues for metre. *Perception*, Vol. 23, pp. 965-976.

Vuust, P., Ostergaard, L., Pallesen, K.J., Bailey, C., & Roepstorff, A. (2009). Predictive coding of music - Brain responses to rhythmic incongruity. *Cortex*, Vol. 45, No. 1, pp. 80-92.

Wald, A. (1955). *Selected Papers in Statistics and Probability*. Stanford University Press.

Wiggins, G.A., Pearce, M.T., & Müllensiefen, D. (2009). Computational modeling of music cognition and musical creativity. In: R.T. Dean (Ed.), *The Oxford Handbook of Computer Music*: Oxford University Press, pp. 383-420.

Windsor, L.W. (1997). Frequency structure in electroacoustic music: ideology, function and perception. *Organised Sound*, Vol. 2, No. 2, pp. 77-82.

Wishart, T. (1985). *On Sonic Art*. York, UK: Imagineering Press.

Wishart, T. (1977). Red Bird. http://www.ubu.com/sound/wishart.html

Wishart, T. (2009). Computer music: Some reflections. In: R.T. Dean (Ed.), *The Oxford Handbook of Computer Music*. New York, NY: Oxford University Press, pp. 151-160.

Zuur, A.F., Ieno, E.N., & Smith, G.M. (Eds.). (2007). *Analysing Ecological Data*. Spring Verlag.