Psychomusicology: Music, Mind, and Brain
2015, Vol. 25, No. 2, 124–137

0275-3987/15/$12.00 http://dx.doi.org/10.1037/pmu0000087

Both Acoustic Intensity and Loudness Contribute to Time-Series Models of Perceived Affect in Response to Music

Kirk N. Olsen, Roger T. Dean, and Catherine J. Stevens
University of Western Sydney

Freya Bailes
University of Hull

In the fields of music and emotion, investigations of causal relationships between acoustic and perceptual parameters have shed light on real-time factors that underpin affective response to music. Two experiments reported here aimed to distinguish the role of acoustic intensity and its perceptual correlate of loudness in affective responses to a diverse set of musical stimuli (Western classical, electroacoustic, and synthesized single-timbre). This was achieved by first subtly distorting the inherently strong psychophysical relationship between loudness and intensity using synthesized reverberation, and then analyzing the consequences of this for perceived affect. Two groups of participants continuously rated loudness ($N = 31$) and perceived arousal/valence ($N = 33$) in response to 3 musical stimuli ranging between 2 and 3 min in duration. Each stimulus consisted of 3 continuous segments with 3 manipulations of the second segment: (a) original acoustic profile; (b) reverberation introduced to decrease loudness but with the intensity profile of the original version closely maintained; and (c) reverberation introduced but the intensity profile increased by a mean of 3 dB SPL. We hypothesized that intensity *and* loudness both make a significant contribution to time-series models of perceived affect. Four types of time-series models are presented: the first allows intensity but not loudness as predictor, the second allows loudness but not intensity, the third allows intensity and loudness, and for conditions of reverberation, the fourth allows for the impact of segment variation. In sum, time-series modeling shows that both intensity and loudness are predictors of perceived arousal and, to lesser extent, perceived valence. However, loudness is often more powerful and sometimes dominant to the point of excluding intensity.

Keywords: affect, loudness, music, reverberation, time-series analysis

Research in the fields of music and emotion has investigated relationships between acoustic properties of music and listeners' perception of affect in response to music. Specifically, the dimensions of arousal and valence that comprise a two-dimensional circumplex model of experienced affect (Russell, 1980, 2003) are commonly measured in experiments investigating the multiplicity of acoustic cues associated with perceived affect (Bailes & Dean, 2012; Dean & Bailes, 2010; Dean, Bailes, & Schubert, 2011; Gingras, Marin, & Fitch, 2014; Olsen & Stevens, 2013; Ritossa & Rickard, 2004; Schubert, 1999, 2004, 2013). With additional use of continuous perceptual measurements, a clearer picture of the real-time factors that underpin affective response to music as it unfolds through time is beginning to emerge.

Kirk N. Olsen and Roger T. Dean, The MARCS Institute, University of Western Sydney; Catherine J. Stevens, The MARCS Institute and School of Social Sciences and Psychology, University of Western Sydney; Freya Bailes, School of Drama, Music, and Screen, University of Hull.

This research was supported by an Australian Research Council Discovery Project grant (DP120102426) held by the second, third, and fourth authors. We thank two anonymous reviewers and the MARCS Institute Music Cognition and Action research group for helpful comments on a draft. Stimuli are available from Kirk N. Olsen, on request.

Correspondence concerning this article should be addressed to Kirk N. Olsen, The MARCS Institute, University of Western Sydney, Locked Bag 1797, Penrith, NSW 2751, Australia. E-mail: k.olsen@uws.edu.au

From causal experiments, time-series models of continuous responses to music from the Western classical tradition have shown that manipulations of acoustic intensity profiles of performed music, as measured by decibels (dB) in SPL, significantly influence listeners' real-time perception of affect and loudness (Dean et al., 2011). A strong impact of acoustic intensity on perceived loudness in music perception is to be expected, as a long tradition of psychophysical research reports an intimate (yet not straightforward) relationship between loudness and acoustic intensity (Canévet & Scharf, 1990; Fletcher & Munson, 1933; Florentine, Popper, & Fay, 2011; Olsen, 2014; Olsen, Stevens, & Tardieu, 2010; Scharf, 1978; Stevens, 1956). One would also expect, therefore, that where intensity predicts perceived loudness *and* perceived affect, loudness might either share with intensity a predictive capacity for perceived affect, or mediate the role of intensity when affective responses are modeled.

Somewhat contrary to this hypothesis, time-series modeling of continuously perceived affect in response to one piece from the Western classical tradition (an extract from Dvorak's *Slavonic Dance Opus 46, No. 1*) included better prediction of perceived affect from acoustic intensity than from continuous loudness (Dean et al., 2011), with loudness not required as a component of the optimized model. The relevance of the surprising exclusion of loudness in this particular model is assessed in the present study. Is this phenomenon specific to the particular Dvorak piece, or is it more general across music with varied complexity and familiarity?

As argued above, we expect that loudness will either share a predictive capacity with intensity or mediate the role of intensity when affective responses to more diverse musical stimuli are modeled. To investigate this hypothesis further, acoustic intensity and perceived loudness must be manipulated in a way that subtly distorts the inherently strong psychophysical relationship, so their potential influences on affect can be distinguished empirically. This is accomplished here with the manipulation of synthesized reverberation, an acoustic cue active for listeners' impressions of auditory space and perceived auditory distance, which in turn has been shown to affect perceived loudness of a sound source (Butler, Levy, & Neff, 1980; Lee, Cabrera, & Martens, 2012; Mershon, Desaulniers, Kiefer, Amerson, & Mills, 1981; Stecker & Hafter, 2000; Zahorik, Brungart, & Bronkhorst, 2005; but see, Zahorik & Wightman, 2001).

Thus, the aim of the present article is to further address the predictive power of continuous intensity and loudness change in time-series models of perceived affect. This is achieved by manipulating continuous loudness perception without concurrent changes in the intensity profile of the stimulus; intensity changes that would normally be deemed necessary to manipulate loudness. Here, one group of participants continuously rated the affective elements of perceived arousal and valence (Experiment 1), while another participant group continuously rated loudness (Experiment 2). As argued above, we predict that intensity *and* loudness can both make a significant contribution to time-series models of affect in response to additional exemplars of Western classical and electroacoustic music, as well as a more simple and controlled single-timbre stimulus with cyclic variations of acoustic intensity. This unchanging-timbre stimulus was chosen to minimize the interaction between timbre and perceived changes in loudness. As the stimuli studied here vary in their level of complexity and event density, it is expected that the reverberation manipulation will have varied effects on loudness that will lead to interstimulus differences in the predictive power of time-series models. Nevertheless, our primary hypothesis challenges the exploratory single-piece (Dvorak) findings of Dean et al. (2011), expecting that a continuous measure of loudness will predict the affective elements of perceived arousal and perceived valence in response to a range of acoustic stimuli. The influence of intensity and loudness are also assessed in conjunction with another global parameter of acoustic streams, timbre, which is represented as spectral flatness (measured as the ratio of the geometric mean to the arithmetic mean of the power spectrum) and found to have predictive impact in previous time-series models of affect (Bailes & Dean, 2012; Dean & Bailes, 2010, 2011). Spectral flatness is included as an additional acoustic parameter for consistency and comparability with these previously published time-series models.

Method

Participants

The group that completed the continuous arousal/valance task consisted of 33 adult participants (26 females and 7 males; M = 21.48 years, SD = 5.39, range = 18–41 years). The group that completed the continuous loudness task consisted of 31 adult participants (24 females and 7 males; M = 19.90 years, SD = 2.07, range = 18–24 years). Both groups of participants were recruited from the University of Western Sydney. All reported normal hearing from a single binary-response question that asked whether each participant experienced a diagnosed hearing impairment. Participants in the affect and loudness groups had a median Ollen Musical Sophistication Index (OMSI; Ollen, 2006) of 103.50 (range = 17–442) and 103.00 (range = 17–336), respectively. An OMSI score <500 means a participant is "not musically sophisticated."

Stimuli

Two prerecorded musical pieces and one synthesized single-timbre stimulus were used as the three stimuli in the study. The first prerecorded performed stimulus was from Mozart's *Piano Concerto 21, K467* (2'18"), performed by Daniel Barenboim and the English Chamber Orchestra, recorded 1969, and digitally remastered in 1986 from HMV 5 86740 2. This stimulus represents the opening of the 3rd Movement, Allegro Vivace. The second prerecorded performed stimulus was composed by Trevor Wishart in 1977, entitled "Red Bird, a political prisoner's dream" (2'18"): this was taken from a recording on UbuWeb of this 45-min piece for tape and includes a strong narrative of obvious human and animate sounds. The final synthesized single-timbre stimulus (2'20") was generated in MAX/MSP using the MSP "Additive Synthesis" object. The timbre of this stimulus closely resembles a synthesized organ. To add temporal variation to this stimulus, ramps with continuous cyclic increases and decreases of intensity (23 dB SPL range between minimum and maximum levels within each ramp of intensity) were generated during the synthesis process with randomly selected durations of up- and down-ramps of intensity varying between 500 and 7,500 ms. These cyclic intensity variations were necessary for the reverberation manipulation (described below) to take effect for an otherwise unchanging continuous synthesized tone.

Three versions of the Mozart, Wishart, and Organ stimuli were presented in the study, with each version divided into three continuous segments. The first version of each stimulus retained the original acoustic characteristics. The original acoustic segments of each stimulus are signified by the symbol A. Therefore, AAA represents a stimulus with three segments, all of which are in their original acoustic format. In a second version of each stimulus, reverberation was added to the middle segment but the mean intensity profile was retained such that there was <1 dB SPL variation between its mean intensity and that of the original. From pilot data, this manipulation of reverberation and intensity using the chosen parameters detailed below caused mean loudness to decrease over the second segment at a magnitude similar to that caused by a 3 dB SPL decrease of intensity. The addition of reverberation to the second segment of each stimulus is signified by the symbol B. Therefore, ABA represents a stimulus with three segments; the first and last retain the original acoustic profile, and the second segment introduces reverberation with the intensity profile of the original version closely maintained. Finally, a third version of each stimulus was designed to retain the reverberation included in the B segment of each ABA version, but return the loudness profile as close as possible to its original form in response to the AAA versions. This was achieved by again implementing reverberation to the second segment of each stimulus, in addition to a 3 dB SPL increase of intensity. This manipulation is signified by the symbol B′. Therefore, AB′A represents a stimulus with three segments; the first and last retain the original acoustic profile, and the second segment introduces reverberation with an overall 3 dB SPL intensity increase relative to the second segment of the AAA version.

Each manipulation was created in MAX/MSP using the freeverb~ external object (version 1.0.1) within a freeverb~ MAX patch. The freeverb~ external object is a simple stereo implementation of the Schroeder/Moorer reverberation model (Moorer, 1979; Schroeder, 1962). It uses eight comb filters per channel in parallel and four all-pass filters per channel in series. The all-pass filters "smooth" the sound and the filters on the right channel are slightly detuned compared with the left channel in order to create a stereo effect. In addition to the freeverb~ external object, the freeverb~ MAX patch contains five continuous parameters relevant to our stimulus manipulations: (a) *room size* (larger values between 0 and 1 result in a longer reverberation tail, and values closer to 1 result in feedback or "room resonance"); (b) *damping* (a value of 0 results in nearly no damping and thus a high level of reflection and longer reverberation, whereas a value of 1 results in high damping and short reverberation); (c) *width* (the stereo width of the reverberation, where a value of 1 nearly gives two separate mono reverberations); (d) *wet* (the level of the reverberation effect, as values between 0 and 1); and (e) *dry* (the level of the unprocessed/original signal, as values between 0 and 1). There is also an overall *output volume*[1] ranging from 0–1.

In each segment of each stimulus, *room size*, *damping*, and *width* were set to 0.5; *dry* was set to 0.3. Therefore, only the parameters *wet* and *output volume* were varied between segment manipulations. In the segments where the original acoustic format was maintained (A segments), *wet* was set to 0.0, and *output volume* was set to 1.0. For the segments where reverberation was added while mean intensity of the original version was maintained (B segments), *wet* was set to 1.0; *output volume* was set to 0.33 for the Wishart and Mozart stimuli and 0.475 for the Organ stimulus. For the segments where the reverberation manipulation was maintained in addition to a 3 dB SPL increase of intensity (B′ segments), *wet* was set to 1.0; *output volume* was set to 0.47 for the Wishart and Mozart stimuli and 0.67 for the Organ stimulus. The specific output volume settings during the addition of the wet signal in B and B′ segments were chosen to maintain the original intensity profile in the B segment, and to provide an overall 3 dB SPL increase relative to the original intensity profile in the B′ segment. The final parameters were chosen after systematic variation of output volume within MAX/MSP until the desired outcome was reached. An increase of 3 dB SPL in the B′ segments was found in a pilot experiment to be the most appropriate manipulation to return mean loudness values back to those measured in response to original versions without reverberation.

The changes to the *wet* and *output volume* parameters in the B and B′ segments in the ABA and AB′A versions of the Mozart and Wishart stimuli were continuously rolled in over 250 ms at 0'46" and rolled out over 250 ms at 1'32". For the Organ stimulus, changes to the *wet* and *output volume* parameters in the B and B′ segments were rolled in over 500 ms at 0'48" and rolled out over 500 ms at 1'34". Each stimulus in the experiment was presented as an .aiff stereo 16 bit audio file with a 44.1 kHz sampling rate. Acoustic intensity and spectral flatness (Wiener's entropy) were measured using Praat (version 5.2.23).

Two practice trials were also presented using stimuli not presented in the main experiment: excerpts of the first movement from Mozart's *Symphony No. 40* (1'18") and of Xenakis's *Orient-Occident* (1'29"). These stimuli did not contain any acoustic manipulations and were presented for participants to become accustomed to the procedure.

Materials and Equipment

Paper materials comprised: (a) the 10-question OMSI (Ollen, 2006) assessing an individual's level of musical sophistication, in addition to basic demographic information; (b) a familiarity question, asking participants to rate how familiar each stimulus was by using five options, from 0: *I have never heard anything like this before* to 4: *I often listen to this piece of music*; and (c) a likability question, asking participants to rate on a five-point scale how much they liked each stimulus, from 0: *Really dislike* to 4: *Really like*, with *Neither like nor dislike* at the midpoint of the scale.

The experiment was conducted in a sound-attenuated booth and stimuli were presented binaurally through Sennheiser HD25 headphones. An Apple MacBook Pro laptop computer (System 10.6.2) using a custom written Java application was responsible for displaying on-screen instructions, the vertical loudness scale for the loudness experiment, and the two-dimensional affect scale for the arousal/valence experiment. The loudness scale was anchored with the labels "Loud" at the top and "Soft" at the bottom, with "Moderate" at the midpoint of the scale. The loudness scale ranged from 0–100, with "soft" corresponding to zero and "loud" corresponding to 100. For the two-dimensional affect scale, the dimension of arousal was represented on the vertical axis with the labels "Very Passive" at the bottom and "Very Active" at the top of the scale. The dimension of valence was represented on the horizontal axis with the labels "Very Negative" on the far left and "Very Positive" on the far right of the scale. Ratings were continuously recorded at a sampling rate of 2 Hz.

Procedure

Participants first read an experiment information sheet, gave written informed consent, and received standardized instructions regarding the task. For the loudness experiment, participants were instructed to move the mouse along the scale to continuously judge the loudness of the music, and completed the two practice trials before the main experiment trials began. For the arousal/valence experiment, detailed on-screen instructions regarding continuous ratings of perceived arousal and valence were presented, followed by training exercises in which participants made separate ratings of arousal and valence in response to affective verbal stimuli. In these training exercises, participants were presented with a series of emotion labels such as "happy," "sad," "surprised," and "angry," and asked to click on the portion of each scale that best represented that emotion. For example, "happy" would fall toward the "Very Positive" and "Very Active" ends of each scale, whereas "sad" would be represented by a response toward the "Very Passive" and "Very Negative" ends of each scale. After the separate arousal and

[1] The term "volume" here refers to its association with intensity and loudness, rather than a space-filling property or apparent size associated with "tonal volume" (Terrace & Stevens, 1962).

valence training exercises, participants went on to complete the two practice trials and then the main experiment trials using the two-dimensional affect scale.

Each experiment was divided into three blocks, with three stimuli in each block. Each block contained one version of each stimulus, but the specific version (either AAA, ABA, or AB′A) was randomly chosen for each block. Across the whole experiment, each stimulus (Mozart, Wishart, Organ) did not repeat sequentially; for example, if a version of the Mozart stimulus was the third presentation in the first block, the first presentation in the second block could not be another version of the Mozart stimulus. Participants were given the opportunity to have a short break between each block. Each experiment took ~40 min to complete.

Statistical Approach

For the purpose of time-series analysis, stationarity of group-mean time-series in response to each stimulus was assessed using the Augmented Dickey-Fuller Generalized Least-Squares test (Dickey & Fuller, 1979), and in each case the time-series required "differencing" to achieve stationarity. This creates a new series corresponding to differences between successive values of the original. A series resulting from differencing *seriesname* (e.g., arousal) is here termed *dseriesname* (e.g., darousal). AutoRegressive modeling with eXternal predictors (ARX) was used for the analysis, as introduced in detail previously (Dean & Bailes, 2010). The Bayesian information criterion (BIC) was used as the basis of model selection and penalizes strongly for the addition of predictor variables to a model (lowest BIC values are best). The statistical program STATA (version 12, STATA Corporation, College Station, TX) was used for all analyses.

The time-series analyses were first undertaken (as summarized above) using group-mean time-series responses derived from aggregating all participants' time-series in relation to each stimulus version (AAA, ABA, AB′A) and response type (loudness, arousal, valence). Such an approach has been detailed in depth in previous papers (Bailes & Dean, 2012; Dean & Bailes, 2010, 2011; Dean et al., 2011). An additional alternative approach that is applied here is to use cross-sectional time-series analysis (CSTSA), where each individual response series is maintained as a separate item and a mixed effects autoregressive analysis can be performed. This approach for analysis of continuous perception of affect has recently been introduced with the mathematical formulation of STATA's version (Dean, Bailes, & Dunsmuir, 2014a, 2014b). We summarize here the key features of the approach.

As a variant of a mixed-effects approach, STATA's xtmixed procedure allows the detection of both fixed effects (which apply to the participant population as a whole) and random effects, in which differences in the behavior of different "units" in the system can be identified and characterized as distributions across the units. Most commonly, units are the participants and the variability between participants can be removed to strengthen the reliability of the fixed effects analysis. In the present case, we considered the different segments as units (treated as A, B, and C representing the first, second, and third segments), to dissect their differences from their generalities. We also used a random-effects approach to assess whether reverberation per se had an influence. Finally, perceived valence was modeled, again using a cross-sectional time-series approach.

Results

Diversity in Stimulus Familiarity and Liking

To assess the level of diversity in familiarity and liking of each stimulus, descriptive statistics for these two parameters across the entire sample of participants are presented in Table 1. Across both experiments, the Western classical Mozart stimulus was rated as most familiar and most liked, whereas the electroacoustic Wishart stimulus and the synthesized organ were rated relatively low on familiarity, and closer to neutral in liking.

Manipulation of Loudness by Reverberation

Group mean time-series loudness responses to AAA, ABA, and AB′A versions of each stimulus are plotted in Figure 1, with each segment demarcated by dashed vertical black lines. To first assess the overall impact of reverberation on loudness, group mean time-series loudness response to all nine conditions (3 stimuli × 3 versions) were standardized into z-scores. Then time-series z-scores were averaged for each segment in each of the nine conditions. Difference scores were then calculated to compare the difference between standardized mean loudness in response to each segment of the AAA baseline version with each corresponding segment of the ABA version. This primarily addressed the degree to which the manipulation of reverberation in the second segment of each stimulus' ABA version affected loudness when the intensity profile was held equivalent to the corresponding second segment of each stimulus' AAA version. These standardized difference scores are shown in the top panel of Figure 2. For each segment, a value greater than zero in Figure 2 can be interpreted as an overall greater mean loudness response to that particular segment of the stimulus' ABA version, relative to the corresponding segment of the stimulus' AAA version. A value less than zero can be interpreted as an overall lower mean loudness response to that particular segment of the stimulus' ABA version, relative to the corresponding segment of the stimulus' AAA version. A value of zero represents no difference in mean loudness between corresponding segments in ABA and AAA versions of each stimulus.

From inspection of the top panel in Figure 2, very little difference between each version's first segment is observed for Mozart, Wishart, and Organ stimuli. This is not surprising, as the first

Table 1
Total Sample Mean Familiarity and Liking Ratings for Each Stimulus

Stimulus	Familiarity	Liking
Mozart	2.65 (.86)	3.90 (.64)
Wishart	1.43 (.80)	2.10 (.93)
Organ	1.32 (.59)	1.78 (.73)

Note. *SD* in parentheses; Familiarity scale ranged from 0: *I have never heard anything like this before* to 4: *I often listen to this piece of music*; Liking scale ranged from 0: *Really dislike* to 4: *Really like*, with *Neither like nor dislike* at the midpoint of the scale.

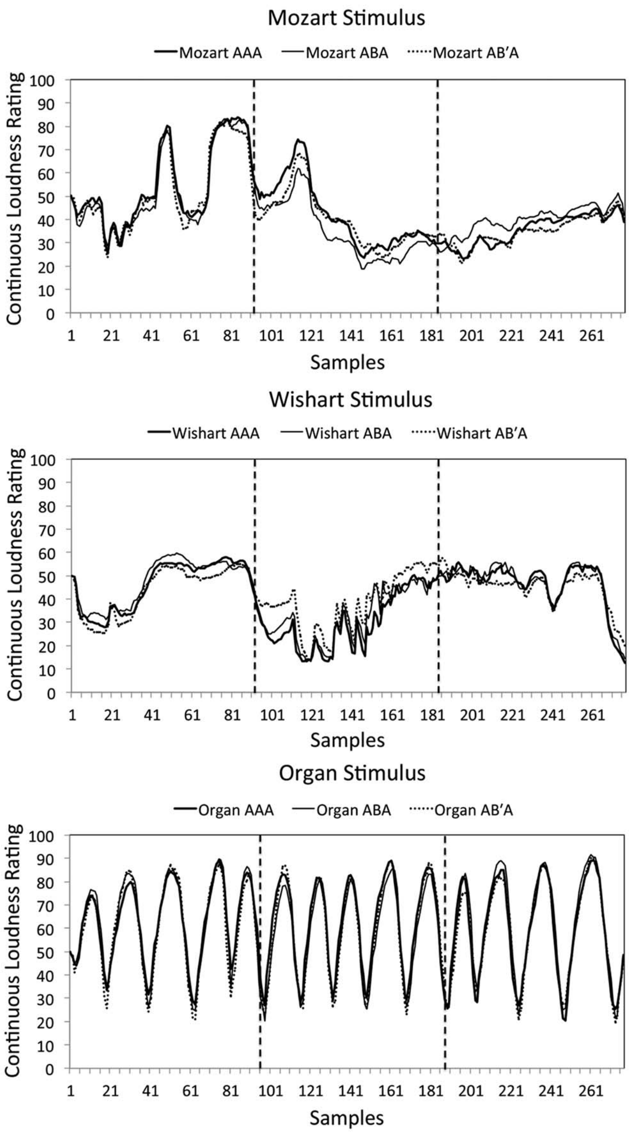

Figure 1. Group mean loudness time-series responses to the Mozart, Wishart, and "Organ" stimuli. Thick solid lines indicate the AAA versions; thin solid lines indicate the ABA versions; dotted lines indicate the AB′A versions. Vertical dashed lines demarcate the three segments within each stimulus. A = original acoustic segments of each stimulus; B = reverberation added to second segment with the intensity profile of the original version closely maintained; B′ = reverberation added to second segment in addition to a 3 dB SPL increase of intensity. On the *y*-axis, zero = "soft" and 100 = "loud." Continuous responses were recorded with a sampling rate of 2 Hz.

segment in AAA and ABA versions of each stimulus contain the original and identical acoustic profile (defined by the symbol A). Mean loudness was overall lower throughout the second segment of the Mozart and Organ stimuli when reverberation was added. This is evident from the negative values observed in the second segment where reverberation was added in ABA versions. The third segment of both AAA and ABA versions of each stimulus retained the original and identical acoustic profiles. Inspection of results from the Mozart and Organ stimuli show that in the third segment, mean loudness between AAA and ABA versions did not return to comparable values, even though each version's acoustic profile in the third segment was identical. As can be seen in the top panel of Figure 2, the mean loudness response to the third segment of the ABA versions of Mozart and Organ stimuli was greater than the mean loudness response to the third segment of the AAA versions. The bottom panel in Figure 2 shows results for a similar comparison, but here between each segment of the AAA baseline version and each corresponding segment of the AB′A version. To summarize, the reverberation manipulation had an impact which was simple and clear for the Mozart and Organ stimuli but more complicated for the Wishart stimulus. As indicated above, our overall purpose was to perturb the relationship between intensity and loudness by means of the reverberation manipulation, and this was achieved.

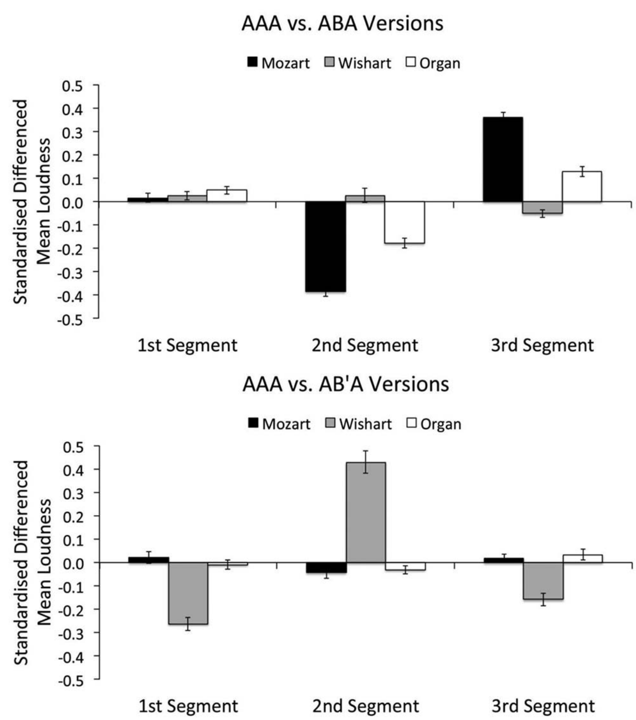

Figure 2. Differences in standardized mean loudness scores (*z*-scores) between AAA and ABA versions in the top panel and between AAA and AB′A versions in the bottom panel. A = original acoustic segments of each stimulus; B = reverberation added to second segment with the intensity profile of the original version closely maintained; B′ = reverberation added to second segment in addition to a 3 dB SPL increase of intensity. For each segment, a value greater than zero can be interpreted as an overall greater mean loudness response to either the ABA version in the top panel, or AB′A version in the bottom panel, relative to AAA versions. A value less than zero can be interpreted as an overall lower mean loudness response to that particular segment of each stimuli's ABA or AB′A version, relative to the corresponding segment of each stimuli's AAA version. A value of zero represents no difference in mean loudness between corresponding segments in ABA/AAA versions (top panel) and AB′A/AAA versions (bottom panel) of each stimulus.

Intensity Versus Loudness: Grand Average Time-Series Analysis Models of Affect

Our earlier exploratory result on Dvorak's Slavonic Dance (Dean et al., 2011) suggested that intensity is the dominant predictor of affect. If this is the case, then loudness in the present study will have little predictive role in models for each stimulus as a whole. In this case, AAA and ABA models of perceived affect expressed in Mozart and Organ stimuli should be reasonably equivalent because intensity does not vary between them, while loudness does. If on the other hand loudness is important, BIC values will be lower in models including loudness because we have perturbed the relationship between intensity and loudness through our experimental manipulation. If the impact of intensity is different in different segments of each stimulus, then models of the perception of A and B segments should be distinct from those of B′. If the impact of loudness is different in different segments of each stimulus, then models for the A and B′ segments will be distinct from those for the B segments. The results of these time-series analyses are shown for each stimulus separately in Tables 2–4. Four types of models for each version of each stimulus are presented and all permit spectral flatness as a potential predictor. The first type allows intensity but not loudness, the second allows loudness but not intensity, and the third allows intensity and loudness according to model optimization. In conditions comprising manipulations of reverberation, a fourth model is presented that allows for any impact of segment variation. In Tables 2–4, the numbers in "predictors" column and the autoregressive component column refer to the specific lags of the predictor or measured response variable darousal, respectively, that significantly contributed to prediction within the specified model. As results are qualitatively similar for perceived arousal and perceived valence, we focus on arousal in this section of analysis using group-mean time-series. We present both arousal and valence analyses in the subsequent section analyzing all individual time-series together using CSTSA.

Comparisons between the models are most readily made on the basis of the differences in BIC, residual squares, or correlations between model and data between versions 1 (AAA) and 2 (ABA) or 1 and 3 (AB′A). On this basis, loudness is generally a better predictor of perceived arousal than is intensity (compare models of Type 2 vs. Type 1). Intensity is a poorer sole predictor for the manipulated versions than the AAA control version in the Mozart and Wishart stimuli. In models of Type 3, where all variables were permitted as predictors, intensity often remained a useful predictor together with the more powerful loudness variable. The exceptions were: (a) the simpler "organ" oscillation stimulus, where intensity ceased to be a significant predictor for the ABA and AB′A versions; and (b) the Wishart stimulus, where intensity ceased to be significant for the AB′A version. Overall, this evidence points to a dominant role of loudness and a subsidiary role of intensity as predictors in models of perceived affect expressed by music.

In case small changes in spectral flatness due to the reverberation manipulations in B and B′ segments of each stimulus were influential (spectral flatness was reduced by a maximum of approximately one unit), segmental analyses were conducted that included intensity, loudness, and like all the other models, spectral flatness. Models of Type 4 in Tables 2–4 show that in no case were the segments mechanistically distinct from the point of view of models of arousal: the addition of separate variables for the B segments of the intensity, loudness, or spectral flatness predictors did not enhance the model over those where these factors were

Table 2
Time-Series Models of Arousal for the Mozart Versions

Stimulus version/modeled affect/model type	Acoustic and perceived loudness predictors	AR	BIC	Prediction squares as % of total squares	Correlation between fit and data
AAA Version					
1. darous/without loudness	L(1–6).dintensity, L(1–6).dspectralf	1, 3	1125.07	70.75	.84
2. darous/without intensity	L(0–4).dloudness	1	1059.92	73.71	.86
3. darous/allowing intensity and loudness	L(1–3).dintensity, L(0–3).dloudness, L(2–4).dspectralf	1	1057.20	76.61	.87
ABA Version					
1. darous/without loudness	L(1–6).d.intensity, L(1–5).dspectralf	1, 2	1234.94	53.73	.73
2. darous/without intensity	L(0, 1, 5).dloudness	1	1123.96	61.60	.78
3. darous/allowing intensity and loudness	L(7).dintensity, L(0, 5).dloudness, L(1, 2).dspectralf	1, 2	1116.30	64.92	.81
4. darous/allowing segment variation	Segments did not provide benefit				
AB′A Version					
1. darous/without loudness	L(1–6).dintensity, L(1–5).dspectralf	1, 2	1247.28	46.20	.68
2. darous/without intensity	L(0, 2, 3, 6).dloudness, L(1–4).d.spectralf	1	1138.98	61.52	.78
3. darous/allowing intensity and loudness	As for model type 2: Intensity did not contribute.				
4. darous/allowing segment variation	Segments did not provide benefit				

Note. AR = autoregressive components; BIC = Bayesian information criteria; 'dspectralf' abbreviates dspectral flatness (and similarly in other tables). Note that dspectral flatness is assessed as a predictor in all models in this paper, but eliminated from some (when not shown) as being unnecessary. All models had white noise residuals free of autocorrelation. All correlations (model:data) shown are significant at $p < .05$. Models that do not differ in BIC by at least 4.6 are not strongly distinct. The correlation values between model and data necessarily give a more positive impression of model quality than do the percent squares explained, but the relation between them remains fairly consistent across all stimuli and their versions. The correlations are given partly for comparison with previous publications.

Table 3
Time-Series Models of Arousal for the Wishart Versions

Stimulus version/modeled affect/model type	Acoustic and perceived loudness predictors	AR	BIC	Prediction squares as % of total squares	Correlation between fit and data
AAA Version					
1. darous/without loudness	L(1–9).dintensity, L(4).dspectralf	1, 5	1146.96	37.67	.61
2. darous/without intensity	L(0, 2, 3, 5).dloudness, L(9*).dspectralf	1	1105.53	39.49	.63
3. darous/allowing intensity and loudness	As for model type 2: Intensity did not contribute.				
ABA Version					
1. darous/without loudness	L(1–6).dintensity, L(9*).dspectralf	2, 3	1220.64	30.29	.55
2. darous/without intensity	L(0, 3, 5).dloudness, L(9*).dspectralf	2, 3	1180.78	36.07	.60
3. darous/allowing intensity and loudness	L(1–3).dintensity, L(0, 3, 5).dloudness, L(9*).dspectralf	3	1173.93	40.25	.63
4. darous/allowing segment variation	Segments did not provide benefit in the models				
AB′A Version					
1. darous/without loudness	L(1–5, 10).dintensity, L(7).dspectralf	4	1267.14	26.52	.51
2. darous/without intensity	L(0, 2).dloudness, L(9*).dspectralf	—	1222.13	32.75	.57
3. darous/allowing intensity and loudness	As for model type 2: Intensity did not contribute.				
4. darous/allowing segment variation	Segments did not provide benefit in the models				

Note. AR = autoregressive components; BIC = Bayesian information criteria. All models had white noise residuals free of autocorrelation. All correlations (model:data) shown are significant at $p < .05$. Predictor lags which were not significant individually but required for the optimal model are indicated with "*".

treated as operating in a constant manner throughout the stimulus. This observation was complemented by the fact that the lags and coefficients—which were functional within the segmental models—were very similar to those within the corresponding nonsegmental models. These results show that spectral flatness was not a mediator of the effects on perceived arousal from manipulations in the B and B′ segments.

Intensity Versus Loudness: CSTSA of Arousal

CSTSA of all nine conditions (3 stimuli × 3 versions) was conducted with possible random effects for each stimulus to test the strength of the conventional time-series analyses and the predictive role of intensity and loudness. The CSTSA analyses also ask: (a) whether the fixed effects of these predictors might differ

Table 4
Time-Series Models of Arousal for the Organ Versions

Stimulus version/modeled affect/model type	Acoustic and perceived loudness predictors	AR	BIC	% predicted squared values	Correlation between fit and data
AAA Version					
1. arous/without loudness	L(10–12, 13*).intensity, L(1, 2, 4).spectralf	1, 2, 4	1280.07	78.26	.99
2. arous/without intensity	L(0, 11, 15).loudness, L(4, 5, 7).spectralf	1, 2, 8	1231.83	81.15	.99
3. arous/allowing intensity and loudness	L(8, 10, 13*).intensity, L(0).loudness	1, 2, 8	1221.21	98.37	.99
ABA Version					
1. arous/without loudness	L(8, 10, 17).intensity, L(1, 2).spectralf	1, 2, 4, 5	1276.11	98.55	.98
2. arous/without intensity	L(0, 8, 13*).loudness, L(2, 5).spectralf	1, 2, 14	1260.20	98.70	.98
3. arous/allowing intensity and loudness	As for model type 2: Intensity did not contribute.				
4. arous/allowing segment variation	Segments did not provide benefit in the models				
AB′A Version					
1. arous/without loudness	L(1, 8, 9–11).intensity, L(1).spectralf	1, 2	1302.12	98.70	.99
2. arous/without intensity	L(0, 17).loudness, L(3, 6).spectralf	1, 2	1244.54	98.89	.99
3. arous/allowing intensity and loudness	As for model type 2: Intensity did not contribute.				
4. arous/allowing segment variation	Segments did not provide benefit in the models				

Note. AR = autoregressive components; BIC = Bayesian information criteria. Organ conditions did not require differencing, as they were already stationary. All models had white noise residuals free of autocorrelation. All correlations (model:data) shown are significant at $p < .05$. Predictor lags which were not significant individually but required for the optimal model are indicated with"*".

between segments (in a segmented model, parameters are permitted to vary between segments, which are here treated in sequence as A, B, and C, as distinct from the corresponding "long" variable which refers to the unsegmented whole); and (b) whether there are random effects between each stimulus. As in the earlier analyses, spectral flatness is considered as a potential predictor in all these models. The models of Table 5 show percentage squares fit and correlations between model and data that are close to the mean values across all nine segments treated individually (cf. Tables 2–4); thus, they are quite good. This result suggests that: (a) the models in operation are quite similar across all stimuli, although there are random effects focused on loudness and autoregression; and (b) segments do show differences in their mode of operation because the addition of segments provides improved predictions from the perceived loudness segments. CSTSA in Table 5 (bottom) also indicates reverberation as a random effects predictor. Reverberation was not a fixed effects predictor, and its random effects could not replace the fixed effects of intensity and loudness. The fact that spectral flatness is eliminated from all models involving loudness segments confirms the earlier indication that it does not mediate the influence of reverberation on perceived affect.

Impact of Intensity on Loudness in Models of Arousal: Vector Autoregressive Analysis

The results so far indicate that intensity and loudness have complementary predictive roles in models of perceived arousal. Indeed, while loudness is generally more important, intensity often retains an influence. However, the possibility that effects of intensity on loudness might mediate intensity's effects on perceived arousal is not directly considered in the analyses above. Here we assessed more directly and more conservatively whether loudness might in fact be the sole mediator of the predictive role of intensity on perceived arousal. This is accomplished by taking the grand average data for perceived arousal for each of the nine conditions (3 stimuli × 3 versions) and use Vector Autoregression (VARX) to: (a) model such interrelations between predictors; and (b) assess whether influences of both intensity and loudness on perceived arousal remain in optimized models.

Table 5
Models of Darousal by Cross Sectional Time-Series Analysis of All Stimulus Versions

Parameter modeled affect/model type	Acoustic and perceived loudness fixed effects predictors	AR	Random effects components by stimulus	*SD* Residual	BIC	% predicted squares values	Correlation between fit and data
Without loudness segments							
darouslong/Fixed Effects only	L(1, 3–6).dintenslong, L(0, 8, 17).dloudlong L(2–7).dspecflong	1, 4	—	2.21	10451.06	66.62	.82
darouslong/Random Effects	L(1, 3–6).dintenslong L(0, 17*).dloudlong	1	L(0, 8).dloudnesslong L(1–4).darouslong	2.10	10278.36	70.61	.84
With loudness segments (instead of loudness long)							
darouslong/Fixed Effects only	L(1).dintenslong L(0, 17).dloudnessA L(0).dloudnessB L(0, 17).dloudnessC	1, 4	—	2.25	10427.93	65.44	.81
darouslong/Random Effects	L(1).dintenslong L(0, 17).dloudnessA L(0).dloudnessB L(0, 17).dloudnessC	1	L(0).dloudnessA, L(1–4).darouslong	2.11	10250.19	70.02	.84
With reverberation as predictor (together with loudness long)							
darouslong/Fixed Effects only	As above for "Without loudness segments": reverb no effect	1, 4	—	2.21	10451.06	66.62	.82
darouslong/Random Effects	As above for "Without loudness segments": reverb no effect. i.e. L(1, 3–6).dintenslong L(0, 17*).dloudlong	1	Random effects by stimulus (as above): L(0, 8).dloudnesslong L(1–4).darouslong Random effects by Reverb status: L(1).dintenslong L(1).darouslong	2.08	10277.37	71.22	.85

Note. AR = autoregressive components; BIC = Bayesian information criteria. The use of the term "long" refers to time-series data that are considered as a single complete series, instead of segmented division of time-series responses as a function of A, B, and B′ segments. Random effects were only accepted when the likelihood ratio test comparing their model with that containing only fixed effects was highly significant ($p < .001$). A = original acoustic segments of each stimulus; B = reverberation added to second segment with the intensity profile of the original version closely maintained; B′ = reverberation added to second segment in addition to a 3 dB SPL increase of intensity.
* Indicates parameters were required but not individually significant.

VARX is a multivariate analogue of ARX where vectors of variables are modeled (rather than one, darousal, as above) and assessed for a potentially reciprocal interaction that is bidirectional. For example, we have shown the perception of arousal is often influenced by loudness. However, it might be that perceived loudness is also influenced by arousal. On the other hand, perceived loudness cannot influence acoustic intensity, and so intensity can be treated as an eXogenous variable, as it is in the ARX models above; but loudness may mediate all of the influences of intensity.

We have provided detailed introductions on the use of VARX in the analysis of continuous real-time responses in previous work (Bailes & Dean, 2012; Dean & Bailes, 2010; Dean et al., 2014a, 2014b; Olsen, Dean, & Stevens, 2014). Nevertheless, VARX can be thought of as a multivariate development of ARX that is subject to the same criteria for model selection and for quality of models. There are two main procedural differences here. First, a special VAR "selection order criterion" method (the *varsoc* command in STATA) is used to assess the economical and judicious maximum order of autoregression and lags of predictors. Because of the model interactions and complexity these are normally lower orders than with ARX models. Second, the models discussed are optimized in each case for the same purpose, different from ARX; that is, the simultaneous prediction of the two dependent (endogenous) variables darousal and dloudness. As with ARX, VARX models can be pruned by assessment of individual significance of coefficients, or by comparing overall information criteria during model selection, although the same predictors are necessarily used with respect to both dependent variables (DVs).

To aid comprehension of the VARX data, a complete optimized model for the AAA (original) version of the Mozart stimulus is shown in the Appendix and described below. Table 6 shows the arousal parts of the models for each of the nine stimulus versions under study. In considering the coefficients in the Appendix and Table 6, it is important to note that the scales of dintensity and dloudness vary differently between stimuli. For example, with the very simple intensity oscillations in the Organ stimulus, the range of dintensity is about half the range of dloudness, and thus were they to have equivalent influence on darousal, the corresponding parameters for dintensity would need to be roughly twice as large as those for dloudness. In addition, the spectral flatness measure (Wiener's entropy) used to represent timbral flux ranges from negative infinity to zero (it does not take on positive values), so its coefficients in models need to be interpreted in that light. Detailed analyses of the quantitative impacts of unit proportional changes of the predictors are not shown, but these can be demonstrated by impulse response analyses as also presented in earlier work (Dean & Bailes, 2010; Olsen et al., 2014). Rather, our interest here is whether both intensity and loudness contribute to models of perceived affect, and whether there is any indication of changed relative impact of intensity and loudness between the different manipulations of a single stimulus; changes that are more closely reflected by the relative changes in their coefficients.

Given this caution on quantitative interpretation, we can deduce the following from the results of the Mozart stimulus presented in the Appendix. First, the model gives good fit for both DVs, darousal and dloudness (as shown by the R^2 values). These values may be roughly compared with the "prediction squares as %" values in the preceding ARX tables. Second, lags of darousal, dloudness, dintensity and dspectral flatness are all needed in the model, and highly significant for darousal. All the predictors except darousal are also highly significant for the model of dloudness. As the range of loudness values is considerably greater than intensity, and yet the parameter coefficients are similar, the results are supportive of our previous ARX deductions that dloudness makes a larger contribution to the perceived arousal model than does dintensity.

Table 6 shows the arousal component of the optimized VARX models for each of the ABA and AB′A stimulus versions. Bear in mind that the loudness component of the models is not detailed in this table, but in every case it is both darousal and dloudness that are simultaneously modeled. The data shown are explained fully in the Appendix. The results are clear-cut: a significant predictive role for both acoustic intensity and perceived loudness is retained in every case, confirming that they have a complementary impact in models of perceived arousal. Their relative predictive roles do, however, change subtly between stimulus versions (as reflected by changing lag structure and changing coefficients). Furthermore, the relative goodness of fit from the different models is in parallel to those from the ARX above, and the predictive role of spectral flatness in the Mozart and Organ extracts is confirmed. A role for spectral flatness in the Wishart is no longer detectable, consistent with its nonsignificant coefficients in the models presented in Table 3. The loudness components of the models (not shown) are in agreement with the complete model shown in the Appendix, in that intensity and spectral flatness both contribute with autoregression and in a few cases, there is an influence of perceived arousal in the loudness models.

CSTSA of Valence

Finally, modeling of continuously perceived valence is presented in Table 7. Here, CSTSA models are optimized by taking all nine stimulus versions together by analogy with the model development of arousal presented in Table 5. This approach aimed to directly discern the possible relative influences of intensity, loudness, loudness segmentation, reverberation, and autocorrelation, as well as any random effects. As can be seen in Table 7, results showed that intensity and loudness participate in optimal models, together with spectral flatness. The optimal models are less predictive than those for arousal, as has been observed with several previous pieces of music (Bailes & Dean, 2012). Loudness segments have a slight impact on models of valence, and there is evidence that the fixed effects impact of loudness varies across segments. Furthermore, there are random effects of reverberation expressed through loudness in addition to autoregression and random effects on each stimulus expressed through loudness. The complementary predictive roles of intensity and loudness are thus again supported. While spectral flatness remained a predictor in most of these models, it did not function differently across the three segments of each stimulus, again confirming that it did not mediate the changed relationships with perceived valence in the manipulated segments.

Table 6
Vector Autoregressive (VAR) Analyses of Perceived Darousal Jointly With Dloudness for Optimized Models of Manipulated Versions of Each Stimulus

1. Mozart: ABA (abbreviated below as moz2)					
darousal equation R^2	dloudness equation R^2		Overall model BIC		
.50	.58		2228.4		
DV darousal: Predictors (6)	Coefficient	*SE*	*p*-values	95% Confidence intervals	
L1.dmoz2arousal	.50	.06	.001	.38	.62
L1.dmoz2loudness	.17	.08	.022	.02	.32
L1.dmoz2intensity	.32	.05	.001	.22	.41
L2.dmoz2intensity	.07	.04	.074	−.01	.14
L1.dmoz2spectralflatness	.59	.14	.001	.30	.87
L2.dmoz2spectralflatness	.96	.18	.001	.61	1.31
2. Mozart: AB′A (abbreviated below as moz3)					
darousal equation R^2	dloudness equation R^2		Overall model BIC		
.40	.50		2386.2		
DV darousal: Predictors (6)	Coefficient	*SE*	*p*-values	95% Confidence intervals	
L1.dmoz3arousal	.47	.06	.001	.35	.58
L1.dmoz3loudness	.17	.06	.006	.05	.30
L1.dmoz3intensity	.22	.05	.001	.12	.32
L2.dmoz3intensity	.02	.04	.557	−.06	.11
L1.dmoz3spectralflatness	.42	.15	.006	.12	.72
L2.dmoz3spectralflatness	.56	.19	.003	.19	.93
3. Wishart: ABA (abbreviated below as red2)					
darousal equation R^2	dloudness equation R^2		Overall model BIC		
.28	.29		2467.3		
DV darousal: Predictors (6)	Coefficient	*SE*	*p*-values	95% Confidence intervals	
L2.dred2arousal	.16	.06	.008	.04	.29
L3.dred2arousal	.17	.06	.006	.05	.29
L2.dred2loudness	.09	.06	.096	−.02	.21
L3.dred2loudness	.17	.06	.002	.06	.29
L1.dred2intensity	.18	.02	.001	.14	.23
L2.dred2intensity	.07	.02	.002	.03	.12
4. Wishart: AB′A (abbreviated below as red3)					
darousal equation R^2	dloudness equation R^2		Overall model BIC		
.21	.44		2444.0		
DV darousal: Predictors (6)	Coefficient	*SE*	*p*-values	95% Confidence intervals	
L1.dred3arousal	.11	.07	.087	−.02	.24
L2.dred3arousal	−.05	.07	.466	−.18	.08
L1.dred3loudness	.12	.07	.092	−.02	.27
L3.dred3loudness	.21	.07	.003	.07	.34
L1.dred3intensity	.18	.03	.001	.13	.23
L2.dred3intensity	.07	.03	.016	.01	.13
5. Organ ABA					
darousal equation R^2	dloudness equation R^2		Overall model BIC		
.67	.89		2604.6		
DV darousal: Predictors (7)	Coefficient	*SE*	*p*-values	95% Confidence intervals	
L1.dorgan2arousal	.40	.06	.001	.29	.52
L2.dorgan2arousal	.00	.07	.961	−.12	.13
L1.dorgan2loudness	.26	.07	.001	.12	.40
L3.dorgan2loudness	−.21	.04	.001	−.30	−.13
L1.dorgan2intensity	.39	.11	.001	.16	.63
L2.dorgan2intensity	−.14	.12	.245	−.39	.10
L1.dorgan2spectralflatness	−.96	.44	.029	−1.83	−.10

(*table continues*)

Table 6 (*continued*)

6. Organ AB′A					
darousal equation R^2	dloudness equation R^2		Overall model BIC		
.80	.92		2501.8		
DV darousal: Predictors (8)	Coefficient	*SE*	*p*-values	95% Confidence intervals	
L1.dorgan3arousal	.54	.06	.001	.43	.65
L3.dorgan3arousal	.12	.06	.053	.00	.24
L1.dorgan3loudness	.29	.07	.001	.15	.43
L3.dorgan3loudness	−.33	.05	.001	−.42	−.24
L2.dorgan3intensity	−1.08	.26	.001	−1.59	−.58
L1.dorgan3spectralflatness	−1.60	.35	.001	−2.29	−.91
L2.dorgan3spectralflatness	−2.66	.56	.001	−3.75	−1.57
L3.dorgan3spectralflatness	−1.71	.86	.045	−3.39	−.04

Note. In each case, the VAR jointly modeled two DVs (endogenous variables): darousal and dloudness. Lags of autoregression, dintensity and dspectralflatness were assessed as candidate predictors during model selection. Models did not require an intercept (as is normal for models of differenced variables). All darousal and dloudness models had $p < .001$ and their respective R^2 are shown. The predictor coefficients are only shown for the darousal equations within the joint VAR models; a few of those shown are nonsignificant in the darousal model, but retained because of their significance in the dloudness model component (the data of which are not shown). Note that the BIC values are higher than those for ARX largely because in VARX it is a vector of DVs (here two) that are being modeled, whereas in ARX only one is being modeled. Only BIC for models of exactly the same vector of DVs can be meaningfully compared with each other (as occurs during model selection). BIC = Bayesian information criteria.

Discussion

The present study was designed to compare the predictive power of continuous intensity and loudness in time-series models of perceived affect. This was achieved by manipulating continuous loudness perception with synthesized reverberation, but without concurrent changes in the intensity profiles of three musical stimuli. This work followed from an analysis reported in Dean et al. (2011), where acoustic intensity profiles but not continuous loudness profiles contributed to an optimal model of affect in response to Dvorak's *Slavonic Dance Opus 46, No. 1*. Here, stimuli comprised music from Western classical (Mozart's *Piano Concerto 21, K467*) and electroacoustic (Wishart's *Red Bird, a political prisoner's dream*) genres, as well as a synthesized single-timbre organ-like condition with continuous cyclic increases and decreases of intensity.

Overall, time-series models indicate that continuous loudness is the stronger predictor of perceived affect than acoustic intensity. This result suggests that the lack of requirement for continuous loudness in modeling perceived affect from the Dvorak Slavonic Dance mentioned above cannot be generalized across additional musical stimuli and genres with varied complexity and familiarity. Indeed, and perhaps not surprisingly, it is real-time perceived loudness that plays the more significant predictive role. However, the models presented here do show

Table 7
Models of dValence by Cross Sectional Time-Series Analysis of All Stimulus Versions

d.valence: Approach/model	Fixed effects predictors	AR	Random effects components by stimulus	*SD* (residual)	BIC	% predicted squares values	Correlation between fit and data
Fixed effects—no segments	L(2, 10*).dintenslong, L(0, 2*, 4).dloudlong, L(4).dspecflong	1, 3, 4, 5		1.26	7997.97	15.00	.38
Fixed effects—loudness segments permitted	L(2, 10*).dintenslong, L(0, 4).dloudnessA, L(0, 4).dloudnessB, L(0).dloudnessC	1, 3, 4		1.26	7945.66	15.41	.38
Random effects—no segments	L(2, 10*).dintenslong, L(0).dloudlong, L(4).dspecflong	1, 4	*On stimulus*: L(0).dloudlong *On reverb*: L(0, 2).dloudlong L(1).dvalenlong	1.22	7900.08	21.98	.47
Random effects—loudness segments permitted	L(2, 10*).dintenslong, L(0).dloudnessA, L(0).dloudnessB, L(0).dloudnessC, L(4).dspecflong	1, 4	*On stimulus*: L(0*).dloudlong *On reverb*: L(0, 2).dloudlong L(1).dvalenlong	1.21	7849.58	22.86	.47

Note. AR = autoregressive components; BIC = Bayesian information criteria. The use of the term "long" refers to time-series data that are considered as a complete series, instead of segmented division of time-series responses as a function of A, B, and B′ segments. As before, only Random Effects models with LR test values showing a significant improvement at $p < .001$ are allowed.
* Indicates parameters were required but not individually significant.

that acoustic intensity has some actions which eventually influence perceived affect, and which do not depend on mediation from perceived loudness. This is consistent with the preliminary observations from responses to the excerpt of Dvorak (Dean et al., 2011). Furthermore, spectral flatness again makes some contributions to perceived affect (Bailes & Dean, 2012; Dean & Bailes, 2010, 2011).

Further illumination of the Mozart ARX time-series results can be made by comparison with analyses of the same stimulus reported in a separate but related study. Specifically, the models reported in the present study have a larger concordance between fit and data than do models that include intensity and spectral flatness in the context of investigating listeners' real-time engagement with a piece of music (Olsen et al., 2014). Besides differing participants and overall stimulus selection and diversity, a notable difference between these experiments is that in the present study, each stimulus was presented three times. At the completion of the second block in the present experiment, participants would have heard two versions with two different loudness profiles. At that point, the intensity and reverberation manipulations in the second segment may have focused participant attention on the intensity profile in general across the whole stimulus. This is because the manipulations of intensity, loudness, and reverberation all vary intensity and/or loudness and may guide attention to the intensity profile/loudness response throughout each stimulus.

It is clear from the data presented here that perceptual loudness and physical acoustic intensity have complementary roles in prediction of perceived affective response to music. This conclusion was not only supported by the ARX results, in which the two predictors are assumed to be independent, but also in the VARX results, in which their interactions are directly modeled to detect whether loudness entirely mediated the effect of intensity. The VARX results confirm that the two predictors have complementary roles, and so the question "why" remains. A possible answer may lie in differences in the time-course of perceptual impact from acoustic intensity. When presented abruptly and at relatively high levels, acoustic intensity can elicit fast and automatic physiological arousal responses (e.g., ~80 ms acoustic startle response to a high intensity acoustic stimulus measured from electromyographic activity; Valls-Solé et al., 1995). This and later consequent physiological arousal responses are likely to influence perception of related features such as affect. Loudness, on the other hand, is a psychological phenomenon that requires approximately 300 ms to establish an overall short-term perceptual impression (Chalupper & Fastl, 2002; Glasberg & Moore, 2002). Loudness may thus impact perceived affect differently than the immediate and delayed effects of heightened physiological arousal. The difference between these physiological and psychological pathways may begin to explain why perceptual loudness and physical acoustic intensity have complementary roles in prediction of perceived affect expressed by music. Considerable further work will be required to complete an understanding of this hypothesis.

In terms of the reverberation manipulation, loudness was significantly affected in the Mozart stimulus and the Organ stimulus. For these two stimuli, there were no differences in mean loudness between the first segment in the original AAA version and the first segment of the ABA version. This is because the first segment in AAA and ABA versions of each stimulus contained the original and identical acoustic profile. However, in the second segment where reverberation was added to the ABA version, loudness decreased relative to the second segment in AAA version. After mean loudness decreased in response to the addition of reverberation, mean loudness increased in the third segment of the ABA version relative to the third segment of the AAA version, even though the third segments in both of these versions were once again acoustically identical. Indeed, there seems to be a kind of overcompensation or "overshoot" of loudness at the beginning of the third segment for those stimuli where reverberation initially decreased loudness in the second segment. This could be explained by the transition from the end of the second segment (B) to the beginning of the third segment (A) in the ABA version, which led to an abrupt and continuous increase in loudness due to the removal of the reverberation and ultimately resulting in the perceptual "overshoot." For the Wishart stimulus, reverberation had less impact on loudness. This is probably due to sparse and relatively "empty" nature of the music, where events are more scattered and simplified when compared to the large number of sound sources heard in the orchestral work of Mozart, for example. Future research is required to further understand perceptual effects of reverberation in music, especially in terms of perceived loudness and affect, and such investigations have implications for a variety of fields such as music perception and cognition, psychoacoustics, and audio engineering.

In sum, data from the present study show that both intensity and loudness are predictors of perceived arousal and to lesser extent, valence, but that loudness is often more powerful and sometimes dominant to the point of excluding intensity. Furthermore, the use of reverberation as a tool to vary loudness without a concurrent variation in acoustic intensity was successful in its design and shows promise for future research.

References

Bailes, F., & Dean, R. T. (2012). Comparative time series analysis of perceptual responses to electroacoustic music. *Music Perception, 29,* 359–375. http://dx.doi.org/10.1525/mp.2012.29.4.359

Butler, R. A., Levy, E. T., & Neff, W. D. (1980). Apparent distance of sounds recorded in echoic and anechoic chambers. *Journal of Experimental Psychology: Human Perception and Performance, 6,* 745–750. http://dx.doi.org/10.1037/0096-1523.6.4.745

Canévet, G., & Scharf, B. (1990). The loudness of sounds that increase and decrease continuously in level. *Journal of the Acoustical Society of America, 88,* 2136–2142. http://dx.doi.org/10.1121/1.400110

Chalupper, J., & Fastl, H. (2002). Dynamic loudness model (DLM) for normal and hearing-impaired listeners. *Acta Acustica United with Acustica, 88,* 378–386.

Dean, R. T., & Bailes, F. (2010). 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. (2011). Modelling perception of structure and affect in music: Spectral centroid and Wishart's Red Bird. *Empirical Musicology Review, 6,* 131–137.

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. http://dx.doi.org/10.1080/17459737.2014.928753

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. http://dx.doi.org/10.1080/17459737.2014.928752

Dean, R. T., Bailes, F., & Schubert, E. (2011). Acoustic intensity causes perceived changes in arousal levels in music: An experimental investigation. *PLoS One, 6,* e18591. http://dx.doi.org/10.1371/journal.pone.0018591

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.

Fletcher, H., & Munson, W. A. (1933). Loudness, its definition, measurement and calculation. *Journal of the Acoustical Society of America, 5,* 82–108. http://dx.doi.org/10.1121/1.1915637

Florentine, M., Popper, A. N., & Fay, R. R. (Eds.). (2011). *Loudness*. New York, NY: Springer. http://dx.doi.org/10.1007/978-1-4419-6712-1

Gingras, B., Marin, M. M., & Fitch, W. T. (2014). Beyond intensity: Spectral features effectively predict music-induced subjective arousal. *The Quarterly Journal of Experimental Psychology, 67,* 1428–1446.

Glasberg, B. R., & Moore, B. C. J. (2002). A model of loudness applicable to time-varying sounds. *Journal of the Audio Engineering Society, 50,* 331–342.

Lee, D., Cabrera, D., & Martens, W. L. (2012). The effect of loudness on the reverberance of music: Reverberance prediction using loudness models. *The Journal of the Acoustical Society of America, 131,* 1194–1205. http://dx.doi.org/10.1121/1.3676602

Mershon, D. H., Desaulniers, D. H., Kiefer, S. A., Amerson, T. L., Jr., & Mills, J. T. (1981). Perceived loudness and visually-determined auditory distance. *Perception, 10,* 531–543. http://dx.doi.org/10.1068/p100531

Moorer, J. A. (1979). About this reverberation business. *Computer Music Journal, 3,* 13–28. http://dx.doi.org/10.2307/3680280

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. (2014). Intensity dynamics and loudness change: A review of methods and perceptual processes. *Acoustics Australia/Australian Acoustical Society, 42,* 159–165.

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. http://dx.doi.org/10.1037/pmu0000044

Olsen, K. N., & Stevens, C. J. (2013). Psychophysiological response to acoustic intensity change in a musical chord. *Journal of Psychophysiology, 27,* 16–26. http://dx.doi.org/10.1027/0269-8803/a000082

Olsen, K. N., Stevens, C. J., & Tardieu, J. (2010). Loudness change in response to dynamic acoustic intensity. *Journal of Experimental Psychology: Human Perception and Performance, 36,* 1631–1644. http://dx.doi.org/10.1037/a0018389

Ritossa, D. A., & Rickard, N. S. (2004). The relative utility of 'pleasantness' and 'liking' dimensions in predicting the emotions expressed by music. *Psychology of Music, 32,* 5–22. http://dx.doi.org/10.1177/0305735604039281

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39,* 1161–1178. http://dx.doi.org/10.1037/h0077714

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110,* 145–172. http://dx.doi.org/10.1037/0033-295X.110.1.145

Scharf, B. (1978). Loudness. In E. C. Catrerette & M. P. Friedman (Eds.), *Handbook of perception: IV* (pp. 187–242). New York, NY: Academic Press.

Schroeder, M. R. (1962). Natural sounding artificial reverberation. *Journal of the Audio Engineering Society, 10,* 219–223.

Schubert, E. (1999). Measuring emotion continuously: Validity and reliability of the two-dimensional emotion-space. *Australian Journal of Psychology, 51,* 154–165. http://dx.doi.org/10.1080/00049539908255353

Schubert, E. (2004). Modeling perceived emotion with continuous musical features. *Music Perception, 21,* 561–585. http://dx.doi.org/10.1525/mp.2004.21.4.561

Schubert, E. (2013). Reliability issues regarding the beginning, middle and end of continuous emotion ratings to music. *Psychology of Music, 41,* 350–371. http://dx.doi.org/10.1177/0305735611430079

Stecker, G. C., & Hafter, E. R. (2000). An effect of temporal asymmetry on loudness. *Journal of the Acoustical Society of America, 107,* 3358–3368. http://dx.doi.org/10.1121/1.429407

Stevens, S. S. (1956). The direct estimation of sensory magnitudes-loudness. *The American Journal of Psychology, 69,* 1–25. http://dx.doi.org/10.2307/1418112

Terrace, H. S., & Stevens, S. S. (1962). The quantification of tonal volume. *The American Journal of Psychology, 75,* 596–604. http://dx.doi.org/10.2307/1420282

Valls-Solé, J., Solé, A., Valldeoriola, F., Muñoz, E., Gonzalez, L. E., & Tolosa, E. S. (1995). Reaction time and acoustic startle in normal human subjects. *Neuroscience Letters, 195,* 97–100. http://dx.doi.org/10.1016/0304-3940(94)11790-P

Zahorik, P., Brungart, D. S., & Bronkhorst, A. W. (2005). Auditory distance perception in humans: A summary of past and present research. *Acta Acustica United with Acustica, 91,* 409–420.

Zahorik, P., & Wightman, F. L. (2001). Loudness constancy with varying sound source distance. *Nature Neuroscience, 4,* 78–83. http://dx.doi.org/10.1038/82931

Appendix

A Vector Autoregressive (VAR) Model Output of Perceived Arousal Jointly With Loudness

Mozart AAA (abbreviated below as moz1)

Modeled DVs (endogenous variables): d.moz1arous d.moz1loud Predictors: autoregressive lags(1) of the endogenous variables; IVs (exogenous variables) l(1/2).d.moz1intensity l(1/2).d.moz1spectralflatness BIC = 2187.99

	Parameters	R^2	*p*-values
Equation			
darousal	6	0.70	.001
dloudness	6	0.53	.001

	Coefficient	*SE*	*p*-values	95% Confidence interval	
DV darousal					
Predictors					
L1.dmoz1arousal	.59	.04	.001	.51	.68
L1.dmoz1loudness	.28	.05	.001	.18	.39
L1.dmoz1intensity	.30	.04	.001	.23	.37
L2.dmoz1intensity	.07	.03	.022	.01	.14
L1.dmoz1spectralflatness	.36	.12	.003	.13	.60
L2.dmoz1spectralflatness	.78	.14	.001	.50	1.05
DV dloudness					
Predictors					
L1.dmoz1arousal	.08	.05	.108	−.02	.17
L1.dmoz1loudness	.46	.06	.001	.34	.57
L1.dmoz1intensity	.37	.04	.001	.29	.45
L2.dmoz1intensity	.20	.03	.001	.14	.27
L1.dmoz1spectralflatness	.42	.13	.001	.16	.67
L2.dmoz1spectralflatness	.79	.15	.001	.49	1.09

Note. L1 and L2 indicate lags 1 and 2 respectively, and dseriesname indicates the first differenced form. The output first describes the parameters, overall fit (R^2), and probability (*p*-values) of the two components of the VAR model, the models for darousal and dloudness. Secondly, the output presents the coefficient for each predictor and its standard error, significance, and confidence intervals. Note that the two DVs are modeled together, and hence a predictor may be essential for one of the DVs and have a highly significant coefficient, but also be unimportant with a nonsignificant coefficient for the other (as here in the case of the L1 of darousal). However, by design the two components share the same predictors. The models did not include an intercept (they are not normally required for models of differenced variables). Note that the BIC value is higher than that for the corresponding ARX model, largely because in VARX it is a vector of DVs (here two) that are modeled, whereas in ARX only one is modeled. Only BIC for models of exactly the same vector of variables can be meaningfully compared with each other (as they are during model selection).

Received October 17, 2014
Revision received March 30, 2015
Accepted April 27, 2015 ■