

MODELING PERCEPTIONS OF VALENCE IN DIVERSE MUSIC: ROLES OF ACOUSTIC FEATURES, AGENCY, AND INDIVIDUAL VARIATION

ROGER T. DEAN
*MARCS Institute, Western Sydney University, Penrith,
Australia*

FREYA BAILES
University of Leeds, Leeds, United Kingdom

WE INVESTIGATE THE ROLES OF THE ACOUSTIC parameters intensity and spectral flatness in the modeling of continuously measured perceptions of affect in nine diverse musical extracts. The extract sources range from Australian Aboriginal and Balinese music, to classical music from Mozart to minimalism and Xenakis; and include jazz, ambient, drum n' bass and performance text. We particularly assess whether modeling perceptions of the valence expressed by the music, generally modeled less well than the affective dimension of arousal, can be enhanced by inclusion of perceptions of change in the sound, human agency, musical segmentation, and random effects across participants, as model components. We confirm each of these expectations, and provide indications that perceived change in the music may eventually be subsumed adequately under its components such as acoustic features and agency. We find that participants vary substantially in the predictors useful for modeling their responses (judged by the random effects components of mixed effects cross-sectional time series analyses). But we also find that pieces do too, while yet sharing sufficient features that a single common model of the responses to all nine pieces has competitive precision.

Received: April 25, 2015, accepted January 8, 2016.

Key words: music perception, valence, modeling, acoustic features, agency

MOST PREVIOUS WORK ON PERCEPTIONS OF affect expressed by music has used at least the two dimensional circumplex model of affect developed by Russell, in which one axis is arousal and the other is valence, the latter of which concerns degrees of perceived positivity or pleasantness (Russell, 1980, 2003). Most work has concerned retrospective “summary”

perceptions of the affect expressed in short (often 30 seconds or shorter) sonic extracts and has been well reviewed (Juslin & Sloboda, 2001, 2011; Juslin & Västfjäll, 2008). There is less work on continuous perceptions of affect in music, though it has a long tradition (Coutinho & Cangelosi, 2009, 2011; Madsen & Fredrickson, 1993; McAdams, Vines, Vieillard, Smith, & Reynolds, 2004). The resulting time series of moment-by-moment perceptual data can then be modeled in a bid to understand the dynamic factors that shape perceived affect. In the case of analyses of continuous responses (Schubert, 1996, 1999, 2004), the resultant statistical models of the data have been moderately good in predicting the arousal response, but notably worse with the valence response (Bailes & Dean, 2012; Coutinho & Cangelosi, 2009, 2011; Dean & Bailes, 2010a). Even in the much larger set of studies that measure discrete retrospective perceptions (which are necessarily much simpler, and contain far fewer data elements to be modeled), the same disparity exists, though there are occasional dramatic exceptions, where the use of a large enough panoply of musical components as predictors or simplified musical stimuli has evinced success. For example, using short simple (pitch-centered) tonal items that were algorithmically composed to permit parameter control, 77-89% of variance in the components of the circumplex model, though labeled differently, could be explained by six musical cues: mode, tempo, dynamics, articulation, timbre, and register (Eerola, Friberg, & Bresin, 2013).

Our previous work led us to envisage that continuously perceived valence might be much more individualistic in origin than perceived arousal. We suggest this may partly reflect individuals' overall liking for a piece and for music at large, as suggested by data correlations in the case of both musicians and nonmusicians listening to the music we studied previously. This music comprised electroacoustic pieces by Wishart, Xenakis, and Dean, and piano music by Webern (Bailes & Dean, 2012). It follows that techniques that can model differences between individuals (so called “random effects”: see Results section for further description of the “mixed effects” aspect of our analyses) might provide better models of valence. As we showed recently in detail for perceived arousal (Dean, Bailes, & Dunsmuir, 2014a,

2014b) cross-sectional time series analysis (CSTSA) is a mixed effects technique. So CSTSA might bring models of arousal and valence into a relation approaching equality.

We also previously found indications that perceptions of agency, particularly human agency (Bailes & Dean, 2012; Dean & Bailes, 2010a), could provide predictors that are useful for valence modeling. By agency we here refer primarily to the broad idea that some acoustic events have physical origins that a listener can detect or at least imagine (for example, waves, wind, or breaking glass), and among these, some have human or animal origins (vocal sounds, clapping). Other sounds, such as some electroacoustic components, can be devoid of such agency, particularly when a listener is unable to deduce the means by which the sounds were produced. We also consider the secondary extension of this concept of agency, in which the entry of a singer within a performed sonic texture, or of a pianist (concerto soloist) within an orchestral texture, might be perceived as the addition of a new source of human agency. Naturally, the degree of activity of such an agent is likely to be perceptible and so a measure of this could provide a continuous predictor variable related to agency, and not just a present/absent dichotomous variable.

It is important to distinguish physical properties of an acoustic stream from what is perceived among those properties (Leman, 2008; Leman, Vermeulen, de Voogdt, Moelants, & Lesaffre, 2005). For example, variations in acoustic intensity are perceived as variations of loudness with quite high precision and little delay, whereas timbral changes or pitch fluxes may be more ambiguous and perceived with lower precision. Broadly, the more robust components in terms of their influence on listeners are those such as acoustic intensity, which feature in the cue redundancy hypothesis, a cross-cultural view of their impact (Balkwill & Thompson, 1999; Balkwill, Thompson, & Matsunaga, 2004). In essence this theory suggests that the robust features may operate regardless of the cultural origin of either the music or the listener.

Overall, listeners' perceptions of the acoustic features of music are the factors that might bear most on their perception of affect, rather than the acoustic features per se. To address this we have repeatedly assessed listeners' continuous perceptions of ongoing "change in the sound" and used this measure successfully as a predictor in our models. Since we give no instructions as to what constitutes change, because we wish to understand listeners' perceptions themselves, it has always been apparent that perceived "change" probably comprises many components that eventually can be isolated, so that such a parameter might no longer be necessary for models.

For example, intensity changes are a dominant driver of perceived change, and if they were the only driver they would render the perceived change variable unnecessary. Intensity changes also substantially drive perception of affect in most pieces, and we showed with an empirical intervention that changing the intensity profile of certain pieces could correspondingly change the perception of their expressed arousal (Dean, Bailes, & Schubert, 2011).

Our focus here is on listeners who are not trained musicians, and on choosing works that will be individually unfamiliar, including some genres to which participants may well have been exposed. Thus we are mainly dealing with novel listening experiences. Our previous work showed that the compositional structure and musicological segmentation (in time blocks generally of 15-30 seconds) of the wide range of pieces studied here was largely reflected (implicitly) in our nonmusicians' continuous perceptions of change (Dean & Bailes, 2014). This was true across the range of nine highly diverse pieces: the music ranged from Aboriginal and Balinese, to a group of Western pieces (classical, jazz, minimal music, drum n' bass, and ambient), and included a piece of performance text using a constructed language. There was some emphasis on Western music since 1950. The observed behavior is unlikely to be simply because intensity profiles align completely with the compositional segments, because these are actually constituted by a range of other features, such as harmonic shift and agency change. Scope remains for further dissection of the components responsible for perceived change, ultimately with a view to dispensing with this poorly defined variable. Here we continue to pursue this, in one case again through the consideration of agency.

We sought to gain a broader impression of the utility of acoustic parameters, perceived change, agency, and implicitly identified structural features for predictive models of arousal and valence. Our hypotheses were as follows:

- 1) That acoustic features such as acoustic intensity (and spectral flatness) are the main predictors of perceived affect, while perceived change is subordinate and may be dispensable when both its effectors and musical segments are fully defined. (H1a: Consequently, the fewer the perceptual segments, which are based on perceived change analyses, or musicological segments, the less the influence of perceived change will be in models.)
- 2) That segments defined on the basis of the agency of human-derived sounds influence both perceived arousal and valence. They do so in conjunction with acoustic features.

- 3) That perceived structural blocks in the music influence perceived affect, again in conjunction with acoustic features. (H3a: Given H1-3, it may be expected that pieces that are perceptually homogeneous — that is, lack structural segmentation — will also have more homogeneous acoustic features and hence their perceived affect will be poorly modeled.)
- 4) That there are considerable interindividual differences in responses to the pieces (this can be reflected in individually different autoregressive parameters and/or parameters of responses to specified predictors).
- 5) That liking and overall valence will correlate, and, given hypothesis 4, that taking account of individual differences in valence response mechanisms may allow its enhanced modeling.
- 6) That in spite of the dramatic musical (and cultural) differences between the pieces studied, the perceived affect they induce will all be reasonably predicted by a shared model whose performance is at least comparable with those of models of individual pieces.

Later in the paper, our results permit us to develop an additional hypothesis, which, unlike the above, was not predefined. We refer to these hypotheses below as H1-H7.

We first use conventional univariate autoregressive time series techniques to model the grand average of responses for each piece, as is also routinely done in neuroscience and in previous continuous response work. Then we follow this up using the newer technique of cross-sectional time series analysis (CSTSA), which retains the integrity of each individual participant's responses in the modeling. We have introduced the technique for comparison of models of individual participants' continuous affect responses, and provided a technical elaboration in a recent pair of papers (Dean et al., 2014a, 2014b). CSTSA is also known as "panel" or "longitudinal" analysis: these latter terms are most often used when the data comprise relatively few time points, whereas the present data series each contain around 120 time points. Besides preserving the integrity of every data series, CSTSA also operates as a mode of mixed effects multilevel analysis, allowing us to distinguish fixed effects (representing shared properties of the participants), from random effects, statistical descriptions of variabilities in predictors and/or in autoregression features between participants. Random effects are commonly studied across samples of participants. Thus if the data show a fixed effect dependence on the magnitude of a stimulus for their response to it (for example,

the magnitude of acoustic intensity influencing perceptions of loudness), then there may be two kinds of random effects across individuals. Firstly, they may vary in the level of loudness they perceive to attach to the lowest level of acoustic intensity: this would constitute a random effect across participants on the intercept for loudness response to intensity. Secondly, they may vary in the degree to which increases in intensity induce increases in their perception of loudness: this would correspond to a random effect across participants on the coefficient for the loudness response to intensity. In each case, for any individual, the predictor relevant to their individual response would be the sum of the estimated population fixed effect, and their individual random effect. Thus the analysis often achieves a considerable enhancement of precision and reliability over their conventional counterparts (e.g., in a circumstance where an ANOVA might be conventional). Additionally, the random effects are expressed as sample normal distributions, and thus characterized by two parameters (mean and *SD*), constituting a quite economical addition of degrees of freedom to the models.

In the present work, the random effects we present are all of the second kind, upon coefficients for predictors. This is to be expected, since all but one analyses concern differenced time series, and hence the intercept "bias," the random effect across individuals, has been largely removed. The populations represented by our data are both that of participants and sometimes also that of pieces of music (represented by our nine diverse items). CSTSA provides best linear unbiased predictor (BLUP) estimates of the realization of its model for each individual unit analyzed (in our case, participants or pieces). Thus the random effects reveal the varied nature of action of the factors studied. We have simplified the presentation of the details of the resultant statistical models (such as all the individual coefficients or the standard deviation of random effect terms), so as to make the key interpretative aspects as visible as possible.

Although we chose some commercial music (drum n' bass, ambient) and some potentially familiar classical music (Mozart), our musical sample was intended to include a majority of relatively unfamiliar works (e.g., Aboriginal music, Balinese music, Xenakis, performance text), as we indeed found. The reason for this is that the behaviors in which we are interested include those of coming to grips with new and potentially unfamiliar pieces and styles during music listening. In accord with our hypotheses H4/5 we particularly envisaged that taking account of interindividual variations in response mechanisms would allow valence to be modeled roughly as well as arousal.

TABLE 1. *Perceptually and Musicologically Defined Segments in the Stimulus Pieces*

Piece	Style	Perceptual Segments	Musicological Segments
Eno, <i>Unfamiliar Winds</i>	Ambient	0	0
Glass, <i>Gradus</i>	Minimal music for solo saxophone	1	0
Smith/Dean, <i>Runda</i>	Performance text	1	0
Xenakis, <i>Metastaseis</i>	20 th century, orchestra	1	1
Art of Noise, <i>Camilla The Old Old Story</i>	Drum n' bass	2	0
Miles Davis, <i>Tutu</i>	Jazz	2	2
Munduk	Balinese gamelan	3	2
Mozart, <i>Piano Concerto 21</i>	Classical, orchestra	3	3
Bushfire, <i>Buffalo</i>	Australian Aboriginal music	4	3

Method

PARTICIPANTS

Twenty-one female undergraduate psychology students between the ages of 17 and 34 years ($M = 21.1$, $SD = 5.1$) participated for course credit. One participant achieved a score greater than 500 on the Ollen Musical Sophistication Index (Ollen, 2006) questionnaire (665), while all others had scores less than 500 out of a possible 1000 ($M = 113.6$, $SD = 73.9$), qualifying them as not musically sophisticated, as we intended. We included data from the musically sophisticated participant in the analyses. Written informed consent was gained in accordance with the University of Western Sydney Human Ethics Research Committee's stipulations.

STIMULI

The pieces are described in full in our preceding paper on this set of works (Dean & Bailes, 2014). In brief, the pieces were chosen to represent a wide range of periods (from traditional¹ Aboriginal music to the present day), of cultures (Aboriginal, Balinese, African-American, Popular, Art Music) and genres (such as the piano concerto, jazz, minimal music, ambient music, and performance text). We restricted the number of pieces under study to nine in order to balance stylistic breadth with the tradition within CSTSA to model such a number of items. An implication of hypotheses H2 and H3 is that the fewer perceived or musicologically defined segments a piece has, the more dominant acoustic intensity and/or spectral flatness will be as predictors (in comparison with perceived change or the segmentation itself). Given that we have some pieces that are somewhat homogeneous (such as the ambient and minimal music works) we have organized the results presentation with the pieces ordered first by number of "perceptual segments" based on a changepoint analysis of listeners' continuous

ratings of perceived change, and second by "musicological segments" based on a musicological analysis of the number of segments, from small numbers to larger (as described in the preceding paper, Dean & Bailes, 2014). This results in the sequence (number of perceptual segments, number of musicological segments) as follows:

ACOUSTIC ANALYSES

As previously, we assessed two continuous acoustic properties of the music: intensity (sound pressure level, dB, on a log scale) and spectral flatness (Wiener's entropy), a global acoustic parameter of the perceptual attribute timbre, expressed on a log scale from 0 to minus infinity. Methods for analyzing these have been detailed (Bailes & Dean 2009).

PROCEDURES

Our methods for determining continuous perceptions of musical change, expressed arousal, and valence have been detailed previously (Bailes & Dean 2012). Briefly, at the outset, participants learned to use the computerized continuous response system, by means of a practice trial in each of the "structure" and "affect" tasks. In the structure task participants freely move a computer mouse if they perceive "change" in the sound, and rates of movement are averaged at 2Hz. No guidance was provided as to what constitutes change. The other task provides continuous perceived affect measures, expressed using the 2D-emotion space, around which participants move the mouse to represent their perceptions. Participants heard the pieces twice, once for each of the two counterbalanced tasks. The order of pieces within each task block was randomized.

Listeners were asked to rate their familiarity and liking for each piece after hearing it for the first time, in a counterbalanced order. Likert scales from 1-5 were used: for liking, ranging from *strongly dislike* to *strongly like* with a neutral midpoint; for familiarity, the markers were

¹ The origins of this traditional Aboriginal music are unknown.

labeled such that 1 = *I have never heard anything like this before*, 2 = *I have heard something like this but not this piece before*, 3 = *I have heard this piece*, 4 = *This piece is very familiar to me*, and 5 = *I often listen to this piece of music*

DATA MODELING

Autoregressive time series analysis with external predictors (ARX). We previously presented extensive tutorial and descriptive introductions to this univariate technique and to multivariate vector autoregression (VARX) describing their application in our field (Bailes & Dean, 2012; Dean & Bailes, 2010a). We have focused on modeling continuous perceptual responses to music. In the current study, we first undertook ARX with grand average (unweighted) perceptual series, together with acoustic and structural predictors. In essence, ARX is the primary technique applicable to modeling and analyzing time series that are autocorrelated, such as most continuous measures related to music. An ARX model represents the impact of autocorrelation (where earlier values of a parameter are predictive of subsequent ones over a certain range of time lags); together with the impact of any predictor variables in the form of a so-called “transfer function.” We discuss the overall models involving both components in this paper. One benefit of ARX (or of VARX as discussed in previous work) is that continuous perceptual variables (such as perceived change) may, if appropriate, also be used as predictors of other perceptual responses. One of our purposes here was to seek to eliminate the perceived change variable to the extent possible. We consider the possible influences of continuously perceived change in a piece on the other perceptual responses we measure, as elaborated previously. This allows investigating our hypotheses about the role of perceived structural segmentation in the perception of affect; structural segmentation in turn has been determined by analyses of perceived change series (Dean & Bailes, 2014). Model selection for the ARX models in this paper is primarily based on minimizing the Bayesian Information Criterion (BIC), which penalizes for the addition of increasing numbers of predictors, weighing this against the improved model fit (lower BIC indicates a better model). Data series were stationarized when necessary (by taking the first difference series). In essence stationarizing involves ensuring that variance and covariance are constant across the time series (it does not involve removing them). The first-differenced form of a series *name* is termed *dname* here. All series needed to be stationarized in this fashion, with the sole exception of the perceived change series used in the ARX model of valence for the piece

by Smith/Dean featured in Table 4. Acceptable ARX models were required to give white noise residuals that lacked autocorrelation.

Cross-sectional time series analysis (CSTSA) of individuals’ continuous responses of arousal and valence. We tested our hypotheses further using CSTSA: this mixed effects technique is outlined above and elaborated in Dean et al. (2014a, 2014b). We adopted a standard analytical procedure. Starting with the best model of the grand average data for a particular piece and response (from Tables 3 and 4 below) as the potential fixed effect component of CSTSA, we tested for the possible utility of random effects on intensity, spectral flatness, and perceived change, and/or their segments, and on the autoregressive components (that is, the lags of the response being modeled). When there were such random effects, the fixed effect component was reoptimized by addition or subtraction of terms. Model selection cannot be primarily based on BIC during CSTSA since it cannot deal with random effects parameters. Thus, selection was based on maximizing model log likelihood, and as far as possible minimizing residuals (to two decimal places of their standard deviation). Over-fitting simply to give the best fit in these terms was avoided by attempting to eliminate individual predictors whose coefficients were not significant, and individual random effects whose standard deviation estimates were less than 2.5 times the standard error (because otherwise the real standard deviation value might well be zero). Some predictors/random effects were required for the preferred model even though the coefficients/SD were not individually significant (these are asterisked in the results Tables). All models that included random effects were required to have a highly significant likelihood ratio test against the fixed effects only model ($p < .001$); that is, to produce a dramatic increase in model likelihood.

We used STATA’s `xtmixed` command as the main modeling vehicle, though some analyses were also undertaken or confirmed in R. In `xtmixed`, which is fundamentally a mixed effects multilevel analysis tool, the autoregressive components can be added as predictors by constructing the relevant lagged versions of the variable in question.

Not surprisingly, some best linear unbiased predictions (BLUPs) for individual responses made by the CSTSA models gave residuals that still contained autoregression and were not white noise. This is an index of the fact that individuals did vary substantially in their responses and in the factors that influenced them, as the models show.

TABLE 2. Mean Familiarity, Liking, and Valence Ratings of Each Piece, with Standard Deviations

Artist/Piece	Familiarity <i>M</i> (<i>SD</i>)	Liking <i>M</i> (<i>SD</i>)	Valence <i>M</i> (<i>SD</i>)
Eno	1.90 (0.89)	3.52 (0.93)	18.52 (6.09)
Glass	1.52 (0.83)	3.14 (0.96)	26.89 (3.49)
Smith/Dean	1.10 (0.30)	1.95 (1.07)	-17.67 (6.93)
Xenakis	1.90 (0.83)	2.05 (0.92)	-26.43 (13.75)
Art of Noise	2.19 (0.98)	3.29 (1.15)	22.81 (3.93)
Miles Davis	2.24 (0.94)	3.57 (0.75)	21.32 (3.93)
Munduk	1.76 (0.54)	2.76 (1.14)	7.78 (5.11)
Mozart	2.81 (0.81)	3.81 (0.75)	32.68 (5.31)
Bushfire	2.05 (0.59)	2.90 (1.00)	11.87 (7.73)

Note. Familiarity and liking ratings for the nine pieces are shown for $n=21$ participants, ranked by segmentation (see methods). Valence is measured continuously on a scale from -100 to +100.

Results

LIKING AND FAMILIARITY

Table 2 shows that our expectations that the pieces would be individually unfamiliar were fulfilled, while the works from more widely experienced genres, such as drum n' bass, ambient music, and jazz did attract higher liking ratings than most of the pieces. Only the Mozart piece

attracted a familiarity rating as high as halfway up the scale from 1-5, indicating, "I have heard this piece." There were distinctions between familiarity and liking (also on a 1-5 scale): for liking, five pieces were above the halfway point; for example, Eno and Glass pieces ranked higher in the liking response table than for familiarity, while the Xenakis piece ranked relatively lower. Nevertheless, liking and familiarity correlated ($r = .73$, $p < .03$). Given the differences in the scales used, this result should not be over interpreted. We tested for correlations between mean continuous response valence across all participants, and their mean liking (H5) finding, $r = .94$ ($p < .0005$) was consistent with our hypothesis. Familiarity and mean continuous valence were not significantly correlated, $r = .57$, $p = .11$, consistent with the observed distinctions and the weaker correlation between familiarity and liking.

PERCEPTUAL RESPONSES: MODELING GRAND AVERAGE PERCEIVED AROUSAL AND VALENCE

We first modeled the grand average responses for each piece using the acoustic predictors and continuous perception of change alone (Table 3). This approach primarily addresses our hypothesis H1: that, as observed

TABLE 3. Optimized Grand Average Autoregressive Time Series Analysis Models (ARX) using Continuous Acoustic Features and Perceived Change as Candidate Predictors.

Piece	Modeled response (DV)	Acoustic/perceptual predictors	Autoregressive components	BIC	% Sum of squares fit	Correlation predicted series: data series
Eno	darousal	L4.dspectralf	Ar(1)	568.3	6.79	.24
	dvalence	L5.dintens	Ar(1)	669.3	4.22	.20
Glass	darousal	L5.dspectf	Ar(1)	603.9	6.30	.25
	dvalence	L(3,5,10).dchange	No Ar	587.0	6.41	.25
Smith/Dean	darousal	L9.dspectralf	Ar(1)	461.1	13.18	.36
	dvalence	L8.dintens*, constant	Ar(1,3)	425.3	12.75	.32
Xenakis	darousal	L(2,3,4,8*).dintens, L2.dspectralf	Ar(1)	884.5	23.57	.48
	dvalence	L(4,5,6*).dintens, L1.dchange	Ar(1)	878.2	25.87	.51
Art of Noise	darousal	L6.dintens, L(2,3).dspectralf, l5.dchange	Ar(1)	719.0	10.21	.32
	dvalence	L8.dintens*	Ar(2)	780.7	8.74	.30
Miles Davis	darousal	L(6,8).dspectralf, L(4,7).dchange	Ar(1-3)	761.2	22.00	.47
	dvalence	L8.dintens, L8.dchange	Ar(2,3)	825.2	24.81	.50
Munduk	darousal	L(3-7).dchange	Ar(1,3)	926.4	25.54	.51
	dvalence	L(5*,6*).dintens	Ar(1)	822.8	2.70	.16
Mozart	darousal	L(3,5).dintens	Ar(1)	943.9	11.34	.34
	dvalence	L3.dintens, L6.dspectralf, L7.dchange*	No Ar	874.9	3.59	.19
Bushfire	darousal	L1.dintens, L(3,7*).dspectralf	Ar(1)	545.2	16.09	.40
	dvalence	L(7,9).dintens	Ar(1/2)	646.3	12.41	.35

Note. The response variables are all stationarized by first-differencing. The ARX models all have white noise residuals free of autocorrelation. DV: dependent variable. L(n) indicates lag n of the variable listed after the dot. Ar(n): the list of autoregressive terms. BIC: Bayesian Information Criterion, which can be meaningfully compared only between different models of the same response series. spectralf and intens are abbreviations for spectral flatness and intensity. *predictor required for the BIC-optimized model, but not individually significant at $p < .05$. "Constant" indicates that a constant was required only in this Smith/Dean model. The % sum of data series squares fit is a stringent criterion, the correlation less so: it is presented for comparability with previous work. The sum of squares fit refers only to the modeled part of the series (the first few events cannot be modeled, depending on the number of lags of variables in the model).

TABLE 4. Autoregressive Time Series Analysis Models (ARX) Improved by Incorporating Perceptual Segmentation

Piece	Modeled response (DV)	No. perceived structural changepoints/musicological segments	Model details (if model better)	BIC (if model improved)	BIC of original model	% Sum of Squares fit (if model improved)
Eno	darousal dvalence	0/0			568.3 669.3	
Glass	darousal dvalence	1/0			603.9 587.0	
Smith/Dean	darousal dvalence	1/0	L9.dspectralf1, L8.dchange1, ar(1)	457.8 420.8	461.1 425.3	26.13 13.65
Xenakis	darousal	1/1	L(2,3,8).dintens1, L(2,3,4).dintens2, L2.dspectralf2, ar(1)	868.7	884.5	33.17
Art of Noise	dvalence darousal dvalence	2/0			878.2 719.0 780.7	
Miles Davis	darousal	2/2	L6.dspectralf2, L6.dspectralf3, L10.dspectralf4, L4.dchange4, ar(1-3)	739.0	761.2	25.75
Munduk	dvalence darousal	3/2	L(11).dintens L(3/5,7).dchange3 L(7).dchange4, ar(1,3)	912.6	926.4	26.25
	dvalence		L6.dintens, L(1,3).dchange2, L1.dspectralf1, L(1,4).dspectralf3, L4.dspectralf4, ar(1)	812.6	822.8	18.46
Mozart	darousal dvalence	3/3			943.9 874.9	
Bushfire	darousal dvalence	4/3			545.2 646.3	

Note. DV: dependent variable. Ar(n): the list of autoregressive terms. BIC: Bayesian Information Criterion, which can be meaningfully compared only between different models of the same response series. dpredictor1-*n* describes the predictor separated into segments 1-*n* corresponding to the perceptual segments defined in the perceived change response to the piece (*n* segments is one more than the number of changepoints detected).

previously, changes in acoustic intensity and spectral flatness are recurrent external predictors of continuously perceived arousal and valence.

Arousal is only moderately well modeled in every case, given the stringent criteria presented, and the two pieces with no perceptual segments are very poorly modeled (consistent with H3a). Changes in intensity are the most common significant predictor, sometimes complemented or replaced by spectral flatness. Perceived change contributes to only three of the nine arousal models. Valence is, unexpectedly, as well modeled as arousal with just two exceptions (Munduk and Mozart, where the models are much worse). Again, intensity and spectral flatness contribute to valence models, and perceived change contributes in four cases. H1 is strongly confirmed by these results, and they are also in agreement with our suggestion (H1a) that the fewer the change segments, the less common the perceptual parameter of change would be as a predictor. Thus for the pieces with no perceptual segments (the first three in the table), only 1 of the 6 optimized models involved perceived change as a predictor; whereas for the remaining pieces (the final six pieces in the table,

with two models per piece), each involving at least some perceived segmentation, 7 of the 12 models involved change.

Tables 3 and 4 show the parameters that were included in successful models of perceived arousal and valence. Table 3 reveals models of valence that are only slightly worse in general than those for arousal, whereas in the pieces we studied elsewhere (i.e., three pieces of electroacoustic music; Wishart, Xenakis, and Dean; and Webern, an atonal piano work), several cases showed valence models that were much inferior in quality to those for arousal. In the present work the valence models were mostly still worse than those for arousal, but not always and not substantially so: in all but two cases % squares explained for arousal and valence were within 5% of each other. The two exceptions, Munduk and Mozart, have models that fail to correlate more than .20 with the data. They share only the feature of clear-cut human agency changes (singing entries, soloist entries). This suggests that further investigation of influences of structural segmentation and agency may improve our understanding of these pieces in particular, and we pursue this in what follows. Thus it is of interest

to consider whether these models, and indeed the models in general, are enhanced by the use of distinct segments of the piece as separate predictors (H2/3). As we will see in Table 4, the valence model of Munduk (but not Mozart) is improved by taking account of this segmentation.

Subsequently, we next assessed our Hypotheses 2/3 that perception of large-scale structural elements in the pieces — as defined by changepoint-segments (which determine whole segment differences) in the perception of change response series (Dean & Bailes, 2014; Dean, Bailes, & Drummond, 2014) — influence perception of affect. The changepoint analysis we developed undertakes a principled assessment of whether a time series of perceived change is best considered a single segment, or benefits from subdivision into several segments. It does this by comparing autoregressive models of the series as a whole, versus models of optimized sets of segments, within defined probabilistic limits, and hence allows determination of whether the whole series does or does not comprise a sequence of segments. Note that the name of the process (and the R package employed), ‘changepoint’ analysis, is slightly misleading, since the analysis does not focus on change at a point, but rather, distinction between successive segments. For all pieces in which perceptual segments had been detected, we assessed whether the acoustic predictors differed in their influences across these segments by determining whether their predictive impact could be enhanced comparing models with or without distinct segmented predictors. Table 4 indicates the previously detected perceptual and musicological segmentations, and shows what differences were detected by this new approach. Only models that provide improvement through the use of segments are specified. Comparing the BIC for a given response between Table 4 and Table 3 indicates the basis for this. The BICs in Table 4 are better (lower) than the corresponding value in Table 3, so the segmented variable(s) specified can effectively replace (and improve on) the corresponding individual unsegmented variable. Because the addition of variables itself increases the BIC through the penalty system, this means that the segmented variables are more powerful individually as predictors than their unsegmented competitors.

It can be seen in Table 4 that for three of the eight pieces in which perceived segments were detected, these were influential in models. For the three pieces where perceptual segments were detected but not musicological segments (Smith, Glass, Art of Noise) segmental influence was detected in one. Overall, the data are consistent with the idea that perceptual segments may influence perceptions of arousal and valence (H2/3), but

at this level of analysis, continuously perceived change is a more powerful predictor (probably simply because it contains so much more data, and includes the data that itself is indicative of the perceived segmentation), often rendering the perceived segmentation secondary. When segments were effective in the models, this was because the impacts of the acoustic parameters and/or the perceived change differed between the segments.

At this stage of the analyses, Eno and Glass, together with the Mozart valence response series, remain particularly poorly modeled. It should also be noted that perceived change is limited in Eno and Glass. The simplest interpretation is that perceptions of affect in these pieces are primarily influenced by other factors than changes in intensity and/or spectral flatness. However, an alternative possibility (Hypotheses 4/5) is that assessing interindividual variation in responses and response mechanisms using CSTSA can provide a better model still based on the acoustic predictors. This is part of the objective of the next section.

CSTSA OF INDIVIDUAL PIECES

In this section we assess first whether the grand average models for each piece as just defined, are confirmed, modified, or overturned when the analysis is done on all the individual participant data series taken together. More importantly, the approach also allows an assessment of interindividual variation in response, quantitatively and qualitatively. For CSTSA models, the significance and coefficient value of individual predictors, the model log likelihood, and the overall fit become the primary selectors towards model parsimony (see Method). The likelihood ratio test indicates whether the random effects improve the fixed effects alone model, and only models where $p < .001$ for this test are shown. When the model in Table 5 shows substantial improvement over those in Tables 3/4 — which is the case for the first three pieces that show no pre-defined musicological segments — this shows that interindividual variation was very large but well modeled. In Table 5, the only data series that remains very poorly modeled is the members of the Mozart perceived valence series set.

It may be helpful to summarize the interpretation of CSTSA models such as those in Table 5. In each case, autoregressive lags of the modeled response (dependent variables arousal and valence respectively) are required. The “fixed effect” predictors are those that apply to the group of listeners as a whole, as in most analyses. In contrast, the “random effects” represent differences between individual listeners. This forms a distribution of differences in relation to the specified effectors. Thus, in the first entry for Eno, the random

TABLE 5. Optimized Cross-Sectional Time Series Analysis (CSTSA) Models of all Individual Response Series Using Only Continuous Acoustic Features and Perceived Change as Candidate Predictors

Piece	Modeled response (DV)	Acoustic/perceptual Fixed Effect predictors	Autoregressive lags of DV	Random Effects	% Sum of squares fit	Correlation fitted values: modeled data
Eno	darousal		1,2	L5.dchange, L(1-3).darousal	20.21	.45
	dvalence		1,2	L3.dintens, L(1,2).dvalence	25.89	.44
Glass	darousal	L7*.dchange	1	L(1-4).darousal	16.00	.40
	dvalence	L5.dintens	3	L(4,5).dintens, L(1,4).dchange, L(1,2,4).dvalence	13.26	.37
Smith/Dean	darousal	L9.dspectralf, L(1,3).dchange	1	L(9).dspectf, L(1-2).dintens, L(1/5).darous	21.40	.47
	dvalence		1	L2.dchange, L(1-3,5).dvalence	33.58	.58
Xenakis	darousal	L(6,8).dintens, L(0).dchange	1-3	L(2,3).dintens, L(3).dchange, L(1-5).darousal	24.74	.50
	dvalence	L5.dintens	1,2	L(1-5).dintens, L(1,3,4).dspectf, L(0,2).dchange, L(1,2).dvalence	16.18	.40
Art of Noise	darousal	L6.dintens,	1,2,5	L(0,1,3,4).dchange, L(0-2).dchange3, L(1-5).darousal	20.37	.45
	dvalence	L8*.dintens	1,2	L(0).dchange, L2.dchange1, L2.dchange2, L(1,3-10).dvalence	18.91	.44
Miles Davis	darousal		1,3	L(6,8,10*).dspectralf, L(2).dchange, L(1/2).darousal	39.32	.63
	dvalence	L(8).dintens, L(7,8,10*).dchange	5,8	dchange, L(2-4,7,9).dvalence	27.13	.52
Munduk	darousal	L(6,7).dchange, L(4,5).dintens	1	L4.dchange2, L4.dchange3, L(1,3).darousal	13.09	.36
	dvalence	L2*.dintens	1	L(1-5).dvalence	13.42	.37
Mozart	darousal	L5.dintens,	1,6	L4.dintens, L4.dspectralf, L(1-10).darousal	19.91	.45
	dvalence	L8*.dchange	1,3	L(1,2,4).dvalence	6.55	.26
Bushfire	darousal	L3.dspectralf	1,3	L(1,4,10).dintens, L(1,2).darousal	15.44	.39
	dvalence	L(7).dintens	1	L(1,2,3).dvalence	10.30	.32

Note. DV: dependent variable. darousal refers to the first difference of arousal; dvalence and dchange similarly to the first differences in valence and change. dchange1 (or 2, 3) refer to successive segments of the dchange series as used in earlier models. The optimized CSTSA models provide best linear unbiased predictors (BLUPs) for each individual response series, which passed Bartlett's white noise residuals test, but in some cases autocorrelations remained. Asterisks indicate individual predictors which were required for the optimal model but whose coefficients were not individually significant. Only correlation coefficients significant at $p < .05$ are shown; but these need to be treated with caution, since autocorrelation remains in some of the BLUPS, and hence the coefficient is not entirely reliable.

effects on L5.dchange mean that across the population of listeners there were substantial differences in the degree to which this parameter influenced the outcome perceived arousal, and that representing this in the model enhanced it significantly. The random effects on L(1-3). darousal similarly indicate that listeners differed substantially in the degree to which the specified autoregressive lags of the modeled darousal (the dependent variable) were influential.

The models shown in Table 5 broadly confirm those from grand average analyses: showing the importance of intensity and perceived change for models of arousal, and lesser roles for spectral flatness. For valence, perceived change was generally less important than for arousal, whereas other factors were similar. Note that Eno and Glass, both lacking perceptual segments and

previously poorly modeled, now allow quite good models. However, Mozart valence remains inadequately modeled. The results strongly support hypotheses H1, H4, and H5.

For every piece and response, random effects were significant, and it is recognizing and modeling this that permits all of the individual data series for a particular piece/response to be described together in single models of reasonable quality. The random effects are most commonly expressed on the autoregressive component, but also often on the acoustic variables belonging to the transfer function, and the perceived change continuous (or segmented) variable. Perceptual segments are important locales of random effects (i.e., variations in responses across individuals) for several of the pieces in which they were detected (but not for those lacking

musicological segments, whether or not there were any perceptual segments). Interestingly, when segments are involved in random effects it is through the continuous (segmented) perceived change response, showing that this is an important mediator of interindividual differences as well as fixed effects. These results support H4/5 (and are reasonably consistent with H3a).

Note that the random effect component for an individual is added on to the corresponding fixed effect component, which is for everyone, when there is such a fixed effect. The consequence is that predictors of perceived affect found either or both in the fixed and random effects parts of the model are impacting on the final model output, and hence putatively on the mechanism of response. Hence it is fair to say overall that acoustic and perceptual change parameters are confirmed as important and widespread contributors to models of continuously perceived affect. Indeed, in Table 5, every model requires either perceived change or an acoustic variable at the minimum as a predictor (H1). Thus, we later test the general applicability of these types of model (Hypothesis 6) across the pieces in a different way, using CSTSA to assess how the different pieces require the predictors to be modified in a communal model.

ANALYZING THE IMPACT OF AGENCY: DISSECTING PERCEIVED CHANGE FURTHER

We next investigate a specific aspect of H2, concerning human agency. Continuous perception of change in the music has been found largely to reflect the musical structure of the piece, including the sectionalization of a piano concerto represented by the entries of the soloist. But as indicated in the introduction, perceived change most probably rolls together many different perceptions. For example, changes in perceived loudness and timbre, which we have studied here through their acoustic counterparts intensity and spectral flatness, can contribute to this, as we showed previously. Note again that intensity and spectral flatness are not simply and entirely congruent with perceived loudness and timbre. Thus, it is of interest to continue to apportion the perception of change, ultimately with a view to being able to dispose of it, representing it fully in more specific subcomponents. As we have noted already, one subcomponent that it seems to represent is agency, in the specific sense of the apparent presence or absence of a particular human activity (such as vocal sounds, or a concerto soloist), and possibly the distinct narrative agency of musical development (such as perhaps exposition, development, or tonal shift). Another subcomponent is rhythmic flexibility, which has been proposed

to be a predictor (Chapin, Jantzen, Kelso, Steinberg, & Large, 2010), but not previously analyzed in our studies. In Chapin et al. (2010), removing expressive timing and dynamic variations together resulted in diminished affective responses, so their relative effects were not discriminated. Thus in this section we assess whether the agency of the soloist in our Mozart piano concerto extract can be quantitated, so as to provide a predictor which enhances our modeling of perceived arousal and valence.

We aimed to go beyond simply providing a predictor variable that defined segments of agency (such 0 for the absence of the soloist, 1 for the presence). As for the structural predictors analyzed already, this would provide such a limited amount of information as to be unlikely to be informative beyond segmenting the models. Rather we chose to assess whether the degree of activity of the soloist might be predictive. The degree of effort of a human agent might well be represented by the number of events per unit time the agent initiates: this would here correspond to event density of the piano. The degree of expressive complexity and density of an agent might relate to the variability of the duration of such events, which may also create variations in event density (number of events initiated per unit time). Thus, we measured event density with a 2 Hz sampling rate to provide a predictor that could be assessed in our models, using event counting in the music analysis and display software Sonic Visualiser (with some assistance from its onset detection algorithms). After a brief orchestral introduction, the piano soloist performs four segments in alternation with the orchestra, including the final segment of our extract. Virtually all the piano event durations are in the range minim (half note) to semiquaver (sixteenth note), and there are also some trills, arbitrarily taken here as providing a note event density one unit higher than the observed semiquaver maximum density (which is 5 events per time slice). The tempo of the performance fluctuates very slightly and there are only two substantial cases of rubato (dramatic elongations of a particular note at the expense of others, with minimal overall tempo change in the phrase), which are not represented in our analysis.

So our data-driven Hypothesis 7, an extension of H1-3 above, is that event density will be a predictor of perceived arousal and valence, improving our models and subsuming some of the predictive impacts of perceived change. The Mozart piece is pertinent for this analysis, as it is in general poorly modeled, and shows weak influences of perceived change (and only on perceived valence). The analysis is conducted on the grand average response series, developing from Table 3 (since the

TABLE 6. *The Influence of the Event Density Time Series in Optimized ARX Models of the Mozart Piano Concerto Extract.*

Modeled response (DV)	Model	Acoustic/perceptual predictors	Autoregressive components	BIC	% Sum of squares fit	Correlation predicted series: data series
darousal	Original	L(3,5).dintens	Ar(1)	943.9	11.34	.34
	Using soloist event density	L(5).dintens, L(4*, 14).devents	Ar(1)	873.0	9.66	.31
dvalence	Original	L3.dintens, L6.dspectralf, L7.dchange*	No Ar	874.9	3.59	.19
	Using soloist event density	L(9*, 10*, 12*).devents	No Ar	849.0	1.15	.11

Note. DV: dependent variable. Ar(n): the list of autoregressive terms. BIC: Bayesian Information Criterion, which can be meaningfully compared only between different models of the same response series. An interpretive limitation here is that the models of arousal and valence including devents as predictor are respectively 9 and 6 events shorter than those without it. devents refers to the first difference of the measured series of piano event density (attacks/0.5 s).

TABLE 7. *Commonalities and Differences Among the Pieces in Relation to Models of Affect: CSTSA of Grand Average Responses to All Pieces Taken Together, with Random Effects by Piece.*

Modeled response (DV)	Model Type	Acoustic/ perceptual Fixed Effect predictors	Autoregressive lags of DV	Random Effects	% Sum of squares fit	Correlation fitted values: modeled data
darousal	Fixed effects only	L(2*,5,9*).dintens, L(0-7).dchange	1,5		11.57	.34
	Random effects allowed	L(5,9*).dintens, L(0-7).dchange	1,5	L(2,3).darousal	15.15	.39
dvalence	Fixed effects only	L(8).dintens, L(3,4).dspectralf	1,2		5.60	.16
	Random effects allowed	L(3,4).dspectralf	1,2	L(5,8).dintens, L(1-3).dvalence	11.41	.34

Note. DV: dependent variable. darousal refers to the first difference of arousal: dvalence and dchange similarly to the first differences in valence and change. The better darousal model had white noise residuals in BLUPs for all pieces. The valence models did not, and thus were incompletely specified.

perceptually defined segments were not distinct in terms of their time series model components for the Mozart), and concerns event density of the solo piano part. Overall, H7 was upheld: model quality (BIC) was improved, though precision of fit was not. Furthermore, perceived change was no longer a required predictor in the optimal model for valence, thus being absent for both models of arousal and valence in this case (in support of H1). This suggests that indeed we had succeeded in separating the components of perceived change. For valence, even intensity was no longer required in the optimal model, but its degree of fit remained poor, and this observation is not powerful. The result suggests that indeed event density may well be influential particularly on the perception of arousal, and that this hypothesis will bear further investigation by means of experimental perturbation studies manipulating event density, such as we have previously conducted with acoustic intensity (Dean et al., 2011). In the case of event density, such experiments will require specially generated stimuli, or substantial segment by segment speeding and slowing of audio stimuli.

TESTING FOR COMMONALITIES AND DISTINCTIONS BETWEEN CORE PREDICTORS ACROSS GRAND AVERAGE RESPONSES FOR ALL PIECES
In Table 7 the grand average response series for each of the nine pieces are taken together, in a distinct CSTSA modeling approach that permits random effects to operate on the basis of the piece as unit, rather than the individual participant as unit (as in the analyses summarized in Table 5). This asks: do the pieces share predictors (mainly expressed in the fixed effects), and hence potentially mechanisms, and do they reflect different emphases on some of those mechanisms (in the random effects)? In order to make these interpretations as clear as possible we show separately cross-sectional models that only permit fixed effects, and those that additionally allow random effects, both for the arousal and the valence continuous perception responses.

The results support the implications of earlier models, that intensity and spectral flatness are respectively common and less common predictors for both perceived arousal and valence, while perceived change is subordinate, partly in that it is here a frequent predictor for arousal only (H1). Thus, they support the idea of

considerable commonality in mechanisms of perception of arousal and valence across the pieces. It must be remembered that the number of pieces under study is limited to only nine. However, they also show substantial inter-piece differences through the random effects, expressed in autoregression for both arousal and valence, and in the roles of intensity for valence (H6). Valence modeling is now comparably good to that of arousal, providing random effects are included (but note again that here they are random effects on piece). For both responses the model/data correlations achieved are quite similar to the mean of the corresponding values in Table 4 (though far fewer data items are involved in Table 6, which concerns grand average data series, than in Table 4).

Discussion

The analysis above provides support, to varying degrees, for all seven proposed hypotheses. While recognizing that our findings are limited by the relatively small number of pieces studied, the work continues to support the dominant role of the surface acoustic properties intensity and spectral flatness in modeling listeners' perceptions. Two features of the work seem to us most important: 1) the variety of evidence we provide that continuously perceived valence can (now) be modeled almost as well as perceived arousal; and 2) that we have indicated a path towards further separating out the acoustic and musical components which drive listeners' perceptions of change, such that this perceptual variable might be made redundant.

We first take these two features in turn. The variety of pieces studied here, rather greater than in our previous work, immediately encourages confidence in the valence modeling, since it is often even initially comparable in quality with that of arousal, whereas this was not generally true in models of the four pieces we studied in earlier work (Wishart, Xenakis, Dean, and Webern), perhaps because these were generally more dissonant with the average experience of nonmusician listeners than those we studied here. Furthermore, we here subsequently enhanced models of perceived affect — notably valence — by confirming the substantial interpersonal differences in how it is perceived, consistent with our finding of links between liking and continuous valence (and consistent with Hypotheses 4/5). Using CSTSA, we were able to attribute these interpersonal differences appropriately by means of random effects components in our models, and consequently to enhance their quality and/or precision. CSTSA is now directly dealing with a huge amount of data, as in the

models of Table 4: nothing is omitted or averaged, unlike most previous papers on continuous response measures. So it is not surprising these models appear more complex, but very encouraging that they perform comparably well to simple models of grand average responses. Most important is that they demonstrate the reasonable modeling of both perceived arousal and valence.

Considering the second feature (roles of perceived change), it is also apparent that the need to include the perceived change time series as a predictor in the models of affect (Table 3 here) may be slighter than seemed from the earlier analyses of the four previous pieces. Furthermore, the analysis (Table 6) of the Mozart piece, the only piece to which we could readily attach a continuous variable potentially representing degrees of agency, suggested that with appropriate analyses, the change variable could indeed be apportioned to its component drivers, and perhaps be made redundant. In other words, in this case we have possibly succeeded in analyzing further the multiple components (beyond acoustic intensity change and spectral flatness changes) that drive perception of overall change, as we sought. This of course requires much further and more general assessment.

Given that perceived change provides our means of assessing perceived structural segmentation, the same arguments can be applied to suggest that the influences of segments per se should potentially be subsumed into those of the constituent features that lead to their perception. This is among the aims behind our previous use of specific timbral features as predictors: for example, in detailed studies of Xenakis *Metastaseis* (Dean & Bailes, 2013) further investigated here, and in our demonstration of roles for human and animistic sounds (H2 here) (Bailes & Dean, 2012; Dean & Bailes, 2010b). This was also found to have bearing in other pieces, where perceptions of vocal agency seemed to be predictive of affective responses (Bailes & Dean, 2012). In the present work these aspects are extended into an analysis of possible levels of agency of the identifiably human agent, the piano soloist, in perceptions of Mozart's Piano Concerto No. 21 (Table 5).

In this Mozart work, we have taken piano event density as potentially a simple surrogate of the degree of human agency (not, of course, as a precise predictor of it). Thus, we consider the differenced form of the event density series as a simplified representation of variation and variability in agency. The differenced series was found to be influential in models, and could replace other predictors. Caution needs to be exercised, in that this Mozart response is poorly modeled here, but nevertheless

the results are at least consistent with the possibility that the predictive role of the piano event density might represent the impact of human agency (H2). If so, the agency is at least partly expressed through rhythmic density (and variety). The best models described here remove only 19.9% and 6.5%, respectively, of the squared data value series for arousal and valence. Yet the piece scores more highly than any other in our set for both familiarity, and liking (above mid-point). In other work consistent with these ratings, we showed that the Mozart piece induced a relatively high mean engagement in listeners, judged by a continuous response measure (Olsen, Dean, & Stevens, 2014). The listeners were a different group of student listeners from those studied here, but of similar ages. Using this measure of continuous engagement as a predictor in models like those described, the ARX arousal model was slightly enhanced (to 22.5% of squares), but for valence was substantially improved (to 17% of squares), bringing the piece roughly into parity with others in the present paper. Thus, engagement levels — sometimes related to liking and familiarity — may form an important predictor in models of affect for more familiar works.

Other models here extend the concept of human agency to one of structural or narrative agency, in which music structural features provide a changing pattern of potentially perceptible agency, in turn potentially influencing continuous perceptions of affect (H3). Maus

(1997) has particularly emphasized such an interpretation of music structure as agency, consistent with this argument. If this line of thought is to be applied to the structural narratives of primarily timbral works such as Xenakis *Bohor* (electroacoustic) and even *Metastaseis* (orchestral), then it will be necessary to consider what constitutes a perceptual event (as one routinely assumes a piano note attack in Mozart to be) in this music. Given greater clarity about the nature of perceptual events in this music, it may well be possible to assess perceptions of degree of agency, or effort, that would correspond to the real or imagined physical process of their generation (as we and others have hypothesized previously, Dean & Bailes, 2010b). We plan to investigate this perceptually.

Author Note

We appreciated funding from the Australian Research Council and the ongoing statistical and time series analysis advice from, and collaboration with, Professor William Dunsmuir, Department of Mathematics and Statistics, University of New South Wales.

Correspondence concerning this article should be addressed to Roger T. Dean, The MARCS Institute, University of Western Sydney, Locked Bag 1797, Penrith, NSW 2751, Australia. E-mail: Roger.Dean@western-sydney.edu.au

References

- BAILES, F., & DEAN, R. T. (2009). Listeners discern affective variation in computer-generated musical sounds. *Perception*, 38, 1386-1404.
- BAILES, F., & DEAN, R. T. (2012). Comparative time series analysis of perceptual responses to electroacoustic music. *Music Perception*, 29, 359-375.
- 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, 43-84.
- BALKWILL, L.-L., THOMPSON, W. F., & MATSUNAGA, R. (2004). Recognition of emotion in Japanese, Western, and Hindustani music by Japanese listeners. *Japanese Psychological Research*, 46, 337-349.
- CHAPIN, H., JANTZEN, K., KELSO, J. A. S., STEINBERG, F., & LARGE, E. (2010). Dynamic emotional and neural responses to music depend on performance expression and listener experience. *PLoS One*, 5(12), e13812.
- COUTINHO, E., & CANGELOSI, A. (2009). The use of spatio-temporal connectionist models in psychological studies of musical emotions. *Music Perception*, 27, 1-15.
- COUTINHO, E., & CANGELOSI, A. (2011). Musical emotions: Predicting second-by-second subjective feelings of emotion from low-level psychoacoustic features and physiological measurements. *Emotion*, 11, 921-937.
- DEAN, R. T., & BAILES, F. (2010a). Time series analysis as a method to examine acoustical influences on real-time perception of 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, 148-159.
- DEAN, R. T., & BAILES, F. (2013). Event and process in the fabric and perception of electroacoustic music. In M. Solomos (Ed.), *Proceedings of the International Symposium Xenakis. La musique électroacoustique / Xenakis. The electroacoustic music* (Intervention11, pp.12). Paris: Centre de Documentation Musique Contemporaine. Retrieved from http://www.cdmc.asso.fr/sites/default/files/texte/pdf/rencontres/intervention11_xenakis_electroacoustique.pdf
- DEAN, R. T., & BAILES, F. (2014). Influences of structure and agency on the perception of musical change. *Psychomusicology: Music, Mind, and Brain*, 24, 103-108.

- DEAN, R. T., BAILES, F., & DRUMMOND, J. (2014). Generative structures in improvisation: Computational segmentation of keyboard performances. *Journal of New Music Research*, 43, 224-236.
- DEAN, R. T., BAILES, F., & DUNSMUIR, W. T. M. (2014a). Shared and distinct mechanisms of individual and expertise-group perception of expressed arousal in four works. *Journal of Mathematics and Music*, 8, 207-223.
- DEAN, R. T., BAILES, F., & DUNSMUIR, W. T. M. (2014b). Time series analysis of real-time music perception: approaches to the assessment of individual and expertise differences in perception of expressed affect. *Journal of Mathematics and Music*, 8, 183-205.
- DEAN, R. T., BAILES, F., & SCHUBERT, E. (2011). Acoustic intensity causes perceived changes in arousal levels in music: An experimental investigation. *PLoS One*, 6(4), e18591.
- EEROLA, T., FRIBERG, A., & BRESIN, R. (2013). Emotional expression in music: Contribution, linearity, and additivity of primary musical cues. *Frontiers in Psychology*, 4, 7-18.
- JUSLIN, P. N., & SLOBODA, J. A. (2001). *Music and emotion: Theory and research*. New York: Oxford University Press.
- JUSLIN, P. N., & SLOBODA, J. (2011). *Handbook of music and emotion: Theory, research, applications*. Oxford: Oxford University Press.
- JUSLIN, P. N., & VÄSTFJÄLL, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences*, 31(05), 559-575.
- LEMAN, M. (2008). *Embodied music cognition and mediation technology*. Cambridge, MA: MIT Press.
- LEMAN, M., VERMEULEN, V., DE VOOGDT, T., MOELANTS, D., & LESAFFRE, M. (2005). Prediction of musical affect using a combination of acoustic structural cues. *Journal of New Music Research*, 34, 39-67.
- MADSEN, C. K., & FREDRICKSON, W. E. (1993). The experience of musical tension: A replication of Nielsen's research using the continuous response digital interface. *Journal of Music Therapy*, 30, 46-63.
- MAUS, F. E. (1997). Narrative, drama, and emotion in instrumental music. *Journal of Aesthetics and Art Criticism*, 55, 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*, 22, 297-350.
- OLLEN, J. (2006). *A criterion-related validity test of selected indicators of musical sophistication using expert ratings* (Unpublished doctoral dissertation). Ohio State University, Columbus.
- OLSEN, K. N., DEAN, R. T., & STEVENS, C. J. (2014). A continuous measure of musical engagement contributes to prediction of perceived arousal and valence. *Psychomusicology: Music, Mind, and Brain*, 24, 147-156.
- RUSSELL, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39, 1161-1178.
- RUSSELL, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, 110, 145-172.
- SCHUBERT, E. (1996). *Continuous response to music using a two dimensional emotion space*. Paper presented at the 4th International Conference of Music Perception and Cognition, Montreal.
- SCHUBERT, E. (1999). Measuring emotion continuously: Validity and reliability of the two-dimensional emotion-space. *Australian Journal of Psychology*, 51, 154-165.
- SCHUBERT, E. (2004). Modelling perceived emotion with continuous musical features. *Music Perception*, 21, 561-585.