Comparative Time Series Analysis of Perceptual Responses to Electroacoustic Music

Freya Bailes and Roger T. Dean
University of Western Sydney, Penrith, Australia

This study investigates the relationship between acoustic patterns in contemporary electroacoustic compositions, and listeners' real-time perceptions of their structure and affective content. Thirty-two participants varying in musical expertise (nonmusicians, classical musicians, expert computer musicians) continuously rated the affect (arousal and valence) and structure (change in sound) they perceived in four compositions of approximately three minutes duration. Time series analyses tested the hypotheses that sound intensity influences listener perceptions of structure and arousal, and spectral flatness influences perceptions of structure and valence. Results suggest that intensity strongly influences perceived change in sound, and to a lesser extent listener perceptions of arousal. Spectral flatness measures were only weakly related to listener perceptions, and valence was not strongly shaped by either acoustic measure. Differences in response by composition and musical expertise suggest that, particularly with respect to the perception of valence, individual experience (familiarity and liking), and meaningful sound associations mediate perception.

Received July 27, 2010, accepted July 26, 2011.

Key words: electroacoustic music, perceptions of change, affect, sound intensity, time series analysis

Perhaps the most fundamental questions in music cognition are which acoustic features of music drive our perceptual and affective responses, and why. Previous research has demonstrated meaningful and reliable patterns between specific events in music and listener perceptions of those events (Krumhansl, 1996; Leman, Vermeulen, De Voogdt, Moelants, & Lesaffre, 2005). However, not all acoustic features map onto listener perceptions directly (Leman et al., 2005; Mathews, 1979; Sloboda & Lehmann, 2001; Stecker & Hafter, 2000), and not all musical styles have been explored. Here we specifically seek a potentially mechanistic understanding of what influence particular acoustic features may have on listener perceptions; in contrast to studies such as Coutinho and Cangelosi (2009) and Korhonen, Clausi, and Jernigan (2006) which instead use a wide range of acoustic variables to generate a complex model that can predict perceptions, notably of emotion, for purposes of music information retrieval and its commercial application.

With respect to electroacoustic music, which is not constructed on the basis of a series of discrete notes within a tonal system, little is known as to which of its acoustic properties are psychologically relevant for the listener (Bailes & Dean, 2009a; Windsor, 1995). This is echoed in the absence of established aural training methods for electroacoustic musicians (Tsabary, 2009). Our previous work has focused on this form of contemporary music, examining listener perceptions of change in the sonic texture, and corresponding ratings of affective expression (Bailes & Dean, 2009b). The stimuli used were short, artificially generated sequences, and here this work is extended to explore listener perceptions of ecologically valid, pre-existing compositions. A published example of the time series analysis of an extended electroacoustic piece details the time series methods applied in the current work (Dean & Bailes, 2010a). While referring the reader to this earlier paper, we have tried to make the present paper comprehensible in its own right. A secondary aim of the current work is to illustrate that such methods can be informative with a wider range of music (e.g., Webern piano music).

A body of research has already identified many of the acoustic properties that underpin listeners' perceptions of musical structure, where this can be defined broadly as patterns of similarity and difference. For instance, listeners segment musical events based on contrasts in textural density, registral change (Deliège, Mélen, Stammers, & Cross, 1996), attack, brightness (McAdams, Vieillard, Houix, & Reynolds, 2004), fundamental frequency, spatial position, temporal fluctuation (Grimault, Bacon, & Micheyl, 2002), and amplitude. It has been argued that late twentieth-century composition lacks meaningful

Music Perception volume 29, issue 4, pp. 359–375. issn 0730-7829, electronic issn 1533-8312. DOI: 10.1525/mp.2012.29.4.359

units with which to structure a work (Keller, 2000). Nevertheless, listeners are able to segment atonal compositions, relying on what have been described as "surface" features of the music, such as change in register, timbre, dynamics, event density, articulation, and tempo (Cambouropoulos, 2006; Imberty, 1993; McAdams, Vieillard et al., 2004). Many of the surface features found to be perceptually valid in these note-based compositions are the essential musical properties in electroacoustic music (Bailes & Dean, 2007; Tsabary, 2009). Indeed, since "deep" or hierarchical note-based structures are lacking, listeners cannot be guided by tonal and metric schemata, and online or real-time processing of so-called surface features is fundamental (Imberty, 1993). In this respect, electroacoustic compositions are an ideal medium to study the acoustic properties that shape real-time perceptions of structure, namely perceptions of change in sound.

A further level of engagement with a musical work is the extent to which it communicates emotion to the listener. A distinction can be made between the induction of emotion, and the perception of emotion (Dibben, 2004; Juslin & Laukka, 2004; Kallinen & Ravaja, 2006), and while the two frequently co-occur, research has shown that listeners are able to identify an emotion expressed in music without feeling it (Dibben, 2004). The perception of emotion is the subject of the current research, which is concerned with identifying the cues to its communication (not its induction). If emotion is perceived primarily in response to anticipated features of the music (Grewe, Nagel, Kopiez, & Altenmüller, 2007; Huron, 2006; Sloboda & Lehmann, 2001), perhaps highly schematic tonal melodies have greater communicative potential with respect to emotion than electroacoustic works that may not elicit strong expectations. The question of the reception of electroacoustic works by an uninitiated audience has been addressed by Landy (2009), and his work suggests that an understanding of a composer's intentions can greatly facilitate a listener's appreciation. Composers frequently (though not necessarily) aim to communicate through their music, and this may well include an affective "content." There is evidence to suggest that listeners not only perceive emotion in computer-processed sounds, but also that when such sounds are embedded within an otherwise instrumental work, they may be rated as the most emotionally forceful component (McAdams, Vines, Vieillard, Smith, & Reynolds, 2004). Sounds may be emotional by mere association with past affective experience (e.g., the "episodic memory" proposed by Juslin & Västfjäll, 2008) or their evocation of meaningful events (e.g., sounds reminiscent of crying, or a slamming door, Bradley & Lang, 2000). Here the familiarity of the listener with the sounds in question is key.

Based on the literature (Leman et al., 2005), it can be predicted that the greater a listener's familiarity with a musical style, the greater the emotion perceived to be expressed. That is not to say that a lack of familiarity with both a particular piece of music and its genre prevents any judgment of affective quality (Balkwill & Thompson, 1999). For example, using a stylistically unfamiliar piece of contemporary classical music (*The Angel of Death* by Roger Reynolds), Dubnov, McAdams, and Reynolds (2006) found a link between the temporal profile of listeners' judgments of emotion, familiarity ratings of the music through time, and the structure of the piece. The capacity to process unfamiliar material may well differ depending on musical experience. For example, when detecting the onset of new musical ideas in *The Angel of Death*, "nonmusicians" were said to be selecting very short ideas compared to musicians, apparently because of difficulties in constructing perceptually substantial sequences from such unfamiliar material (Lalitte et al., 2004). Indeed, musicians rated the style of the piece in question as more familiar than nonmusicians. However, there was no difference in the number of subsections that participants identified within each theme. Similarly, Deliège (1989) did not find any difference between professional musicians (knowledgeable in contemporary music) and nonmusicians with respect to their segmentation of a contemporary composition (Berio's *Sequenza VI*). On the whole, the literature seems to point to differences by musical expertise only insofar as these impact perceptual fluency and cognitive processing. Thus, a nonmusician might perceive change in sound in an unfamiliar musical style much as a musician might, though potentially at an attenuated rate. However, effects of expertise might still be expected to impact on both the cognitive and affective apprehension of unfamiliar music. For example, electroacoustic practitioners might be expected to be more sensitive to certain aspects of timbral and noise-sculpting (e.g., spectral filtering) than acoustic instrumentalists, whose potentially powerful capacity for timbral control operates in different ways.

An experiment was devised to study listeners' real-time perceptions of electroacoustic compositions, testing a set of five hypotheses. The first hypothesis, in line with the literature reviewed above, is that the musical structure that listeners perceive, measured as continuous perceptions of change in sound, is a function of the surface features of the music, specifically acoustic measures of intensity and spectral flatness. Our purpose is to develop an analysis of the impact of different acoustic features on specific perceptual parameters, by progressively complementing the statistical correlative approach of the present paper with causal intervention studies. Thus it is logical to start with "global" features such as intensity and spectral flatness. These are both among the very few parameters that can be considered relevant to virtually all kinds of music, and hence potentially widely influential. In contrast, tonality, metre, pulse, and even pitch may not be universally important. While intensity is commonly studied (e.g., Leman, et al., 2005; Sloboda & Lehmann, 2001), spectral flatness requires a little more introduction.

As described in Dean and Bailes (2010a), spectral flatness is a global parameter of timbre: every spectral component, symmetric or otherwise, impacts it. Spectral flatness is the ratio of the geometric mean to the arithmetic mean of the power spectrum. Noise is indicated by high values, while an infinitely narrow peak has a spectral flatness of minus infinity. It is one of the basic spectral audio descriptors in the MPEG-7 standard (see MPEG-7 Overview, 2004). It is noted here that low values can signal the presence of tonal components. Unlike spectral centroid, which represents the spectrum's central frequency, even symmetrical changes in the power spectrum alter spectral flatness. Other advantages to the use of spectral flatness in our work include its utility in music that is largely textural/timbral, such as electroacoustic music, as well as its suitability for the study of music that is based on discrete pitches, such as tonal Western music.

While electroacoustic compositions generally lack a tonal or metrical structure, given their lack of "note-centredness," it should not be assumed that listeners rely on surface features as cues to their musical structure in a one-to-one mapping. For example, the relationship between sound intensity and the subjective perception of loudness is biased towards an overestimation of the loudness of increasing versus decreasing intensity (Neuhoff, 2001; Olsen, Stevens, & Tardieu, 2010). Accordingly, increases in intensity, and even fluctuations of particularly high intensity, might be particularly salient and receive high ratings of perceived change in sound. Correspondingly, we have observed duration asymmetries in crescendi versus decrescendi in a wide range of electroacoustic and other music (Dean & Bailes, 2010b). As for spectral flatness, changes in this measure might not only represent a timbral dynamic, but could also be perceived as indicative of event density, again affording particularly high ratings of perceived change in sound. While acoustic intensity and spectral flatness can be easily quantified, their integrative effect on listener perceptions of musical structure is not known. Moreover, if acoustic and listener definitions of structure differ, it is important to investigate both as potential predictors of perceptions of the affective properties of music.

The second hypothesis states that the perception of arousal is a function of the acoustic intensity of the music. Arousal is used as one of two affective dimensions in this study, following in the tradition of Russell's circumplex model of affect (Russell, 1980), and Schubert's two-dimensional emotion space (e.g., Schubert, 2004). As a dominant determinant of perceived loudness (Geringer, 1995), intensity is predicted to influence perceptions of arousal much as loudness has been argued to relate to arousal (Schubert, 2004). Valence is the second of the two affective dimensions under study, measured on a continuum from very positive to very negative. Past work has identified a relationship between timbral properties and valence. For example, regression analyses by Leman et al. (2005) uncovered a link between listener ratings of roughness and valence. Countinho and Cangelosi (2009) describe a qualitative association between sharpness and valence resulting from a canonical correlation analysis. Our previous work isolated spectral flatness as an informative global measure of timbre, and found a correlation between spectral complexity and ratings of perceived positive valence (Bailes & Dean, 2009b). Accordingly, our third hypothesis is that the perception of valence is a function of the spectral flatness of the music.

Hypothesis four posits that not all perceived affect could be accounted for by the acoustic measures of intensity and spectral flatness, with different patterns emerging for different compositions. For example, the composer of one of the works studied, Trevor Wishart, deliberately emphasizes the human voice in his music (Brattico & Sassanelli, 2000), suggesting an identifiable sound source with particular affective connotations (Bailes & Dean, 2009c). The particular piece in question, *Red Bird*, is known for its narrative form, based on the transformations of common, everyday sounds (Windsor, 1995). Such semiotic associations have an affective reach beyond the acoustic measures of intensity and spectral flatness, as demonstrated in Dean and Bailes (2010a). An alternative approach to the psychophysical is provided by ecological psychology, which expresses perception as an active exchange between an organism and its environment. According to this perspective, a listener perceives not so much the acoustic qualities of the music as the information that such sounds afford. In other words, they perceive the sounds' meaning (Dibben, 2001; Windsor, 2000). Alternative ecological variables to the acoustic variables of intensity and spectral flatness will therefore be tested as substitute predictors of listener perceptions of affect.

The final hypothesis states that participants' familiarity with the music will differentially influence their reliance on physical cues to judge its affective content (Balkwill & Thompson, 1999). However, it is predicted that judgments of structure as expressed by perceptions of change in sound will be qualitatively similar for all participants, even though musical experts will be quantitatively more subtle in their discrimination than nonmusicians.

Method

PARTICIPANTS

Thirty-two participants took part, with ages ranging from 18 to 58 years (median 25.5 years). The nonmusician group (NM) was recruited from an undergraduate psychology course in exchange for 70 min of course credit (12 female, 4 male). The musician group (M) included participants selected for their music training; they were offered a CD of British-Australian jazz in exchange for their participation (4 female, 12 male). Of the musically trained, eight were deemed to be expert (EA) in contemporary music from their role as computer music composers or sound engineers. Participants completed the Ollen Musical Sophistication Index (OMSI - Ollen, 2006), where scores under 500 represent less musically sophisticated, and scores of 500 or more represent more musically sophisticated. The nonmusicians scored a mean of 136, while the musicians scored a mean of 789.

STIMULI

An excerpt of approximately 3 min duration was selected from each of four different compositions. Each was presented as a .aiff stereo audio file, 44.1kHz sampling rate, 16 bit.

Anton Webern (1937) 'Piano Variations Op. 27'. The second (*sehr schnell*) and third (*ruhig fliessend*) movements were excerpted from a performance by Glenn Gould (1964) in the film "The Alchemist." The excerpt begins at about 1'45. This piano work is included as an instrumental, pointillist contrast.

Roger Dean (2003) 'soundAFFECTS'. An excerpt from about 3'00 to 6'00 of this sound composition features filtered noise and is part of an audiovisual work for performance and for the web (Brewster, Smith, & Dean, 2004).

Iannis Xenakis (1962) 'Bohor'. A 4-track work for tape, from which 3 min of a stereo recording was excerpted. The recording was excerpted from EMF CD 003.

Trevor Wishart (1977) 'Red Bird, a political prisoner's dream'. As described in Dean & Bailes (2010a), 3 min was excerpted (22'30 to 25'30) from a recording on UbuWeb of this 45 min piece for tape, which has a strong narrative.

Practice stimulus. For use as a practice trial, a 1 min excerpt ('30 to 1'30) was taken from "Surfacing" from the CD "Reconnaissance" by Oren Ambarchi and Martin Ng.

ACOUSTIC ANALYSES

These methods, using the software Praat (Boersma & Weenink, 1992-2007), have been described in detail previously (Dean & Bailes, 2010a). In brief, intensity corresponds to unweighted sound pressure level (SPL) in dB. Spectral flatness is measured as Wiener Entropy, expressed on a log scale from 0 to minus infinity.

PROCEDURE

Participants performed a "change in sound" task and an "affect" task, in a counterbalanced order. These tasks required participants to continuously rate music while listening, rather than provide a retrospective, global rating of each piece. A substantial advantage of continuous data collection is the ability to study the evolving pattern of listener perceptions (Bailes & Dean, 2009b). For each task, participants were presented with the stimuli over *AKG K271* Studio headphones, seated at an *iMac5.1, OS10.4.11*. The screen resolution was 1680 x 1050 pixels, and the experimenter set the volume in advance.

Each of the four pieces was presented once in a random order, for each task. Participants were able to initiate each trial at their own pace by moving the cursor to a central square on the computer screen. One practice trial was included at the start of each task. After the presentation of each piece, listeners rated their familiarity with the piece and/or style of piece, and their liking of it, with these two questions asked in a counterbalanced order.

There was a break between the two tasks for participants to fill out the OMSI and a questionnaire gathering information on age, gender, nationality, familiarity with musical genres, and debrief questions.

Affect task. Instructions and training were provided on a computer screen using EmotionSpaceLab 1.02a (Schubert, 1996). A template replaced the default faces of the original program with the labels "very active," "very passive," "very positive," and "very negative."

> Participants were first given the following instructions:
>
> After a questionnaire and a short tutorial you will be asked to make judgements about emotional stimuli (that is, words, pictures and sounds). Your

> responses will be made by moving a slide on the screen using the mouse. Instructions will be provided on screen, but occasionally the researcher will ask you to read instructions on card instead of on screen. The final stage of this task is to judge the emotion conveyed by pieces of music. During the first trial, you may ask the experimenter about anything that is unclear.

They then followed on-screen directions, taking them through a training in how to use the emotion space that included words and faces, using valence and arousal axes first separately and then together.

Change in sound task. Participants were asked to indicate whether the music seemed to change. The following instructions were given:

> You are going to hear a piece of music over headphones. Your task is to detect whether the music changes and to indicate this by moving the mouse **during** any perceived change. The greater the change, the faster you should move the mouse. For example, it may be that you wish to make a scrubbing motion with the mouse to indicate a strong and sudden change in the music. The smaller the change, the slower you should move the mouse. For example, it may be that you wish to move the mouse only slightly to indicate a subtle change in the music. Please move the mouse for the **duration** of any change. If you DON'T think the music changes, keep the cursor still. Please try to maintain your CONCENTRATION throughout each piece. Your mouse movements will be recorded while the sound is playing. There is NO NEED TO CLICK the mouse while the sound is playing.

Data Analysis

The data fall into three different categories: continuous measures through time (sampled every 500 ms), continuous measures averaged across time, and liking and familiarity ratings per piece. Summary statistics are provided to reveal series means and coefficients of variation (CV: the ratio of standard deviation/mean). For the acoustic time series, all values are positive (intensity) or negative (spectral flatness); thus, their CV can be measured directly. However, for the perceptual time series, the measurement scale ranges from -100 to +100. For such series, CV is determined as a function of series constructed as the measured values plus 100, so that all series values are positive, and taken in conjunction with the mean, the CV then fairly reflects the degree of variability.

Most if not all time series representing the acoustic (intensity and spectral flatness) and perceptual (change in sound, arousal, and valence) measures require special techniques (such as our choice, time series analysis) in order to analyze the autoregressive components. Since the data are not independent, they cannot properly be investigated by the most conventional techniques (e.g., simple regression techniques), because these rely on an assumption that data points are independent (appropriate methods evaluated in detail in Dean & Bailes, 2010a). In order to make this paper as complete as possible in its own right, we offer an explanatory glossary (see Appendix) of the key terms relevant to time series analysis as applied here, and provide some pointers to primary and review references on this highly developed array of techniques; we refer readers to our earlier paper for more specific discussion of their application to musical time series. In brief, first perceptual measures were averaged across participants for each sampled time point, producing an average response profile for the duration of each piece. The issues behind this decision, which is the most parsimonious and conservative approach, are discussed in detail in Dean & Bailes (2010a). Second, data were transformed (see "differenced" in Glossary) when necessary to render each series stationary (see "stationarity" in Glossary). A series resulting from differencing *seriesname* is called *dseriesname* in this paper. ARIMA and ARIMAX models (see Glossary) were developed to assess the statistical impact of the exogenous acoustic variables on the measured perceptual variables, and the influence of autocorrelation. The Akaike Information Criterion (AIC) (Akaike, 1974) was used as the main basis for model selection (see "information criteria" in Glossary). ARIMAX allows the assessment of the influence of multiple lags of autoregression and of exogenous variables; it is well established already that many acoustic variables, such as intensity, may operate with a lag of up to about 5 s (e.g., Schubert, 2001, 2004), which correspond to 10 lags in our analyses (one lag represents 500 ms). As discussed in Dean & Bailes (2010a), ARIMAX models were sometimes improved in terms of AIC by the inclusion of lags of some variables whose coefficients individually were not significant, but only up to two of these individually insignificant lags have been allowed.

ARCH models (see "heteroskedasticity" in Glossary) were also assessed, such that the influence of exogenous variables on temporal variance as well as on the mean model could be demonstrated (Enders, 2004). Third, the time series were modeled using vector autoregression (see "VAR" in Glossary; also Enders, 2004; Hamilton, 1994) to assess conservatively the inter-relationship between all the acoustic and perceptual measures, and to provide impulse response functions (see Glossary) represented as fractional error variance distribution estimates (FEVD: as discussed in Dean & Bailes, 2010a) that reveal which of the factors (acoustic and perceptual) appear as statistically significant influences on a given perceptual response. Confidence limits for the FEVD were determined by bootstrapping the residuals, since this does not require prior knowledge of their distribution (which is not available). The test of Granger Causality (Granger, 1969, and see Glossary), which depends on these VAR analyses, was also used to infer the plausible statistical significance and directions of these relationships. All the models discussed here showed white noise residuals, the key criterion of their appropriateness (see "criteria of satisfactory times series models" in Glossary). All the coefficients in the ARIMAX /ARCH models presented in the tables were significant at $p < .05$ unless otherwise indicated.

The continuous response data were also averaged across time for each participant, for comparison with ratings of liking and familiarity per piece.

Results

Dean and Bailes (2010a) illustrate the effective prediction of perceptual time series by ARIMA(X) modeling in the case of the piece by Wishart, by tracing the predicted (modeled) time series and comparing them with the measured ones. Here some further illustrations of modeled time series are provided, but we mainly summarize the key parameters of the models in tabular form.

ACOUSTIC CHARACTERISTICS OF THE PIECES

Table 1 summarizes the mean and coefficient of variation for the acoustic features of the pieces under study. Both Xenakis and Dean are electroacoustic pieces with relatively homogeneous textures that change gradually – in the former case with metallic bell-like sounds, in the second with noise filtered by a continuing succession of waves. Wishart's piece, as discussed in Dean and Bailes (2010a), involves much greater dynamic contrasts, and includes overlapping segments of human and animate sounds as part of its long-term narrative structure. The Webern piano work involves rather continuous rhythmically repetitive patterns with significant dynamic change. As expected, it shows the lowest spectral flatness (i.e., narrow peaks in the spectrum) corresponding to its relative timbral purity, though with quite a substantial coefficient of variation. Xenakis and Dean show the lowest coefficients of variation for both acoustic parameters. Intensity and spectral flatness are not significantly

TABLE 1. Means and Coefficients of Variation for the Intensity and Spectral Flatness Measures of the Four Stimuli

	Intensity (dB)		Spectral Flatness (Wieners)	
Piece	*M*	CV	*M*	CV
Wishart	52.45	0.25	-9.06	0.20
Xenakis	60.53	0.04	-6.83	0.07
Dean	63.74	0.04	-3.73	0.12
Webern	49.27	0.20	-10.13	0.16

cross-correlated in any of the works, as judged after prewhitening (see Chapter 11 of Box, Jenkins, & Reinsel, 1994, and Glossary), and hence were potential useful exogenous variables for the intended analyses of their influence on perceptual responses.

PERCEPTIONS OF CHANGE IN SOUND

Table 2 summarizes the basic characteristics of perceived change during the four pieces.

Time series analyses demonstrate that listener perceptions of change in sound are a function of the intensity profile of the music for all four pieces ($p < .001$ using Granger Causality Wald tests (Wald, 1955) for Wishart (model $df = 3$), Xenakis (model $df = 2$), Dean (model $df = 4$), and Webern (model $df = 10$). The Granger Causality tests confirm that there is no statistical evidence of the reciprocal causality (the "endogenous" perceptual variable causing the "exogenous" acoustic variable – see Glossary). Note that the stationarized versions of each variable are used in the Granger Causality tests. Table 3 summarizes the ARIMA(X) model results, and compares the best models based solely on autoregression (ARIMA) with those using the exogenous variable also (ARIMAX). The Table indicates which series have been differenced, by prefacing the variable name with 'd.' Note that the AIC and *SSE* (sum of squared errors) can only be compared between different models of the same perceptual series, while the correlations provide some crude comparison between the quality of models of different series. In summary, Table 3 shows that the AIC was lower for ARIMA(X) models

TABLE 2. Means and Coefficients of Variation of the Change in Sound Ratings for the Four Pieces

	Change in Sound	
Piece	*M*	CV
Wishart	0.14	0.74
Xenakis	0.08	0.41
Dean	0.21	0.49
Webern	0.11	0.60

TABLE 3. ARIMA(X) Time Series Models for Perceived Change Based on Acoustic Variables

Piece and modeled (stationary) variable	Model autoregressive and predictor inputs	AIC	Model: Sum of squared errors for the whole series	Observed correlation for the whole series (model: observed)	Ex post prediction of last 100 time points (correlation model: observed)
Wishart dchange	ar(1,2,3)	-1160.8	1.2	0.33	0.28
	(L1,2).dintensity, ar(1,2,3)	-1299.2	0.8	0.62	0.59
Xenakis change	ar(1,2,*10*)	-2006.7	0.1	0.80	0.62
	(L1).intensity, ar(1,2) constant	-2029.0	0.1	0.82	0.65
Dean dchange	ar(1/5) ma(6,17)	-1230.9	0.7	0.42	0.30
	(L1,2,3).dintensity, ar(1,2,3,4,5) ma(6,9)	-1248.0	0.6	0.49	Theil's U >1
Webern dchange	ar(1,2,3) ma(4,5,9,10)	-1432.4	0.5	0.41	Theil's U >1
	L(*0*,1).dintensity, ar(1,2,3) ma(4,5,9,10)	-1580.4	0.3	0.66	0.63

Note. ar(n) indicates the influence of autoregressive lag n; Ln indicates the influence of lag n of the specified exogenous (predictor) variable; and ma(n) indicates a moving average error autoregression of lag n. Any variables which are not significant at $p <. 05$ are in italics. All models presented are significant at $p <. 0001$. Correlations for 'ex post forecasts' are provided whenever Theil's U (see Glossary) is < 1, i.e., the forecast is better than a naïve method forecast. Ex post forecasts involve the prediction of time points that have not been used to develop the model, and are expected to be poorer than those for the modeled data.

than for ARIMA models, confirming the substantial influence of the addition of intensity as an exogenous variable, which was also associated with higher correlations between modeled and observed data.

Figure 1 (see color plate section) illustrates the influence of intensity on perceived change in sound for each of the four pieces, using the best ARIMAX model (with intensity as the exogenous variable).

Spectral flatness measures of the music did not Granger-cause ratings of change in sound in any of the pieces studied, and correspondingly they were not useful predictors in ARIMAX or VAR models.

PERCEPTIONS OF AFFECT (AROUSAL AND VALENCE)

Table 4 provides summary statistics for the perceptual response time series.

The low CV for arousal in Xenakis and Dean parallels the low CV for the intensity profile of these pieces in comparison with the other two, suggesting that acoustic properties could only have limited impact on arousal in these pieces. Accordingly, arousal and valence were in most cases not Granger-caused by the measured acoustic properties of the music (intensity or spectral flatness) for either Xenakis or Dean when VARs were run with each pair of acoustic and perceptual variables. Given that the CV for spectral flatness is higher for Dean than Xenakis, it is perhaps not surprising that the only candidate exception was that dspectral flatness was just significant in Granger causing darousal for Dean ($p = .05$, model $df = 2$), but as Table 5 shows, it did not provide a useful ARIMAX model, as evidenced by Theil's U values larger than one (indicating that the model is no better than a Naïve Method 1 forecast).

TABLE 4. Summary Statistics for the Real-time Perceived Affect Responses

Piece	Arousal *M*	Arousal CV	Valence *M*	Valence CV
Wishart	11.89	0.19	-24.46	0.33
Xenakis	14.83	0.04	-6.32	0.08
Dean	28.84	0.07	-18.84	0.11
Webern	9.40	0.16	9.54	0.05

As hypothesized, and in accord with the larger CV of the acoustic properties, particularly intensity profiles of the Wishart and Webern pieces, Granger Causality measures showed that intensity significantly shaped the perceived arousal of both Webern (model $df = 5$) and Wishart (model $df = 4$) compositions ($p < .001$). Table 5 summarizes the salient ARIMA(X) results, and illustrates that models of darousal for both Webern and Wishart pieces were improved by the inclusion of intensity, seen by lower AIC values for ARIMA(X) than ARIMA models. There was reciprocal Granger Causality between darousal and dspectral flatness for Webern ($p <. 001$ in both senses, VAR model $df = 5$). Given that dspectral flatness is necessarily exogenous, we proceeded only to model its influence on darousal in ARIMA(X).

TABLE 5. ARIMA(X) Models of Perceived Affect, with One Exogenous (Acoustic) Variable

Piece	Modeled (stationary) variable	Model autoregressive and predictor inputs	AIC	Model: Sum of squared errors for the whole series	Model: Observed correlation for the whole series (model: observed)	Model: Ex post prediction of last 100 time points (correlation model: observed)
Wishart	darousal	ar(1,3,5,14,15)	1282.6	594.6	0.55	Theil's U >1
	darousal	L(1,2,3,4,5,6,7,8,9,10,*17*). dintensity ar(1,3) ma(20)	1112.8	396.1	0.70	Theil's U >1
	dvalence	ar(1)	1171.1	460.5	0.65	0.73
	dvalence	L(1,2,3,4,5,6,7,8,9,10,11,12,*15*) .dspectralf, ar(1,4)	1062.6	342.3	0.68	Theil's U >1
Dean	darousal	ar(1,6) ma(7)	993.0	Theil's U >1	Theil's U >1	Theil's U >1
	darousal	L(2,3,4,*5*).dspectralf, ar(1,6) ma(7) noconst	964.6	Theil's U >1	Theil's U >1	Theil's U >1
Webern	darousal	ar(1) ma(5,8,10)	1451.4	973.1	0.60	0.43
	darousal	L(0,1,2,3,4,5,6,15).dintensity, ar(1) ma(5,8,10)	1328.9	747.6	0.68	Theil's U >1
	darousal	L(1,2,3,*4*,*5*).dspectralf, ar(1) ma(5,8,10,23)	1398.5	862.7	0.65	0.39

Note. Lags which are not significant at $p <. 05$ are in italics. Note that we assessed possible models of both darousal and dvalence (indicated in column 2) for each piece, and nonsignificant models are not displayed.

For perceived valence, the literature provided no strong basis for a predicted relationship with either measured acoustic factor, and for Xenakis, Dean, and Webern, no Granger Causalities were observed. Perceptions of valence in the composition by Wishart were Granger Caused by the spectral flatness profile of the music ($p < .005$, $df = 4$), though the reciprocal causality was also just significant (at $p < .05$). Moreover, the ARIMA(X) model of dvalence for Wishart, with spectral flatness as an exogenous variable, is only a marginal improvement on the ARIMA model, further failing to satisfy Theil's U in an ex post prediction (see Table 5).

VECTOR AUTOREGRESSIVE ANALYSES: A) FURTHER APPROACHES TO MODELING PERCEIVED VALENCE

As detailed in Dean and Bailes (2010a), VAR permits a conservative approach to assessing the possible mutual influences between the acoustic and perceptual variables, which necessarily occur simultaneously. Even though the acoustic variables are by definition exogenous (they cannot be influenced by the perceptual variables), this statistically conservative approach allows for such a possibility, and its occurrence might give cause for question and statistical concern. We commence with VAR assessments involving both acoustic variables taken together in relation to each perceptual variable in turn. For Granger-causal relationships the fractional error variance decomposition (FEVD; see impulse response function – IRF – in Glossary) is estimated. Put simply, this is an assessment of the influence of unit change in each predictor upon the specified variable, and influences in every possible direction (including endogenous variable upon exogenous acoustic variable) can be considered. Those that indicate an influence whose bootstrapped confidence limits do not include zero are significant and are presented. Some examples of FEVDs are shown, and others are summarized in terms of the lag 8 (representing 4 s) FEVD value for a given interaction between variables, providing its confidence limits confirm that it is significant. This is of particular interest for the assessment of influences on perceived valence, which for three of the pieces in this study are poorly modeled by the approaches so far. Table 6 shows that there is Granger causality for Wishart valence perception, though the corresponding FEVD coefficient is not significant.

VECTOR AUTOREGRESSIVE ANALYSES: B) ECOLOGICAL TIMBRAL VARIABLES

In previous work on Wishart's *Red Bird* (Dean & Bailes, 2010a) we began to analyze specific ecological contributions to the detected impact of spectral flatness upon perceived valence, and its heteroskedasticity. We demonstrated a clear cut influence of animistic sounds upon valence, which indeed largely subsumed the impact of spectral flatness. In other work on NoiseSpeech (Bailes & Dean, 2009c) and NoiseSong (Bailes & Dean, 2011), we have assessed the perceptual clustering and detected affective correlates of vocal and vocal-related

TABLE 6. Impulse Response Functions from VAR Analyses of Valence

Piece	*df* (model)	Significant Granger causalities	Significant FEVD values (lag 8)
Wishart	4	dintensity -> dchange***	0.28
	4	dintensity -> darousal***	0.21
	4	dspectralf -> dvalence*	
Xenakis	2	dintens -> dchange***	0.30
	2	dspectralf -> dchange***	
Dean	5	dspectralf -> darousal*	
	5	dintensity -> dchange****	0.13
Webern	5	dintensity -> darousal***	0.17
	5	dintensity -> dchange**	0.29
	5	dspectralf -> dchange**	0.04

Note. Only models in which either or both Granger Causality and lag 8 FEVD values are significant are summarized. For Granger Causality: * $p < .05$, **$p < .01$, ***$p < .001$; for FEVD, only values in which the 95% confidence limits do not reach zero are shown.

acoustic and electroacoustic sounds. Given this, one of us (FB) observed that elements of *soundAffects* might be construed as related to vocal sounds, even though the composer (RTD) did not use vocal sources. Therefore, we decided to investigate briefly the possible impact of perceived vocal quality on the affective responses to this piece.

Three musicians (FB and two others) continuously rated the perceived vocal quality (ranging from "0" = "no voice" to "100" = "some vocal quality" to "200" = "pure voice"), using the same computer interface as for change and affect. Our raters reported substantial and changing vocal quality (mean 51, range 18 to 100). Since our raters were musicians, we assessed whether their continuous perceptions of vocal quality constituted significant predictors of either the M or EA group responses. We performed VARs using the differenced VocalQuality mean series and dspectral flatness together to model either the darousal or the dvalence series for both participant groups. In accord with our previous work on NoiseSong, there were no Granger-causal relations of valence with dVocalQuality or dspectral flatness, and arousal was related with dVocalQuality ($p < .05$), but not simultaneously with dspectral flatness.

VECTOR AUTOREGRESSIVE ANALYSES: C) MULTIVARIATE ANALYSES

Finally, multivariate VAR analyses were done involving all acoustic and perceptual variables. The results are presented in this section for those cases in which there is Granger Causality as judged by the complete VAR model (with all variables simultaneously assessed). Consistent with the analyses presented in Dean and Bailes (2010a), for the Wishart piece Granger causality is observed in the multivariate VAR of darousal upon dchange. This means that there was a statistical indication that darousal might influence dchange, all other interactions considered. We express this in shorthand as "darousal -> dchange." Similarly, we observed the following Granger causalities for Wishart: dintensity -> dchange, dvalence -> darousal, dintensity -> darousal, and dspectralf -> dvalence (in each case $p < .01$ and model $df = 4$). IRF analyses with bootstrapped error estimations confirm that only some of these causalities produce substantial and significant FEVD values, in agreement with the earlier analyses. These are presented in Figure 2. If several variables provide significant FEVD influences upon a particular other variable, this indicates that they are expected to do so independently of a change in each other. Note that the influence of dspectral flatness on dvalence is not significant in this analysis, given the inputs of the other perceptual and acoustic variables, and in concordance with its lack of significant FEVD in Table 6.

For the Webern, Granger causality applies to dintensity and dspectralf -> dchange, dintensity and dchange -> darousal, darousal -> dvalence ($p < .05$, model $df = 5$). Of these, the statistically significant step 8 values are shown in Figure 2. For the Xenakis piece, significant Granger causalities are: dintensity and dspectralf -> dchange ($p < .05$, model $df = 4$). Figure 2 shows that only the IRF step 8 value for dintensity as an impulse upon dchange was significant. For the Dean piece, Granger causalities are: dintensity -> dchange, and dvalence -> darousal ($p < .01$, model $df = 5$). Figure 2 shows the only corresponding significant FEVD value (dintensity on dchange).

Perhaps most salient among these highly multivariate analyses are the indications that perceived arousal and perceived change are in some cases small but significant mutual influences. In these cases, listeners' perceptions of arousal

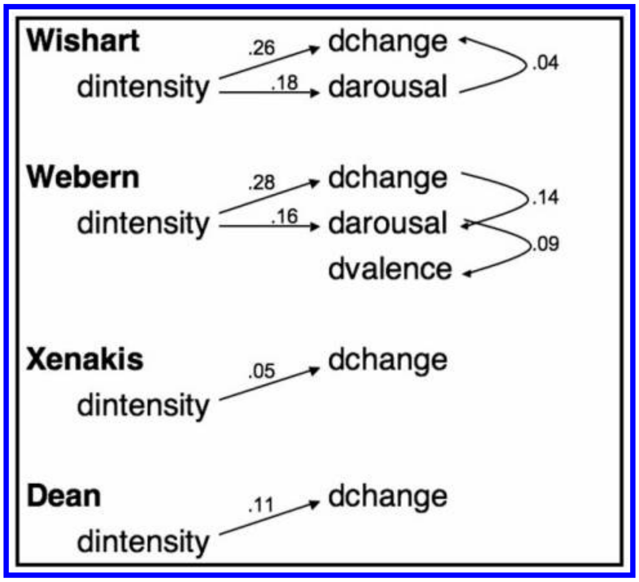

FIGURE 2. Figure highlighting the statistically significant FEVD values (at step 8) resulting from IRF analyses with bootstrapped error estimations. Arrows show which variables impact on others.

in the music are underpinned by its perceived structure. In addition, using the most conservative analyses possible, the results confirm that intensity is a key influence on change and arousal. Valence is probably influenced mostly by factors other than intensity or spectral flatness, as demonstrated specifically for the animistic sounds in the Wishart piece, discussed in Dean and Bailes (2010a).

VAR analyses have also been undertaken using the undifferenced series, as discussed in our method development paper (Dean & Bailes, 2010a), and the results from these, which do generally permit achievement of white noise residuals for the response series, are consistent with those described already, and are not presented separately.

INDIVIDUAL DIFFERENCES

Analyses so far have been conducted on the mean response series across all participants. Figures 3 and 4 (see color plate section) show that this is entirely reasonable in that the separate curves for the three participants groups (nonmusicians, musicians, or expert musicians) are very similar for each parameter for the Wishart and Webern pieces, chosen to represent respectively sound-based and note-based instrumental extremes. In the case of the fine-grained perception of change, the graphs only show the first 100 s, for clarity, and as with the other figures in which this is the case, this section is representative of the whole. It is interesting that in the case of perceived change, there are clear differences in the scale of response (which would be minimized by standardizing the series) with, for example, the expert group showing attenuated responses to the Webern in comparison with the other two groups.

Indeed, we expected differences between participant responses, and that these would be informative. To this end we further analyzed responses to Wishart and Webern. For each group and perceptual response we measured the standard deviation across participants at each time point to construct new series. We undertook VARs with stationarized series (first difference) derived from the standard deviation series, and assessed Granger Causality when response series, intensity, and spectral flatness were included. We considered it reasonable that influences on the mean model for a perceptual parameter might be distinct from (or overlapping) influences on the variability between individuals. Regarding change, Table 7 shows that intensity was important for both series, while spectral flatness was not. Spectral flatness was causal for variance of valence of both pieces. Neither feature influences the arousal variance of Webern, while spectral flatness again has a role in relation to the arousal variance of Wishart.

EXPERTISE AND ITS INFLUENCE ON PERCEPTUAL RESPONSES

Table 8 summarizes the group mean responses and coefficients of variation for the three perceptual variables. There were sometimes differences in the autocorrelation structure of the series from the different participant groups, meaning that the number of lags (in the range 2-4, representing 1-2 s) over which previous information influenced the ongoing response also varied: this can be viewed as representing variations in speed of response, or of the time in which information is cognitively incorporated. Related to the second of these interpretations, it was hypothesized (see Introduction) that subtlety of discrimination, particularly for change in sound, might follow the order E > M > NM. For two of the pieces, change sensitivity (as judged by CV), fitted this hypothesis. For the third (Wishart), E > M ~= NM. For the fourth (Dean) there was little difference among the three groups. Perhaps correspondingly, the NM group shows the highest mean value for change in each case. Given the evidence above that intensity significantly influences perceived arousal, the E group arousal may be understood as reflecting the CV for the intensity time series of these pieces: thus the E arousal CVs for Wishart and Webern are both high (c.f. high intensity CV), while Dean and Xenakis are low (c.f. low intensity CV).

It is notable in the data above that mean valence is poorly explicable on the basis of intensity and spectral flatness, though more specific timbral parameters of agency seem powerful predictors for the Wishart piece. In parallel with this, CVs for valence responses in Table 8

TABLE 7. Granger Causalities for the Impact of dintensity and dspectral Flatness Upon the Stationarized (Differenced) Time Series of Interindividual Variation (as Standard Deviation) for Webern and Wishart Perceptual Responses

	Webern			Wishart		
	Change (*SD*)	Arousal (*SD*)	Valence (*SD*)	Change (*SD*)	Arousal (*SD*)	Valence (*SD*)
dintensity	0.000	*ns*	*ns*	0.000	*ns*	0.001
dspectral flatness	*ns*	*ns*	0.013	*ns*	0.008	0.002

Note. p values are shown for significant causalities in the 3-endogenous variable VARs; *ns* = not significant.

are also highest for the Wishart for all participant groups, but data discussed so far provide no obvious explanation unless it is the animistic sounds that convey this high CV for valence. Given their influence on the variance of valence demonstrated in the preceding paper (Dean & Bailes, 2010a), this is highly plausible. It is also consistent with the influence of spectral flatness on interindividual valence variability (Table 7).

LIKING AND FAMILIARITY

Familiarity ratings for each piece were gathered from each participant according to the following scale:

"I have never heard anything like this before."

"I have heard something like this but not this piece before."

"I have heard this piece."

"This piece is very familiar to me."

"I often listen to this piece of music."

TABLE 8. Perception of Change and Affect by the Different Participant Groups

		Change in sound		Arousal		Valence	
Piece	Group	*M*	CV	*M*	CV	*M*	CV
Wishart	E	0.07	0.95	7.99	0.28	-17.79	0.26
	M	0.18	0.79	15.60	0.19	-20.73	0.28
	NM	0.18	0.81	11.98	0.17	-29.66	0.41
Xenakis	E	0.02	1.18	41.82	0.07	9.41	0.07
	M	0.07	0.75	25.73	0.08	1.06	0.06
	NM	0.11	0.47	-4.11	0.05	-17.88	0.12
Webern	E	0.02	0.89	1.36	0.22	10.95	0.07
	M	0.10	0.75	18.37	0.17	7.43	0.06
	NM	0.15	0.64	8.88	0.14	9.87	0.07
Dean	E	0.10	0.56	49.83	0.11	-6.93	0.10
	M	0.20	0.62	40.95	0.13	4.34	0.09
	NM	0.27	0.55	12.29	0.09	-36.38	0.21

Note. E = Expert musicians; M = musicians; NM = nonmusicians.

Table 9 shows the distribution of familiarity responses for each of the four pieces of music, subdivided according to expertise.

Interestingly, for the M and NM groups, mean valence was significantly correlated with both liking, $r(30) = .54$, $p < .01$ (M group) and $r(62) = .47$, $p < .001$ (NM group), and familiarity, $r(30) = .34$, $p < .05$ (M group) and $r(62) = .26$, , $p < .05$ (NM group), for the relevant piece. Thus, local timbral and intensity structure are not the major influences on valence (as above), but enculturated experience and stylistic preference may be. Notably, liking and familiarity per se do not strongly correlate for any group. There were no significant correlations for the E group, or for arousal/liking, and arousal/familiarity.

Discussion

All five hypotheses were supported from our findings, but the results revealed considerable complexity, and hypothesis three – that valence is a function of spectral flatness – received very limited support. Rather, it seems that more specific timbral parameters, and particularly those of ecologically identifiable features, can subsume most of the effect of spectral flatness. We consider each hypothesis in more detail here.

A highly robust finding in this study is that perceived structure (change in sound) is a function of the acoustic intensity of the music, but not its spectral flatness. The impact of intensity on ratings of structure is consistent with a related finding by Leman et al. (2005), who found that the structural cues identified by musicologists in their study were best explained by measures of loudness and articulation. It is striking that a property of what has traditionally been considered the musical surface can be instrumental in shaping listeners' perceptions.

In partial support of our second hypothesis, results show that perceptions of arousal were also strongly influenced by the intensity pattern of the music for two pieces out of four (Webern and Wishart). This is consistent with work by Schubert (2004), who found that loudness was positively associated with changes in perceived arousal, and our recent findings (Dean, Bailes, &

TABLE 9. Count of Familiarity Ratings for Each Piece by the Different Participant Groups

Piece	Group	I have never heard anything like this before	I have heard something like this but not this piece before	I have heard this piece	This piece is very familiar to me	I often listen to this piece of music
Wishart	E	0	6	1	1	0
	M	1	7	0	0	0
	NM	12	4	0	0	0
Xenakis	E	0	7	1	0	0
	M	1	7	0	0	0
	NM	15	1	0	0	0
Webern	E	0	7	1	0	0
	M	0	8	0	0	0
	NM	5	10	1	0	0
Dean	E	0	7	1	0	0
	M	1	7	0	0	0
	NM	12	4	0	0	0

Note. E = Expert musicians; M = musicians; NM = nonmusicians

Schubert, 2011) of a causal relationship of intensity on perceptions of arousal, where one same intensity pattern has a similar impact on ratings of arousal, regardless of the musical content. Clearly, however, perceived arousal cannot always be explained in terms of sound intensity. Perceived arousal in Dean and Xenakis was not significantly shaped by intensity. It could be that the more limited variation in this acoustic parameter for these two pieces, as measured by intensity CV, is responsible for this finding. Indeed, in Dean et al. (2011), we superimposed the intensity profile of another piece of music onto the Dean composition, and subsequently found that this more varied intensity influenced perceptions of arousal. Musical experts appear to mirror this pattern, exhibiting large variation in arousal ratings for those pieces with the broadest range of intensity (Webern and Wishart), but not for those pieces with the smallest range (Dean and Xenakis). An analysis of Dean's *soundAffects* revealed a relationship between continuous ratings of the vocal quality of the music and its affect. This more ecological feature was found to relate significantly to perceptions of affect: dynamic changes in vocal quality were Granger-causal with ratings of perceived arousal.

The hypothesis that perceptions of valence would be a function of spectral flatness received only limited support. Similarly to Schubert (2004), who examined spectral centroid, no consistent influence of timbre was found, and valence was less well explained in terms of measured acoustic properties than the dimension of perceived arousal. Korhonen et al. (2006) had a similar finding, with only 21.9% of their model variance for valence explained by a raft of musical features, versus 78.4% for arousal. Interestingly, these authors speculated that models of valence might to some extent depend on arousal, while perceptions of arousal can be modeled independently. Our multivariate VAR analysis of Webern supports this idea, showing that darousal had a significant impact on dvalence, but not vice versa (Figure 2).

While spectral flatness did not predict mean ratings of valence in any piece in the current study, analyses of the variability of participant response for Webern and Wishart revealed that spectral flatness of the music influenced interindividual valence variability. Valence was found to correlate with overall liking for the piece, which is a pattern also observed by Grewe et al. (2007). However, this was not the case for expert respondents. Given that piece and expertise differences in valence were found, it seems that other factors underline listener judgments of this affective parameter. This is consistent with Leman et al. (2005), who reported that valence ratings differed more across participants than ratings of activity (arousal).

Schubert (2001) has suggested that greater intersubject agreement for arousal than valence as measured by his two-dimensional emotion space might indicate that participants prefer the former, or somehow respond to this dimension by default. However, it is plausible that a more predictable response pattern for arousal than valence is more than a response bias to the measurement tool. Wishart's *Red Bird (A political prisoner's dream)* is arguably stronger in "mimetic discourse" than "aural discourse" (Emmerson, 1986), and as demonstrated in Dean and Bailes (2010a), the evocation or otherwise of animal sounds more directly relates to ratings of valence than either physical parameter considered here. According to Bradley and Lang (2000), the valence of sounds serves as an index of a motivational system; appetitive or defensive. Motivation to respond to animal sounds would be greater than any propensity to respond to either intensity or spectral flatness in the abstract.

Overall, physical measures of the music were less able to account for listeners' perceptions of affect than for their perceptions of change in sound. As argued above, the inevitable limits to the range of each acoustic parameter exploited in the compositions might in turn limit the extent of emotional expression possible (Schubert, 2004). For some of the compositions, sounds with familiar associations were dominant, and these would be expected to signify beyond the intensity and spectral flatness information (Erlich, Lipp, & Slaughter, 2007; Windsor, 1995). It is worth reflecting that sound recognition (Brattico & Sassanelli, 2000) is a primary concern of the auditory system, and an organism's perceptions of its environment are inherently meaningful according to ecological accounts (Clarke, 2005; Dibben, 2001; Windsor, 2000). Ambiguous sounds (Brattico & Sassanelli, 2000) with respect to their genesis might in themselves be perceived as particularly affective, due to uncertainty. Again, acoustic measures do not capture such perceptions.

Inevitably, the listener brings her or his particular sonic experience to bear in perceiving change in sound, and particularly in perceiving affect (Cupchik, 2001). As hypothesized (and unlike De Vries, 1991), musical background and familiarity with the music(al style) impacted listener perceptions of affect, but less for perceptions of change in sound. For musicians and nonmusicians, mean valence ratings positively correlated with their ratings of liking and their ratings of familiarity with the music. This is consistent with the established relationships in the literature between positive affect and liking, and positive affect and perceptual fluency (e.g., Witvliet & Vrana, 2007). However, the expert musicians did not display this pattern, and it could be speculated that their greater experience equipped them to dissociate expressed affect from their personal familiarity with and liking for the music. Future research should attempt to disentangle this finding, controlling for the inevitably higher levels of familiarity of the expert group, and hence the limited variance in the correlation data.

The finding of a lesser influence of musical background and familiarity on perceptions of change in sound is consistent with studies employing other measures of perceived surface structure, such as similarity ratings of thematic subsections, in which no important differences were found between musicians and nonmusicians (e.g., McAdams, Vieillard et al., 2004). However, as predicted, there was some evidence (from two of the pieces) that the subtlety of discriminating change in sound may increase with musical experience. Extensive additional work is underway to determine whether experts really perceive more fine-grained changes in sound than those with less expertise, or whether they simply indicate their response with a greater range of movement, implying greater refinement.

In conclusion, the perception of the basic structural parameter of change in sound can be predicted well by acoustic measures of the music, in particular intensity. This is so regardless of participants' level of musical expertise, in line with previous studies of structural perception (Vines, Krumhansl, Wanderley, & Levitin, 2006). However, the perception of affect is additionally mediated by the listener's interpretation of the meaning of the work, by their expertise, and by their familiarity with the music. Other instances of listeners responding to music primarily in terms of their liking rather than its acoustic properties have been reported (Lalitte & Bigand, 2006). Liking and acoustic properties are united by the bimotivational structure of appetitive and defensive responses to sound proposed by Bradley and Lang (2000). The current study limited its investigation to just two acoustic properties of sound. It is surely the case that listener perceptions of both structure and affect (Gabrielsson & Lindstrom, 2001) are optimally explained by the contribution of many different factors. Our time series analysis approach provides a powerful way of delineating further, as we have already shown by identifying roles for animate and vocal-like sounds in two pieces.

Author Note

This research was supported by Australian Research Council Discovery grant DP0453179. Our thanks to Mary Broughton and William Dunsmuir for assistance with data analysis, and to the *New Music Network* and *SOMA* for contributing CDs as a thank you to our participants.

Correspondence concerning this article should be addressed to Dr. Freya Bailes, MARCS Institute, University of Western Sydney, Locked Bag 1797, Penrith NSW 2751, Australia. E-MAIL: f.bailes@uws.edu.au

Appendix – Glossary

Some of the Key Terms in Time Series Analysis.

As requested by the journal, we seek to make this paper as free-standing as possible, so that by means of this Appendix the reader may be able to grasp the key concepts of the time series analyses we apply. These terms and methods are discussed more fully in our preceding methods development paper (Dean & Bailes, 2010a) and in the books referenced therein; we refer the reader to this earlier paper for further understanding. Many of the explanations given here are quoted directly from our earlier paper; they are presented in alphabetical order.

ARIMA(X): autoregressive integrated moving average regression. X refers to the addition of exogenous variable(s) to the autoregression-only ARIMA.

Criteria of satisfactory Time Series Models: 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. NB. If there were still autocorrelation in the residuals from the model, this would represent unmodeled information. See also "Theil's U."

Dickey-Fuller test for stationarity: The Augmented Generalized Least-Squares test for stationarity (Dickey & Fuller, 1979) uses interpolated critical values derived from empirical data. 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 nonstationary series, this expectation is not true. The "augmented" part of the test allows for its autocorrelation structure, and the critical values are empirically derived to enhance the power of the test.

Differencing: This means calculating the difference between successive values in a time series, thus creating a new series with one fewer point than the original. A differenced linear time series without (measurement) error has a constant value, and differencing a time series that 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. It is important to apply specific tests of stationarity, such as versions of the Dickey-Fuller test (q.v.), and multiple differencing may be used if necessary (this was not required in the present work). 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. For a series *seriesname*, we name the first difference series *dseriesname*.

Endogenous and Exogenous variables: These are statistical analogues of the dependent and independent variables of psychology. In the present work, perceptual features were the endogenous, and acoustic the exogenous variables, but in some aspects of VAR (q.v.) the most conservative statistical approach treats them all equally, as if they are endogenous i.e., as if they are all capable in principle of exerting mutual influence.

Granger Causality: A statistical correlation that may be indicative of a causal relationship (Granger, 1969).

Heteroskedasticity: A stationary series (q.v) may still show transient changes in variance that are not sufficient to breach the overall criterion of weak stationarity. Such variance is described as heteroskedasticity (sometimes spelled heteroscedasticity). This variance may be 'conditional', in other words, influenced by changes in an exogenous variable, as in ARCH models (Autoregressive Conditional Heteroskedasticity models).

Impulse Response Function (IRF): After a VAR, this analysis assesses a forecast-error variance decomposition (FEVD) (Lütkepohl, 2007), which is an indication of the impact of a unit change in each endogenous variable on a given output, lag by lag, independent of the impact point in the time series. Similar "dynamic multipliers" may be calculated in IRFs for exogenous variables.

Information Criteria: These estimate the quality of a model, by using both its precision (based on the likelihood estimate), and penalizing for the number of parameters involved. The number of parameters dictates the degrees of freedom consumed by the model (as shown in the tables). Only Information Criteria values for a specific individual data set prediction can be compared; and lower values indicate a better model. We routinely determine both the Akaike (AIC) and Bayesian Information Criterion (BIC), and the BIC can penalize for the number of parameters more strongly than does the AIC. In most cases, the BIC and AIC lead to the same interpretations with our data.

Pre-whitening: A technique used to provide a valid assessment of cross-correlation between series. It comprises establishing a purely autoregressive statistical time series model of one series, so that the residuals from the model are white noise, i.e., free of autocorrelation (tested with a measure by Bartlett, 1966). The resultant autoregressive model, in terms only of its autoregressive lag structure and coefficients, is then used to model the second time series, generating a further time series of residuals from the second series. 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 the first series. 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.

Stationarity: To minimize spurious detection of correlation between pairs of time series data, which may be solely due to shared autocorrelation structure, "weak stationarity" needs to be established, if necessary by transformation of the original series. This means in essence removing trends in the data, such that mean, variance and covariance are all unaffected by a change of time origin. Stationary series may still possess some autocorrelation (reflected in the constant covariance structure). It is important to apply specific tests of stationarity, such as versions of the Dickey-Fuller test (q.v.). When cross-correlation is to be directly investigated, pre-whitening (q.v.) may be required.

Theil's U: this test determines whether the model under consideration 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 (see Theil, 1966).

Time Series Analysis: a large and highly developed battery of methods specific to data that are not independent, often organized sequentially in time, autoregressive and time dependent, which thus contravene the assumptions of most statistical approaches

VAR: Vector autoregression. VAR co-relates multiple time series simultaneously, treating them either as potentially mutually influential ("endogenous", q.v.), or potentially independent of others, rather input ("exogenous", q.v.) variables. We adopt primarily the statistically less restrictive yet more conservative approach in which variables are all treated as endogenous. Comparative VARX studies in which the acoustic variables are treated as exogenous (X) are made in each case; these simplify the vector decomposition involved in the computations.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*, 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, 36*, 83-93.

Bailes, F., & Dean, R. T. (2009a). Empirical studies of computer sound. In R. T. Dean (Ed.), *The Oxford handbook of computer music* (pp. 473-492). New York: Oxford University Press.

Bailes, F., & Dean, R. T. (2009b). Listeners discern affective variation in computer-generated musical sounds. *Perception, 38*, 1386-1404.

Bailes, F., & Dean, R. T. (2009c). When is noise speech? A survey in sonic ambiguity. *Computer Music Journal, 33*(1), 57-67.

Bailes, F., & Dean, R. T. (2011). The perceived affective expression of computer-manipulated sung sounds. *Computer Music Journal, 35*(1), 1-15.

Balkwill, L.-L., & Thompson, W. F. (1999). A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues. *Music Perception, 17*, 43-64.

Bartlett, M. S. (1966). *An introduction to stochastic processes* (2nd ed.). Cambridge: Cambridge University Press.

Boersma, P., & Weenink, D. (1992-2007). Praat: Doing phonetics by computer. Amsterdam, The Netherlands: University of Amsterdam.

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

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

Brattico, E., & Sassanelli, F. (2000). Perception and musical preferences in Wishart's work. *Journal of New Music Research, 29*, 107-119.

Brewster, A., Smith, H., & Dean, R. T. (Artist). (2004). soundAffects. *Text*, *8*(2).

Cambouropoulos, E. (2006). Musical parallelism and melodic segmentation: A computational approach. *Music Perception, 23*, 249-267.

Clarke, E. F. (2005). *Ways of listening: An ecological approach to the perception of musical meaning*. New York: Oxford University Press.

Coutinho, E., & Cangelosi, A. (2009). The use of spatio-temporal connectionist models in psychological studies of musical emotions. *Music Perception, 27*, 1-15.

Cupchik, G. C. (2001). Theoretical integration essay: Aesthetics and emotion in entertainment media. *Media Psychology, 3*, 69-89.

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

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

Dean, R. T., & Bailes, F. (2010b). A rise-fall temporal asymmetry of intensity in composed and improvised electroacoustic music. *Organised Sound, 15*, 147-158.

Dean, R. T., Bailes, F., & Schubert, E. (2011). Acoustic intensity causes perceived changes in arousal levels in music. *PLoS One, 6,* e18591.

Deliège, I. (1989). A perceptual approach to contemporary musical forms. *Contemporary Music Review, 4*, 213-230.

Deliège, I., Mélen, M., Stammers, D., & Cross, I. (1996). Musical schemata in real-time listening to a piece of music. *Music Perception, 14*, 117-160.

Dibben, N. (2001). What do we hear, when we hear music? Music perception and musical material. *Musicae Scientiae, 5*, 161-194.

Dibben, N. (2004). The role of peripheral feedback in emotional experience with music. *Music Perception, 22*, 79-115.

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, 74*, 427-431.

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, 57*, 1526-1536.

Emmerson, S. (1986). The relation of language to materials. In S. Emmerson (Ed.), *The Language of Electroacoustic Music* (pp. 17-39). New York: Harwood Academic Publishers.

Enders, W. (2004). *Applied econometric time series* (2nd ed.). Hoboken, NJ: Wiley.

Erlich, N., Lipp, O., & Slaughter, V. (2007, April). *Physiological correlates of affective sounds*. Paper presented at the 34th Australasian Experimental Psychology Conference, The Australian National University, Canberra.

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* (pp. 223-248). London, UK: Oxford University Press.

Geringer, J. M. (1995). Continuous loudness judgements of dynamics in recorded music excerpts. *Journal of Research in Music Education, 43*, 22-35.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica, 37*, 424-438.

Grewe, O., Nagel, F., Kopiez, R., & Altenmüller, E. (2007). Emotions over time: Synchronicity and development of subjective, physiological, and facial affective reactions to music. *Emotion, 7*, 774-788.

Grimault, N., Bacon, S. P., & Micheyl, C. (2002). Auditory stream segregation on the basis of amplitude-modulation rate. *Journal of the Acoustical Society of America, 111*, 1340-1348.

Hamilton, J. D. (1994). *Time series analysis*. Princeton, NJ: Princeton University Press.

Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation*. Cambridge, MA: MIT Press.

Imberty, M. (1993). How do we perceive atonal music? Suggestions for a theoretical approach. *Contemporary Music Review, 9*, 325-337.

Juslin, P., & Laukka, P. (2004). Expression, perception and induction of musical emotions: A review and a questionnaire study of everyday listening. *Journal of New Music Research, 33*, 217-238.

Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences, 31*, 559-621.

Kallinen, K., & Ravaja, N. (2006). Emotion peceived and emotion felt: Same and different. *Musicae Scientiae, 10*, 191-213.

Keller, D. (2000). Compositional processes from an ecological perspective. *Leonardo Music Journal, 10*, 55-60.

Korhonen, M. D., Clausi, D. A., & Jernigan, M. E. (2006). Modeling emotional content of music using system identification. *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics, 36*, 588-599.

Krumhansl, C. L. (1996). A perceptual analysis of Mozart's Piano Sonata K. 282: Segmentation, tension, and musical ideas. *Music Perception, 13*, 401-432.

Lalitte, P., & Bigand, E. (2006). Music in the moment? Revisiting the effect of large scale structures. *Perceptual and Motor Skills, 103*, 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, 22*, 265-296.

Landy, L. (2009). Sound-based music 4 all. In R. T. Dean (Ed.), *The Oxford handbook of computer music* (pp. 518-535). New York: Oxford University Press.

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, 34*, 39-67.

Lütkepohl, H. (2007). *New introduction to multiple time series analysis*. Berlin and Heidelberg: Springer.

Mathews, M. V. (1979). Perception of crescendos as a function of duration. *Journal of the Acoustical Society of America, 65*, S123-S123.

McAdams, S., Vieillard, S., Houix, O., & Reynolds, R. (2004). Perception of musical similarity among contemporary thematic materials in two instrumentations. *Music Perception, 22*, 207-237.

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, 22*, 297-350.

MPEG-7 Overview § ISO/IEC JTC1/SC29/WG11N6828 (2004). http://mpeg.chiariglione.org/standards/mpeg-7/mpeg-7.htm

Neuhoff, J. G. (2001). An adaptive bias in the perception of looming auditory motion. *Ecological Psychology, 132*, 87-110.

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

Olsen, K. N., Stevens, C., & Tardieu, J. (2010). Loudness change in response to dynamic acoustic intensity. *Journal of Experimental Psychology: Human Perception and Performance, 36*, 1631-1644.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*, 1161-1178.

Schubert, E. (1996, August). *Measuring temporal emotional response to music using the two dimensional emotion space*. Paper presented at the 4th International Conference for Music Perception and Cognition, Montreal, Canada.

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

Schubert, E. (2004). Modeling perceived emotion with continuous musical features. *Music Perception, 21*, 561-585.

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, 19*, 87-120.

Stecker, G. C., & Hafter, E. R. (2000). An effect of temporal asymmetry in loudness. *Journal of the Acoustical Society of America, 107*, 3358-3368.

Theil, H. (1966). *Applied economic forecasting.* Amsterdam: RandMcNally.

Tsabary, E. (2009). Which aural skills are necessary for composing, performing and understanding electroacoustic music, and to what extent are they teachable by traditional aural training? *Organised Sound, 14*, 299-309.

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., & Levitin, D. J. (2006). Cross-modal interactions in the perception of musical performance. *Cognition, 101*, 80-113.

Wald, A. (1955). *Selected papers in statistics and probability*. Stanford, CA: Stanford University Press.

Windsor, L. (2000). Through and around the acousmatic: The interpretation of electracoustic sounds. In S. Emmerson (Ed.), *Music, electronic media and culture* (pp. 7-35). Abingdon, Oxon: Ashgate Publishing Group.

Windsor, W. L. (1995). *A perceptual approach to the description and analysis of acousmatic music.* London, UK: City University.

Witvliet, C. V. O., & Vrana, S. R. (2007). Play it again Sam: Repeated exposure to emotionally evocative music polarises liking and smiling responses, and influences other affective reports, facial EMG, and heart rate. *Cognition and Emotion, 21*, 3-25.

color plates

Comparative Time Series Analysis of Perceptual Responses to Electroacoustic Music

Freya Bailes and Roger T. Dean

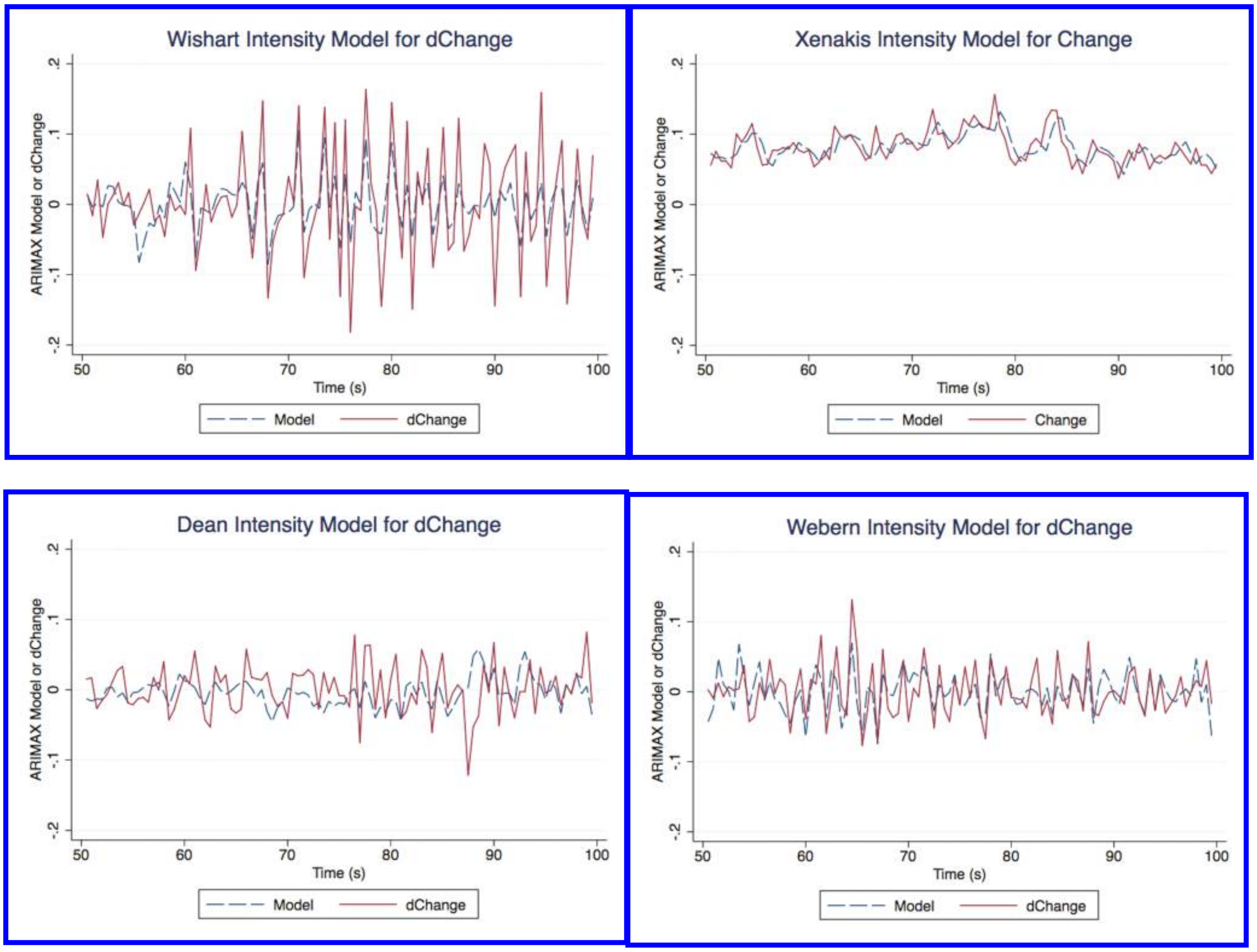

FIGURE 1.

Influence of intensity on perceived change in sound per piece from the best ARIMAX models, with intensity as the exogenous variable. Time is in seconds. Note: In all figures we show the first 100 s only for clarity. All models, however, are representative of the fit achieved throughout the time series.

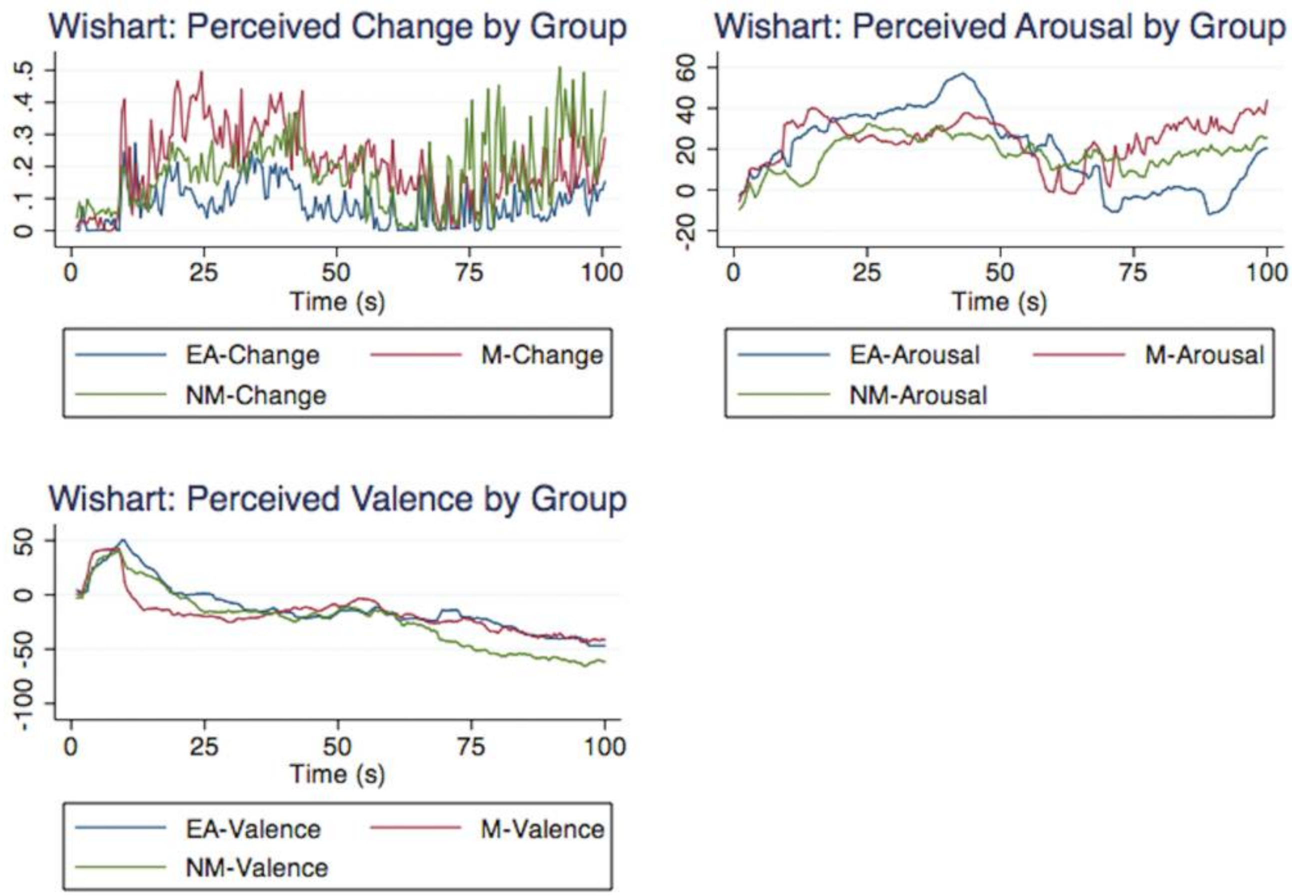

FIGURE 3.

Perceived a) change in sound, b) arousal, and c) valence for Wishart, by group of musical expertise. Time is in seconds. EA = expert in contemporary music; M = other musicians; NM = nonmusicians.

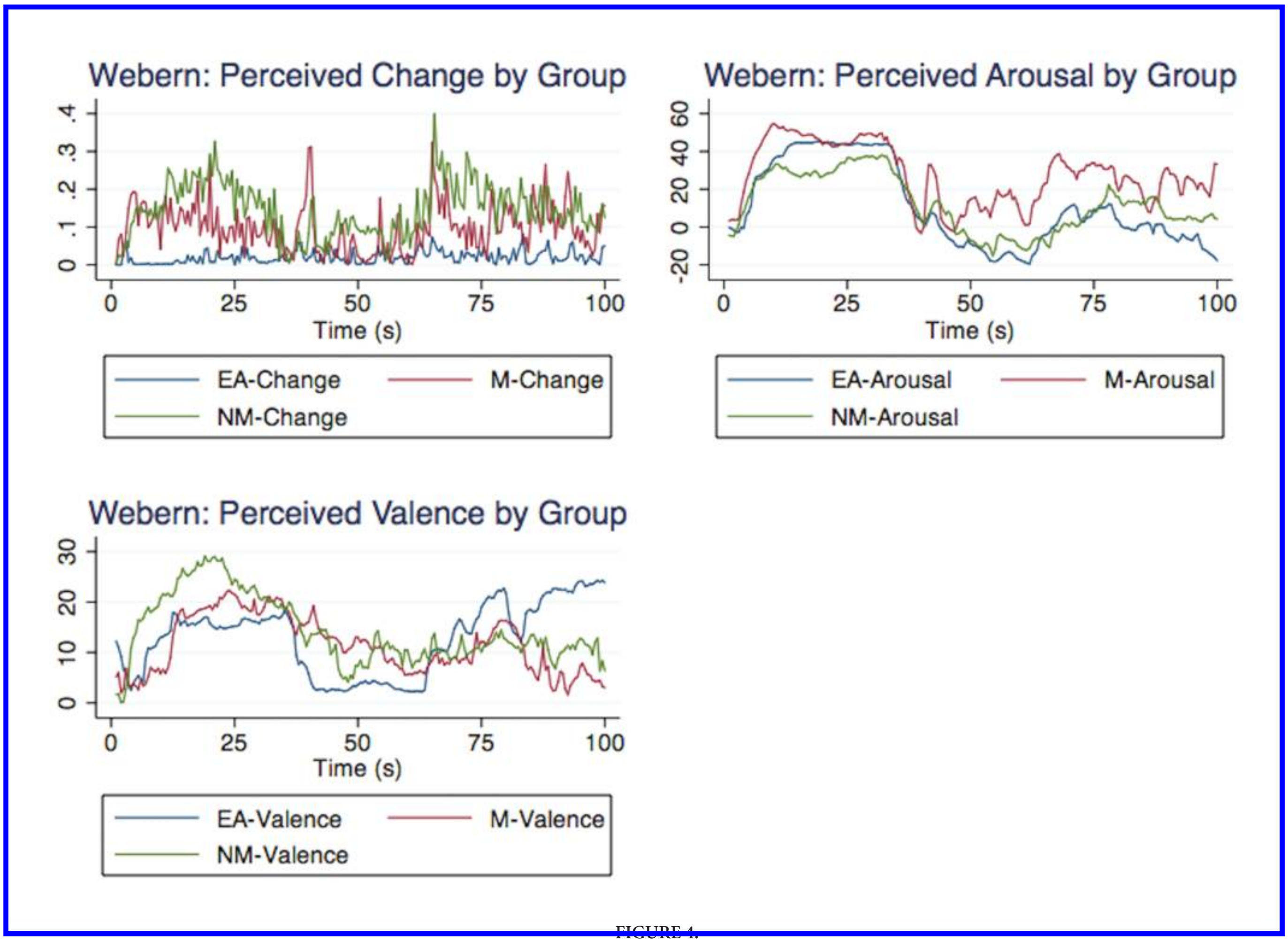

FIGURE 4.

Perceived a) change in sound, b) arousal and c) valence for Webern, by group of musical expertise. Time is in seconds. EA = expert in contemporary music; M = other musicians; NM = nonmusicians.