Linking melodic expectation to expressive performance timing
and perceived musical tension

Running head: Linking melodic expectation to timing and tension

Bruno Gingras[1], Marcus T. Pearce[2], Meghan Goodchild[3], Roger T. Dean[4],

Geraint Wiggins[2], and Stephen McAdams[3]

[1] Institute of Psychology, University of Innsbruck, Austria

[2] Queen Mary, University of London, United Kingdom

[3] Schulich School of Music, McGill University, Canada

[4] MARCS Institute, University of Western Sydney, Australia

Corresponding author:

Bruno Gingras

Institute of Psychology, University of Innsbruck

Innrain 52f, Innsbruck, Austria A-6020

Email: brunogingras@gmail.com

ABSTRACT

This research explores the relationships between the predictability of musical structure, expressive timing in performance, and listeners' perceived musical tension. Studies analyzing the influence of expressive timing on listeners' affective responses are constrained by the fact that, in most pieces, the notated durations limit performers' interpretive freedom. To circumvent this issue, we focused on the unmeasured prelude, a semi-improvisatory genre without notated durations. In Experiment 1, twelve professional harpsichordists recorded an unmeasured prelude on a harpsichord equipped with a MIDI console. Melodic expectation was assessed using a probabilistic model (IDyOM) whose expectations have been previously shown to match closely those of human listeners. Performance timing information was extracted from the MIDI data using a score-performance matching algorithm. Time-series analyses show that, in a piece with unspecified note durations, the predictability of melodic structure measurably influences tempo fluctuations in performance. In Experiment 2, another 10 harpsichordists, 20 nonharpsichordist musicians, and 20 nonmusicians listened to the recordings from Experiment 1 and rated the perceived tension continuously. Granger causality analyses were conducted to investigate predictive relationships among melodic expectation, expressive timing, and perceived tension. Although melodic expectation, as modeled by IDyOM, modestly predicted perceived tension for all participant groups, neither of its components, information content or entropy, was Granger causal. In contrast, expressive timing was a strong predictor and was Granger causal. However, because melodic expectation was also predictive of expressive timing, our results outline a complete chain of influence from predictability of melodic structure via expressive performance timing to perceived musical tension.

Keywords: performance; tension; expectations; entropy; information content; timing

INTRODUCTION

Music is a uniquely human and cross-culturally universal means of conveying information using sequential streams of auditory events (Clarke, 1985, 1988; Palmer, 1989, 1996). Although there are major differences between language and music as modes of communication, both can be viewed as a means of transferring information (Koelsch et al., 2004). As in linguistic communication, musical communication involves the elaboration of a musical message (composition), its utterance (performance), and finally its perception and cognitive appraisal by one or more recipients (listeners). Music is an important part of many people's everyday lives (Sloboda, 2010), and the information it conveys can profoundly affect the mood, emotions, and well-being of listeners (Panksepp & Bernatzky, 2002). As such, research on music cognition broadens our perspective on human cognition in general by considering aspects that share functionalities with other types of cognition, but more importantly by considering its specificities.

Kendall and Carterette's (1990) theoretical model of musical communication, based on traditional Western art music, expresses this three-way relationship between the composer, the performer(s), and an audience. Following this model, we posit that a performer's interpretation of a musical composition (the message) is influenced by the structure of the composition, and that the listener's response to a particular performance of a piece reflects both the musical structure as notated in the score and the specific features of the performer's interpretation.

The purpose of this paper is to study the processes involved in the musical flow of information from the composer through the performer to the listener. The vast majority of empirical research in the field focuses exclusively on one or two of these aspects of musical

communication (e.g., Gabrielsson & Juslin, 1996; Juslin & Timmers, 2010). Here we explicitly attempt an empirical study of all three components in the communicative chain. In doing so, we first have to ask *what* is communicated and second, *how* it is communicated. It is the primary aim of the present study to address the latter question. First, however, we must establish our assumptions about the content of musical communication. In doing so, we begin with the listener, and then consider the structure of the composition, and finally the role of the performer.

The listener: Emotion, tension, expectation

Previous research suggests that listeners value music primarily for its impact on mood and emotion (Juslin & Laukka, 2004) and the content of musical communication is thought to be primarily affective rather than referential (Hanslick & Cohen, 1854; Meyer, 1956). There exist several mechanisms by which music can have an effect on the emotional state of listeners (Juslin & Vastfjall, 2008). Of these, we focus on expectation as a vehicle for studying the communication of emotion by music because it is more directly based on the structure of the music than other mechanisms (such as episodic memory or evaluative conditioning, see Juslin & Vastfjall, 2008), thereby facilitating the establishment of a link between compositional structure, performance and perception. Musical structures generate expectations for how the music will continue, and these expectations change dynamically as the musical structure unfolds in time. At any given moment, the music may confirm or violate the expectations of a given listener to different degrees. Real-time responses to music can be used to measure this continuous process (Krumhansl, 1996, 2002; Toiviainen & Krumhansl, 2003; McAdams, Vines, Vieillard, Smith, & Reynolds, 2004). In particular, the perception of musical tension, including its connection to aesthetic responses by listeners, has been studied extensively (Farbood, 2012; Fredrickson, 1995,

2000; Madsen, 1997; Madsen, Brittin, & Capperella-Sheldon, 1993; Madsen & Fredrickson, 1993).

Expectation is thought to be an important cognitive mechanism in language processing (Cristia, McGuire, Seidl & Francis, 2011; DeLong, Urbach & Kutas, 2005; Hale, 2006; Levy, 2008; Saffran, 2003), visual perception (Bar, 2007; Bubic, von Cramon & Schubotz, 2010; Egner, Monti & Summerfield, 2010) and motor sequencing (Wolpert & Flanagan, 2001). It is also an important psychological process determining the aesthetic experience of music listeners (Hanslick & Cohen, 1854; Huron, 2006; Meyer, 1956). In particular, violations of expectation are thought to lead to increases in tension (Meyer, 1956), as reflected in increased physiological arousal (Egermann, Pearce, Wiggins, & McAdams, 2013; Steinbeis, Koelsch, & Sloboda, 2006). Analysis of behavioral ratings of tension experienced while listening to Mozart's Piano Sonata in E-flat major, K. 282 (Krumhansl, 1996, 2002) showed that tension ratings covaried both with surface features, such as pitch height (register), and features associated with the stability of a pitch within a given tonal framework (tonal stability) or the tendency for certain pitches in a melody to move toward other pitches to a greater or lesser extent (tonal attraction) (Lerdahl, 2001). The latter features are likely to contribute to an experience of expectation violation in listeners enculturated with Western tonal music (Lerdahl & Krumhansl, 2007). However, a full model of musical expectation depends not only on the degree of tonal stability but also on the schematic effects of long-term musical exposure and on the presence of repeated or elaborated structures (e.g., motifs, theme and variation) within a musical work (Pearce, 2005). The goal of the present research is to examine the hypothesized relationship between the expectations generated by such a model and behavioral ratings of tension in musical listening.

The composition: Musical structure, predictability and expectation

Expectation can be viewed as a process of probabilistic prediction based on knowledge of musical structure acquired through statistical learning (Huron, 2006; Meyer, 1956; Pearce, 2005; Pearce & Wiggins, 2012). Note that similar processes are thought to operate in language perception (Saffran, 2003; Cristià et al., 2011; Hale, 2006; Levy, 2008) and other areas of perceptual processing (Friston, 2010; Perruchet & Pacton, 2006). There is empirical evidence that listeners acquire expectations for musical events through implicit statistical learning of regularities through long-term exposure to music (Krumhansl, 1990), and listeners' expectations also reflect the learning of statistical properties of the local context (Oram & Cuddy, 1995; Saffran, Johnson, Aslin, & Newport, 1999; Tillmann, Bharucha, & Bigand, 2000; Tillmann, Bigand, & Pineau, 1998). This process of statistical learning allows listeners to generate probabilistic predictions about forthcoming musical events, dependent on the prior musical context and previously acquired schematic expectations for the musical style in question (Krumhansl, Louhivuori, Toiviainen, Järvinen, & Eerola, 1999; Pearce, Ruiz, Kapasi, Wiggins, & Bhattacharya, 2010; Pearce & Wiggins, 2006). Consistent with an approach based on statistical learning, melodic pitch expectations vary between musical styles (Krumhansl et al., 2000) and cultures (Carlsen, 1981; Castellano, Bharucha, & Krumhansl, 1984; Eerola, 2004; Kessler, Hansen, & Shepard, 1984; Krumhansl et al., 1999), throughout development (Schellenberg, Adachi, Purdy, & McKinnon, 2002), and across degrees of musical training and familiarity (Krumhansl et al., 2000; Pearce et al., 2010).

In this study, we use a computational model of auditory expectation, IDyOM (Information Dynamics of Music), to specify precise, quantitative measures of structural predictability for each note in a melody. The model itself has been presented (Conklin & Witten, 1995; Pearce,

2005) and tested (Egermann et al., 2013; Hansen & Pearce, 2014; Omigie, Pearce & Stewart, 2012; Omigie, Pearce, Williamson, & Stewart, 2013; Pearce, Conklin, & Wiggins, 2005; Pearce et al., 2010; Pearce & Wiggins, 2004, 2006), so we restrict ourselves to a summary of features that are most relevant for the present study.

The central feature of the model is that it learns about sequential dependencies between notes in an unsupervised manner through exposure to melodies. In the context of this research, the model focuses specifically on the pitch of each note in a melody. Given such a representation of music, the model returns a conditional probability distribution of pitches for each note, given the preceding sequence of pitches. Two information-theoretic measures are derived from the predictive distribution generated for each note: *entropy* is a measure of the uncertainty of the predictive distribution associated with the melodic context (maximum entropy occurs for a distribution in which every pitch is equally likely to follow the context) whereas *information content* (IC, the negative log probability) is a measure of the unexpectedness of the note that actually follows.

Entropy reflects the uncertainty of the prediction before the next note is heard, whereas information content reflects how unexpected the next note is once it actually happens. If entropy is high, then every possible continuation of the sequence is equally likely to the model (i.e., it is very uncertain about the continuation). Alternatively, a very low entropy signifies that one note is highly expected, and other possibilities are very unexpected (i.e., the model is very certain). In this latter situation, if the certain prediction is correct then the next note is highly expected and the information content is low; if, however, the certain prediction is incorrect then the next note is unexpected and the information content is high. Therefore, accurate predictive models are associated with both low entropy predictions and low information content outcomes. For

example, a listener who is sure that the pitch of the next note in a melody will be a middle C (261.63 Hz) and not any other note has low predictive uncertainty at that point in time (before the next note is heard); another listener to the same melody who expects to hear any pitch between middle C and the A above it (440 Hz) with equal likelihood has higher predictive uncertainty. Note that predictive uncertainty is distinct from the expectedness of the event that actually happens. Our first listener is highly certain about the next event in the melody but may be incorrect (and surprised) if the note that actually follows is not a middle C.

The model is based on a Markov or *n*-gram model (Manning & Schütze, 1999, chap. 9), which computes the conditional probability of a note given the $n − 1$ preceding notes in the melody. The quantity $n − 1$ is called the order of the model, and basic Markov models have a fixed order (as a result, they suffer from the problem that previously unseen notes are estimated with a probability of zero). Basic Markov modeling is extended in two ways. First, the model is of variable order, computing a weighted average of probabilities from models of different order. This allows the system to benefit both from the structural specificity of longer (but relatively rare) contexts and the statistical power of more frequent (but less specific) low-order contexts. It also allows the model to generate probabilities for notes that have yet to appear in a given context.

Second, because perceptual representations of music are multidimensional (e.g., Levitin & Tirovolas, 2009; Shepard, 1982), the system has the ability to use a combination of different features, or *viewpoints*, to predict the properties of notes (Conklin & Witten, 1995; Pearce et al., 2005). For example, in predicting the pitch of the next note, the system may combine models trained on representations of *chromatic pitch*, *scale degree* (the position of the pitch in the tonal scale), *pitch interval* (the distance in pitch between two consecutive notes) or *contour* (whether

the pitch rises, falls, or remains the same). Two viewpoints may be *linked* (A × B, where A and B are the source viewpoints), so that notes are represented by pairs of values. For example, a link between pitch interval and scale degree would represent each note as a pair of values representing, respectively, the pitch interval and the scale degree of that note.

A separate model is used for each viewpoint specified, each of which generates a probability distribution. These distributions are subsequently combined using a weighted geometric mean into a single probability distribution governing the pitch of the next note. The viewpoints used may be specified by hand or optimized using a hill-climbing search through the space of possible sets of viewpoints for sets that minimize the information-content (or maximize the predictability) of the music being studied. The hill-climbing procedure starts with an empty set of viewpoints and at each step makes the single addition or deletion of a viewpoint, if any exist, which reduces the average information content of the musical piece by the greatest amount (Pearce, 2005).[1]

In this research we use two sets of viewpoints: first, a single *default* linked viewpoint combining pitch interval and scale degree (i.e., each note is represented as a pair of values – the first reflects the pitch interval in semitones from the previous note, the second reflects the chromatic scale degree from 0, the tonic, to 11, the leading tone). Second, we examine an *optimized* model in which viewpoints were selected through hill-climbing. The entire set of viewpoints considered and the optimized viewpoints selected are listed in the Appendix.

---

[1] It is possible that, during the hill-climbing procedure, the addition of subsequent viewpoints renders an existing one redundant or incompatible, so deleting it increases prediction performance (reducing information content). This was not the case for the optimized model used in the present paper (but see Pearce, 2005, p. 163, for an example).

Regardless of the model's configuration, its output is always a conditional probability distribution governing the pitch of the next note in a melody. As noted above, the entropy of this distribution reflects the predictive uncertainty of the model, whereas the information content (the negative log probability) of the note that actually follows reflects the unexpectedness of that note.

The performer: Improvisation and expressive performance timing

We turn now to the performer, the expressive intermediary between the musical structures created by the composer and the acoustic signal perceived by the audience. A substantial body of research on music performance has shown that performers' interpretive choices are informed by the formal structure of the musical pieces. For instance, it is well-established that performers, similarly to speakers (Cooper & Danly, 1981), tend to slow down at sectional boundaries, a process called "phrase-final lengthening" (Clarke, 1989; Palmer, 1989; Repp, 1990; Todd, 1985), and that the magnitude of this deceleration tends to correspond to the hierarchical importance of the boundaries (Shaffer & Todd, 1987). Similarly, the music communication hypothesis (Clarke, 1985, 1988; Palmer, 1989, 1996; Repp, 1992, 1995) suggests that tempo fluctuations are a means of conveying information about the grouping structure of a piece. Thus, the performer plays a crucial role in clarifying the musical structure for the audience.

Although tempo fluctuations on both local and large-scale levels have been identified as a major expressive device in music performance, most pieces in the Western repertoire specify a regular rhythmic structure, which greatly restricts the performers' latitude regarding note durations. Consequently, the link between information-theoretic measures derived from the music score and performance timing cannot be analyzed optimally. A few studies have addressed

this limitation by focusing on improvised performance (e.g., Large, Palmer, & Pollack, 1995), but this has the drawback, from the viewpoint of Kendall and Carterette's (1990) model of musical communication, of amalgamating the composer and performer functions, making it impossible to isolate the individual characteristics associated with each component. In order to circumvent this issue and investigate performance-induced effects on listeners' aesthetic responses, we focused on one particular genre of lute and harpsichord compositions, cultivated especially in France between 1650 and 1720, known as the *unmeasured prelude*. The unmeasured prelude is a semi-improvisatory genre that does not specify a rigid metrical framework—a pattern of strong and weak beats—and even leaves the durations of individual notes unspecified in many instances. It thus allows for a much wider variability in terms of expressive timing (i.e., variations in tempo made by a performer, which may differ between performances) than most Western musical genres (Gingras, Asselin, & McAdams, 2013), providing an ideal repertoire for investigating the links between information-theoretic measures derived from the musical score, expressive timing in performance, and listeners' aesthetic responses.

Following the musical communication hypothesis described above, we hypothesize that performers (just like listeners) are influenced by the predictability of the music they are performing. In perception, it has been shown that unexpected events take longer to process than expected ones (Bharucha & Stoeckig, 1986, 1987; Omigie et al., 2012; Pearce et al., 2010). We predict that this effect will extend to performance such that structural expectations influence expressive timing. In particular, we hypothesize that unexpected structures will be associated with a slowing of tempo and with a greater variability in expressive timing. Our decision to study harpsichord performance was also motivated by a concern to minimize the role of acoustic

intensity changes, since these are thought to be major influences on perception of affect and tension (Ilie & Thompson, 2006; Krumhansl, 1996). Given that the harpsichord has a relatively small dynamic range (Fletcher, 1977; Penttinen, 2006), particularly with respect to the performance of a particular note or chord, we expected that the nature and roles of performance timing would be readily detectable.

The present study

The purpose of the present research is to elucidate the cognitive mechanisms engaged in the model of musical communication proposed by Kendall and Carterette, involving the flow of information from the composer via the performer to the listener. More precisely, the cognitive mechanism we seek to elucidate is *expectation*, which we conceive as a dynamic process of predicting future elements in a sensory stimulus using both the preceding context within the stimulus and the prior experience of the observer (Bubic et al., 2010). We simulate expectation formally in terms of the information-theoretic quantities *information content* (reflecting the unexpectedness of an event) and *entropy* (the predictive uncertainty about which event will follow, before it appears). We derive these quantities using a computational model of auditory expectation (IDyOM) which has been shown to account accurately for listeners' pitch expectations in music (Egermann et al., 2013; Hansen & Pearce, 2014; Pearce et al., 2010; Pearce & Wiggins, 2006; Omigie et al., 2012, 2013). In this work we examine the role of expectation, as a cognitive process, in musical communication. Specifically, we hypothesize, on the basis of previous research (Meyer, 1956; Narmour, 1990; Huron, 2006), that composers introduce patterns that confirm or disconfirm expectations to different degrees, that performers perceive this ebb and flow of predictability and reflect it in their expressive performance, and

that listeners, in turn, experience heightened tension and relaxation in expressive performances at points where the process of expectation is, respectively, disrupted (by high unexpectedness or uncertainty) or stabilized (by low unexpectedness or uncertainty).

We conducted our investigation in two studies (Experiments 1 and 2) in which we report time-series analyses of continuous processing of an unmeasured prelude for harpsichord by Louis Couperin (1626-1661), the *Prélude non mesuré No. 7* (a score excerpt is provided in Figure 1). This prelude comprises a single unaccompanied melodic line in which pitches unfold successively, although harmonies are sometimes implied through the use of arpeggiation (a melodic succession creating the semblance of chords) and sustained notes.[2]

[Insert Figure 1]

Experiment 1 examines relationships between the changing structure of the music (specifically, structural predictability as assessed by the computational model of auditory expectation described above) and expressive timing in performances of the piece. Experiment 2 examines the impact of both structural predictability and expressive timing on the emotional experience of listeners, as measured by continuous tension ratings. As mentioned above, the

---

[2] In terms of musicologically conceived tension, the structure of the piece progresses from maximum stability to maximum instability and back based on its gestural structure (stable arpeggios enclosing less stable melodic passages) (see Goodchild, Gingras, & McAdams, 2015).

unmeasured prelude is an ideal genre to address these issues, due to the absence of a rigid metrical framework.

Melodic expectations, tempo variations in performance, and perceived musical tension exemplify the three vertices of a relationship involving musical structure, expressive strategies, and aesthetic responses. The present study focuses on these three parameters, aiming to quantify the individual components and assess the nature and directionality of their relationships with each other. We hypothesize that unexpected events (those with high IC) and points of predictive uncertainty (high entropy) will be associated with a slowing down of performance timing and greater variability between performers. In turn, we hypothesize that high IC and high entropy, in concert with their influence on performance timing, will be associated with a greater perceived tension in listeners.

EXPERIMENT 1: PERFORMANCE TIMING

Methods

Participants and procedure

Twelve professional harpsichordists, five female and seven male, participated in the study. They were professional harpsichordists from the Montreal (Canada) area or harpsichord students at McGill University in Montreal. Their average age was 39 years (range 21-61). They had played the harpsichord for a mean duration of 22 years (range 6-40). Seven had previously won prizes in regional, national, or international harpsichord competitions.

The harpsichordists were provided with the score of the *Prélude non mesuré No. 7* by Louis Couperin 4-6 weeks prior to the recording session. They performed the piece twice (with no

special directives) using the score. Performances took place in an acoustically treated studio on an Italian-style Bigaud harpsichord (Paris, Heugel) with two 8-foot stops. Only the back stop was used for the experiment. This harpsichord was equipped with a MIDI console, allowing precise measurement of performance parameters (note onset/offset and key press velocity).

The audio signal was recorded through two omnidirectional microphones (MKH 8020, Sennheiser Electronic GmbH, Wedemark, Germany). The microphones were located 1 m above the resonance board and were placed 25 cm apart. The audio and MIDI signals were sent to a PC computer through an RME Fireface audio interface (Audio AG, Haimhausen, Germany). Audio and MIDI data were then recorded using Cakewalk's SONAR software (Cakewalk, Inc., Boston, MA) and stored on a hard disk.

Analysis of the performance data

Note onsets and offsets were extracted from the MIDI data of the 24 performances and matched to the score using an algorithm developed by Gingras and McAdams (2011), which has been shown to be suitable for ornamented harpsichord pieces. The mean error rate per performance, defined as the proportion of wrong notes or missing notes relative to the total number of score notes, was 0.54% (range: 0-2.14%). This low error rate is comparable to the rates reported by Repp (1996) and Goebl (2001) in studies on professional piano performance, suggesting that the performance data collected for the current study were of suitable quality for assessing performance timing in professional harpsichord performance.

Tempo patterns were determined through inter-onset intervals (IOI). IOIs were measured as the time interval between the onsets of successive notes, as obtained from the MIDI data for each performance of the piece. Each of the 140 note events in the *Prélude* was used to generate a

series of IOIs. The IOI for the last note of the piece was taken to be the onset-offset time interval. Although this interval was on average more than four times as long as the IOI of any other event in the piece, it was found by several additional analyses that excluding the last note from the data did not change optimized models. There were very few note events for which no timing data were available (0.60% of the total number of events): 18 wrong or missing notes, plus 2 complex ornamentations for which a precise onset could not be computed. IOIs for these notes were interpolated by using the average of the data for all the other performers, each note onset in the MIDI data having been previously unambiguously matched to a score event in the piece.

The IOI time series thus obtained for both performances by each harpsichordist yielded 12 within-performer average IOI series that were used in all subsequent analyses in Experiment 1. These within-performer IOI series were then themselves averaged across performers for some analyses. Inter-performer variability was assessed event by event to create a time series of the *coefficient of variation* (CV) of the IOIs (the ratio of the standard deviation to the mean for each event considering all 12 within-performer IOI series).

Information-dynamic analysis

The computational model described above (Pearce, 2005) was used to generate an information-dynamic analysis of the *Prélude*, including measures of the information content (unexpectedness) and entropy (uncertainty) for each note in context, given the model. We use two separate models with distinct representations. The first uses a *default* feature consisting of a single linked viewpoint combining pitch interval and scale degree (i.e., each note is represented as a pair of values – the first reflects the pitch interval in semitones from the previous note, the second reflects the chromatic scale degree from 0, the tonic, to 11, the leading tone). Because the

piece is nonmodulatory, the mapping between chromatic pitch and scale degrees was treated as consistent throughout the piece. This viewpoint allows the model to learn about dependent melodic and tonal relations in the sequential structure of the music. Previous research has found support for the influence of these two factors on pitch expectations in tonal melody (Hansen & Pearce, 2014; Pearce et al., 2010; Pearce & Wiggins, 2006; Schellenberg, 1997). The second approach uses an *optimized* set of viewpoints selected using the hill-climbing procedure described above (see also Appendix).

Time-series analysis

A detailed presentation of time-series analysis (TSA) techniques in the study of continuous musical events is available elsewhere (Bailes & Dean, 2012; Dean & Bailes, 2010). This large and long-standing suite of techniques is appropriate for use with data that are serially correlated (thus not independent measures), unlike most common statistical approaches that require independence of the multiple data measures to be studied. TSA is thus a specialized form of regression analysis that can take proper account of autocorrelations, which standard techniques fail to do. Autoregressive (AR) TSA can model the influence of external predictor variables (X), such as the information-dynamic parameters just mentioned, or acoustic properties such as intensity. Data to be modeled need to be statistically stationary, lacking trends. In essence stationarity means that as the data series fluctuates, its mean and variance remain constant, and the autocorrelation between measures made at any fixed time separation remains constant (as revealed in the so-called autocorrelation function). This is most commonly achieved by differencing the time series, that is, constructing a new series comprised of the differences between successive members of the original (which is thereby one value shorter). For a series

called *seriesname,* we refer to this differenced series as *dseriesname*, and of course it can be used to reconstruct the original series, so that models of *dseriesname* also provide predictions for *seriesname* and thus are informative about its predictors. Model selection is based on parsimony and information criteria. Here, we used the Bayesian Information Criterion (BIC; Schwarz, 1978). A lower BIC value indicates a better model for a given data series (BIC values cannot be meaningfully compared between models of different data series). The absolute value of the BIC is not informative. The quality of a model is assessed by a range of criteria, including the nature of the residual series and the degree of fit of the model to the data. Residuals, corresponding to the variance associated with each data point that is not explained by the model, should themselves lack autocorrelation and be normally distributed around 0 (this is also known as a Gaussian white noise distribution). Otherwise they still contain significant information that should be modeled. The degree of fit is assessed using the log-likelihood of the model.

We refer to our standard TSA analyses as ARMAX: AutoRegressive Moving Average models with eXternal predictors. We use moving average (MA) components in some cases, but they can be replaced by autoregressive (AR) terms if necessary by a re-parameterization, and so we do not discuss them in detail. TSA can also assess factors that may have mutual influence, such as perception of loudness and pitch, and consider their possible influences on each other at the same time as their influences on another measured variable. This is done by Vector Autoregression (VAR), in which multiple variables may be constrained either as exogenous (similar to a psychological independent variable) or endogenous (similar to a dependent variable). Note that all variables were initially included in the VAR analyses; the order of addition and removal of variables has no influence on the final models. This permits a direct comparative assessment of the predictive utility of the selected variables. Additional specific details are provided below as

appropriate, and the reader is referred to standard textbooks on time-series analysis, as well as the summary guidance in Dean & Bailes (2010), Bailes & Dean (2012), and Dean, Bailes, & Dunsmuir (2014a,b). All time-series analyses presented here were conducted with the commercial software STATA (StataCorp LP, College Station, TX). All statistical analyses involved principled manual processes and decisions; there was no automation of model selection.

Results

Our goal in Experiment 1 is to examine a hypothesized relationship between variations in expressive timing in performance and changes in melodic predictability based on the information content (IC) and entropy estimated by the IDyOM model. First, we examine the patterns of IC and entropy; second, we describe how we measure variations in expressive timing; and finally, we use time-series analysis to examine the extent to which changes in IC and entropy account for changes in performance timing.

Information-theoretic structure of the piece

As described above, we consider the output of two IDyOM models, using *default* and *optimized* viewpoints. The IC and entropy values estimated by the models are shown in Figure 2. It can be seen that the IC is on average lower for the default compared to the optimized viewpoint (although the two are highly correlated). This is true both in absolute terms and in relation to the entropy, indicating that the optimized set of viewpoints allows the model to predict the pitch of the individual notes in the *Prélude* more accurately than the default viewpoint.

[Insert Figure 2]

Overview of the performance timing data

The average duration of a performance (from first to last onset) was 59.9 seconds for the *Prélude* (range: 39.1-84.2 s). These durations corresponded to a mean tempo of 149.2 events/minute (range: 99.8-214.7). Figure 3 shows the mean IOI for each note event, with the shaded area corresponding to the standard deviation. In contrast to previous studies on measured organ music, which reported high correlations for timing patterns within and between performers (Gingras, 2008), we found a moderate mean correlation of 0.54 (SD = 0.18, *df* = 138) between pairs of performances of the *Prélude* played by different performers. Note that even though these correlations may be influenced by autocorrelation (the tempo is likely to be more similar for nearby note events than for temporally distant ones), the moderate values observed for the *Prélude* imply a fair amount of dissimilarity among the performances. Interestingly, the mean correlation between performances by the same performer reached 0.87 (SD = 0.10, *df* = 138). Overall, these analyses suggest that although both recordings from the same performer generally displayed similar timing patterns, there was a fair amount of variation among performers (see Gingras et al., 2013, for a detailed investigation of the individual variation between performers).

[Insert Figure 3]

Prediction of performance timing and inter-performer timing variability

The IOI time series was not statistically stationary (which, as mentioned, is required for reliable time series analysis). However, its first difference (*dIOI*) was stationary, provided that the IOI of the final note of the piece was excluded; in spite of the exceptional nature of the timing of the final note, its inclusion or exclusion in the time series models did not change the specifications of the optimal form. Thus initial modeling of the IOI used the first differenced versions of the relevant data streams, the dIOI and those from the Information Content(IC)/Entropy analysis in IDyOM. IOI itself showed notable partial autocorrelations at lags 1, 6, and 7 (corresponding to an autocorrelation between the current note duration and those of notes located 1, 6, and 7 events earlier, respectively). Further analyses showed that the periodicities at 6 and 7 events were related to the arpeggio figurations found at the beginning and end of the piece. The opening arpeggio pattern contains a 6+6 subdivision (see Figure 1), whereas the final arpeggio pattern is characterized by a 6+7 subdivision, and these subdivisions were generally clearly delineated by the performers. Moreover, the significant partial autocorrelations at lags 6 and 7 disappeared when these arpeggio patterns were excluded from the analysis. Graphs of the IOI and IC series showed that long IOIs often anticipate peaks in IC by about one event, whereas they seem to coincide with troughs in entropy, consistent with our prediction that (statistically) unexpected events, even though they are already known to performers, are associated with slowing.

Modeling of dIOI indicated that leads (F in Table 1), corresponding to subsequent events of the IDyOM data streams, were more influential than lags (earlier events). For both default and optimized models, dEntropy (the first difference of the Entropy time series) was not influential, but lead 1 of dIC (the first difference of the Information Content time series) was. Entropy components were considered as predictors in all cases, but were dismissed during the model

optimization process on the basis of BIC. In other words, whereas adding a predictor to a model nearly always enhances its fit, this improvement may not be sufficient to justify the inclusion of an additional predictor to the model (as judged here by the BIC), and hence it is excluded.

[Insert Table 1]

Once a good model of a time series is obtained, it shares the autocorrelation features of the data series. Thus measurement of correlation between model and data now becomes informative, although the value should not be viewed as identical in nature to a correlation between sets of independent variables. The correlations were similar for the default and optimized models suggesting little advantage of the optimized set of features. Figure 4 compares the dIOI to its prediction from the information content of the optimized model, showing the good fit achieved.

[Insert Figure 4]

While the dEntropy was not a significant predictor of dIOI, we hypothesized that it might be a stronger predictor of inter-performer variability, and this idea was supported in the best model we obtained. Thus we determined a time series for the coefficient of variation of the IOIs across performers (abbreviated *CV-IOI*) and developed a model of it. CV-IOI was stationary and autoregressive, and hence could be modeled without differencing. By a similar analytical process to that just outlined, we found that in our best model CV-IOI was predicted jointly by lead 1 of IC for the optimized model and leads 1 and 2 of its entropy, with autoregressive terms. This

model (also shown in Table 1) provided satisfactory residuals, and the modeled-observed correlation was 0.53 (see Figure 5 for the comparison of modeled and observed time series). Once again, the model was substantially better than the best purely autoregressive model, suggesting that our predictors do indeed capture variation in the data.

[Insert Figure 5]

Discussion

As expected, changes in information content predict performance timing. Specifically, performers appeared to slow down in anticipation of unexpected notes (notes characterized by a high IC). No differences appeared between the default and optimized versions of the IDyOM model. Entropy was not a significant predictor.

Turning to inter-performer variability, both IC and entropy were predictors. Therefore, performance timing appears to be more variable among performers around points of high uncertainty and unexpectedness in the musical structure. Modeling of two individual performances that represented extremes of duration (not shown) revealed that they could be quite well fit by the standard models shown in Table 1, supporting our interpretations.

An alternative explanation for the link we observed between variations in expressive timing in performance and entropy would be that performers are constrained to slow down at more challenging points in the piece (due to technical difficulties) and that these challenging points likely correspond to points of greater uncertainty in the piece. However, this explanation seems implausible considering that all performers were professional harpsichordists who received the

score of the piece several weeks prior to the recording session, and that the error rates were very low. Furthermore, previous research has shown that, among expert populations such as professional performers, trade-offs between timing and accuracy only become apparent when individuals perform at the limit of their abilities and under strict temporal constraints (Pfordresher, Palmer, & Jungers, 2007). Our subjective impression from listening to the excerpts is that the performers were mostly or fully in control of the timing strategies they employed, so it is very unlikely that the performers would be constrained to slow down at certain points due to technical difficulties (besides, the piece itself would not be considered to be particularly challenging for professional harpsichordists, as author R.T. Dean, himself a professional pianist, can confirm).

In Experiment 2, we examine the impact of both structural predictability and performance timing on the experience of musical tension in harpsichordists, other musicians, as well as nonmusicians listening to the performances analyzed in Experiment 1. To do so, we propose to use IC and entropy as model-based predictors of IOI and musical tension (using our information-dynamic model of melodic structure), and IOI as an expressive, performance-derived predictor of musical tension.

EXPERIMENT 2: PERCEIVED TENSION

Methods

Participants and stimuli

Ten harpsichordists, 20 nonharpsichordist musicians, and 20 nonmusicians were invited to listen to recordings of the performances collected in Experiment 1. The harpsichordist listeners,

eight male and two female, had an average age of 31.6 years (range: 23-38). None of these harpsichordists had participated in Experiment 1. The musician participants, 12 male and eight female, had a minimum of two years in undergraduate music studies and had an average age of 26.7 years (range: 20-47). The nonmusician participants, 12 male and eight female, had less than two years of musical training in early childhood and had an average age of 26.6 years (range: 20-48). Listeners received monetary compensation for their participation in the study.

Only one recording from each performer who participated in Experiment 1 was used, yielding a set of 12 stimuli, each from a different performer. Performances were chosen subjectively as the best one (out of two for each performer) in terms of recording quality and number of performance errors.

Procedure

All nonharpsichordist musicians and nonmusicians, as well as five harpsichordists were tested at McGill University, Montreal, Canada. Because of the difficulty of recruiting a sufficient number of Montreal-based harpsichordists who had not participated in Experiment 1, an additional five harpsichordists were tested at Goldsmiths, London, United Kingdom. Although some minor technical details differed between the two experimental setups, the experimental procedure was identical.

In both cities, the experiment took place in a double-walled IAC sound-isolation chamber; the sound booth models used were model 120-act3 in Montreal and model 1200-A in London (IAC Acoustics, Bronx, NY). In Montreal, participants performed the experiment on a Macintosh G5 computer running OS 10.6.8 (Apple Computer, Cupertino, CA). Sounds were amplified through

a Grace Design m904 monitor (Grace Digital Audio, San Diego, CA) and presented over Sennheiser HD280 Pro headphones (Sennheiser Electronic GmbH, Wedemark, Germany). In London, participants performed the experiment on a MacBook Pro 15" running OS 10.6.3 (Apple Computer, Cupertino, CA). Sounds were played through an Edirol UA-4FX USB external sound card (Roland Corporation, Shizuoka, Japan) and presented over Sennheiser HD202 headphones (Sennheiser Electronic GmbH, Wedemark, Germany).

All participants passed an audiogram to ensure that they had normal hearing. The musical excerpts were presented at a comfortable listening level that was kept constant for all participants. Participants used a slider to continuously rate the perceived musical tension of each performance. The slider was equipped with an elastic rubber band to provide feedback through physical resistance as it was moved toward higher tension ratings. The slider was connected to an AtoMIC Pro analog-to-MIDI converter (Ircam-Centre Pompidou, Paris). The computer interface, programmed in the PsiExp environment (Smith, 1995), provided instructions on the screen and allowed the participant to advance through the trials by clicking on an on-screen button with the mouse. Participants were instructed to hold the slider box on the desk or on their lap and to move the slider forward (away from themselves) to indicate higher perceived tension and backward (toward themselves) to indicate lower perceived tension. They were asked to use the full range of the slider for at least one performance during the whole experiment.

The experiment was divided into two parts. During the first part (two practice trials), the experimenter was present to answer questions. The practice performances were taken from the set of 12 performances that were not used in Experiment 2 (see section "Participants and stimuli"). The participants were instructed that the two practice trials were representative of the range of performances in the main part of the experiment. The data from the practice trials were

not analyzed. The definition of musical tension was left open due to the wide range of participants involved with the experiment. The main experimental block consisted of one performance by each performer, interspersed with 15 s of nature sounds as a short break. The order of performances was randomized for each participant. Once the experiment was completed, participants filled out a questionnaire. The entire experiment lasted approximately 30 minutes.

Data analysis

For each participant, a log file recorded the tension slider ratings continuously over time for each performance. Due to the force exerted by the elastic band when stretched at maximum, a block was inserted into the slider, resulting in MIDI ratings ranging between 54 and 127, which were then range normalized by participant. We used the matched score of the MIDI data of the performances to establish a correspondence between MIDI note onset events and score. The tension values were averaged over each score event.

Statistical models predicting perceived tension on the basis of the performance timing (IOI) and IDyOM parameters (IC and entropy) only made use of the IOI data from the 12 performances heard by participants in Experiment 2 (i.e., within-performer average IOI time series were not used here, in contrast to Experiment 1). Here, we also used VAR to provide the Granger Causality test (Granger, 1969, 2004), which yields a principled analysis of the directionality of the relationships between factors that might be mutually influential in a pair or group of time series (taking account of the autocorrelations). In other words, Granger causality analysis shows which time series (for instance IOI and IC), provide information that can predict future values of other time series, which is the main purpose of our analyses. Granger causality analyses were conducted to determine whether IOI and/or IDyOM parameter time series were

significant predictors of future values of the perceived tension profiles. Other aspects of the time series analysis follow the procedure described in Experiment 1.

Results

As shown in Figure 6, the profiles of perceived tension for the musician and nonmusician groups were somewhat different, whereas the harpsichordists were closely similar to the nonharpsichordist musicians. Thus we first modeled the possible impact of the IOI and IDyOM parameters on the perceived tension profiles of each participant group independently.

[Insert Figure 6]

The tension profiles showed strong autocorrelation. The first differenced profile was stationary in the case of musicians and harpsichordists and could be modeled satisfactorily. VAR analyses showed that IDyOM parameters had predictive influence, and that they provided a fair model of the tension profile. However, when they were taken together with the IOI, the results showed that only the IOI was required for the optimal model of the tension profile, which was significantly improved. In other words, IC or entropy were not *required* predictors. Moreover, unlike IC or entropy, IOI was a Granger causal predictor of the tension profile.

Because the VAR analysis, and in particular Granger causality, provides a direct comparison of the IOI and information-theoretic series as predictors of perceived tension, it is worth illustrating this further, taking as example the case of the musicians as auditors. We optimized a VAR model (of the first differenced variables) in which all four time series (Tension, IC, entropy

and IOI) are considered as endogenous. That is, in the model they may each influence each other from a statistical perspective, and a joint model of all four variables is constructed. This is the most conservative way of evaluating the model, because information content and entropy are computed directly from the score of the *Prélude* and are thus fixed (exogenous), and cannot be influenced by the other components. We obtain a reasonable model of dTension, with $R^2 = .40$ for the fit. The Granger analysis shows that the only Granger causal predictor of Tension is dIOI ($p < .01$); the p-values for dIC and dEntropy are not significant (both p-values $> .31$). Furthermore, the coefficients for the dIC and dEntropy series are also not significant at any lag included in the model (all p-values $>.29$). Thus, dIOI is clearly the only useful predictor. Very similar results obtain for the harpsichordists as auditors, given the slightly different overall model parameters (see below). On the basis of these and other VAR analyses, we then undertook ARMAX models comparable to those presented in Experiment 1.

Table 2 summarizes the best ARMAX time-series modeling results for perceived tension, giving white noise residuals. Figure 7 shows the average differenced profile for tension as perceived by nonharpsichordist musicians and its prediction on the basis of the best ARMAX model. Although harpsichordists showed similar perceived tension profiles to nonharpsichordist musicians, the best model for their perception involved the lead 1 ("F") of dIOI rather than its lag 1 (Table 2). This lead dIOI parameter could not be substituted in models by the IC features that predict dIOI (as shown in Table 1).

[Insert Table 2]

[Insert Figure 7]

In the case of the nonmusicians' tension profile, two differencing steps were required to reach stationarity. As with harpsichordists and nonharpsichordist musicians, only IOI time series showed Granger causality for the tension profiles of nonmusicians. Correspondingly, only IOI series had a significant influence, besides autoregression, in ARMAX analysis of stationarized tension profiles of nonmusicians (Figure 8).

[Insert Figure 8]

Discussion

Our analysis of the perceived tension data indicates that it is expressive performance timing (analyzed here in terms of IOI time series), rather than information-dynamic parameters such as IC and entropy, that directly influences perceived tension for all groups of listeners. Indeed, neither IC nor entropy were Granger-causal nor required predictors in statistical models predicting perceived tension, whereas IOI (more specifically, its differenced values) entered all models. Notably, there is no significant additional influence of IC on the tension values once the effect of expressive timing is taken into account, and only expressive timing was a Granger-causal predictor of perceived tension. However, we have shown in Experiment 1 that IC and entropy predict performance timing itself. Similarly, it was confirmed in the analyses of Experiment 2 that IC and entropy can predict perceived tension; although they were not required predictors, and did not account for any variance above and beyond what is already accounted for

by the IOI values (which they predict as shown in Experiment 1). Thus we may most readily interpret Experiment 2 as showing that the information parameters (extracted from the score as conceived by the composer) influence expressive performance timing (the realization by performers), which in turn influences perceived tension (the reception by listeners), embodying the tripartite communication model of Kendall and Carterette presented in the Introduction.

Table 2 also reveals an interesting difference between the harpsichordists' tension ratings and those of the nonmusicians and musicians. Indeed, unlike nonmusicians and nonharpsichordist musicians, harpsichordists' ratings are predicted only by imminent events (reflected in the F or 'lead' term in the TSA) rather than by immediately preceding events (represented by the L or 'lag' term in Table 2). This suggests that harpsichordists' tension ratings do in fact anticipate upcoming events, perhaps due to their greater familiarity with the musical style.

GENERAL DISCUSSION

To our knowledge, this is the first large-scale study that combines an analysis of expressive strategies in performance, melodic expectations (as predicted by a computational model), and listeners' ratings of musical tension in an attempt to characterize the relationships among these elements of musical communication. Results show that, in a semi-improvisatory musical genre such as the unmeasured prelude, the composed melodic structure, as measured by information-dynamic parameters, has the capacity to predict local tempo variations in performance. These in turn predict the perception of tension by listeners, including harpsichordists, nonharpsichordist musicians, and nonmusicians, in broadly similar ways across groups. More generally, our findings indicate that, when the constraints on expressive timing in performance introduced by notated meter and note duration are removed, the link between melodic structure and expressive

timing becomes statistically salient, thus bringing to the fore the close interconnection between the dimensions of pitch and time in music.

Our results thus substantiate the proposed interactions in the tripartite relationship suggested by Kendall and Carterette (1990): the cognitive processes involved in the communication between composer (via the notated score) and performer, and also the interactive communicative effect of the melodic expectations implied by the score, combined with the expressive timing in performance, on the listener. Furthermore our findings illustrate that these relationships act in a cumulative manner, in which one chain of influence (composer → performer) indirectly affects another (performer → listener).

Our choice of a piece containing a single unaccompanied melodic line, played on an instrument with a narrow dynamic range such as the harpsichord, allows us to avoid the complexities related to polyphony and harmony and variations in sound intensity, without compromising ecological validity. To be sure, our results do not necessarily indicate that melodic structure is the main predictor of musical tension. However, because sound intensity and roughness were presumably very similar for all performances (besides the limited dynamic range of harpsichord music, all recordings were made on the same instrument and under the same acoustical conditions), the experiments presented here allow us to understand more precisely the influence of pitch predictability and performance timing on perceived tension. In future work, it will also be interesting to assess whether sound intensity profiles, even in harpsichord performances, have a bearing on the perception of affect, and whether they themselves are sufficiently varied across the piece to reflect any influence of the flux of information content. Regardless of whether variations in intensity in harpsichord performances are perceived by listeners, the fact that harpsichordists appear to display individual velocity profiles (Gingras et

al., 2013) and that they tend to play an emphasized voice with higher velocity (Gingras et al., 2009) suggests that velocity profiles do play a role, however marginal, in expressive harpsichord performances.

The piece selected for this study exhibits a relatively simple high-level structure, and we predict that influences of this structural layout can also be detected in the performance timings[3]. Furthermore, we surmise that such structural influences generalize in more complex pieces to other aspects of performance and perception, as we have observed in pieces from a range of musical styles (Bailes & Dean, 2012). Once we are able to deal with structures of sufficient complexity, it may be possible to enhance information-theoretic properties of the event structures by larger-scale musicological ones (such as the 'development' or 'recapitulation' structures of classical music sonatas).

Previous research using the IDyOM model has demonstrated that information content accurately accounts for pitch expectations generating while listening to melodies (Pearce, 2005; Pearce et al., 2010; Omigie et al., 2012), whereas entropy predicts listeners' uncertainty about the note that will follow in a given context (Hansen & Pearce, 2014). These previous results suggest that perceptual processing of pitch sequences can be accurately modeled as a cognitive process of probabilistic prediction based on implicit statistical learning through exposure to music. The present results extend this picture for the first time to expressive performance and to emotional experience during listening. Specifically, the results confirm that information content (or unexpectedness) is a good predictor of performance timing and indicate that the entropy (or

---

[3] A companion article compares performers' structural analyses of the *Prélude* with their performance timings and the resulting perceived tension profiles, both on a local level and considering the large-scale structure of the piece (Goodchild, Gingras, & McAdams, 2015).

uncertainty) may also be influential. The results also exhibit a relationship between information content and tension experienced by listeners, although there was no additional variance in the tension ratings accounted for by information-theoretic variables after accounting for performance timing (which was itself predicted by information content and entropy). Current work is extending IDyOM to the analysis of other aspects of music, such as harmony (multiple pitches notated with simultaneous timing in the score, absent in this piece) and meter (entirely lacking from the current piece), and their interactions. This approach will then be used to understand the influence of information-theoretic properties of more complex stimuli including metrical and harmonic structure (in addition to melodic structure, examined here) on performance and perception.

It is interesting to place this piece, which is an example of a genre which codifies prior improvisations, within the current activities that develop from it: free improvisation. This immediately points to the fact that our study could be extended by applying a similar methodology to free musical improvisations, in which not only the meter and duration are not given (as with the piece used in this study), but the melodic structure is elaborated by the performer. A series of related studies on improvisation suggest that this approach could indeed be fruitful (Dean & Bailes, 2014; Dean, Bailes, & Drummond, 2014). Thus we surmise that the melodic structure of an improvisation should also predict timing in performance and perceived tension, especially in the case of music without regular meter such as free jazz and free improvisation, although this is commonly done in circumstances permitting wide ranges of acoustic intensity, which are expected also to be influential.

In conditions of experimentally controlled event timings, it may be theoretically possible to disentangle the relative contribution of the pitch structure and expressive timing to perceived

tension by exposing listeners to artificial isochronous performances (with all the notes having the same duration) and comparing the results to those obtained with actual performances of the same pieces. However, a stimulus with equal-duration pitches (isochronous rhythm) would cause listeners to perceive a regular accent pattern (Povel, 1984), which is not compatible with the specific constraints of this essentially unmeasured musical genre. In order to isolate the contribution of the pitch structure, one might also envision using artificial performances in which all notes have the same pitch height. However, Lerdahl and Krumhansl (2007) have argued that musical expectations (closely related to musical tension) are essentially undefined for repeated pitches, suggesting that IC and entropy may not be appropriate measures in the case of a stimulus with repeated identical pitches. In spite of the above-mentioned drawbacks associated with isochronous performances or performances with monotonic pitches, studies using such stimuli may yield useful insights that would help us determine whether our findings can be generalized to other musical genres. Moreover, a thorough experimental investigation of the impact of systematically controlled temporal deviations on perceived tension would be necessary to proceed beyond our demonstration of Granger causality (a strong directional statistical statement) to a claim of reductive strong causality. Such experimental manipulations are beyond the scope of the present study but should be followed up in future research.

This work has clear implications for music composition, performance and listening, the three poles of the model proposed by Kendall and Carterette. For composition, it confirms that systematic organization of pitch (and presumably other structures) measurably influences the performers' interpretive choices as well as the listeners' emotional responses, and can be used to shape these responses with a degree of predictability. For performance, it suggests that implicit or explicit awareness of the musical structure—of both the broader stylistic conventions to which

a piece relates and of the piece itself—is likely to have a significant impact on the choice of expressive strategies and on the resulting expressive impact on the audience. Regarding the aesthetic appreciation of music by listeners, it again suggests that implicit knowledge of such statistical/structural features influences the tension experienced during musical listening and that familiarity with a musical idiom can transform that influence.

We have focused on music (in fact, a specific piece of music with particular properties) as a domain in which the relationships between performance and perception can be investigated and related to the structure of the stimulus in a controlled way, without compromising ecological validity. However, the results have implications for human perception and performance beyond music. The role of prediction has been established in several areas of psychology such as language processing (Cristia et al., 2011; DeLong et al., 2005; Hale, 2006; Levy, 2008; Saffran, 2003), visual perception (Bar, 2007; Bubic et al., 2010; Egner et al., 2010) and motor sequencing (Wolpert & Flanagan, 2001). Furthermore, probabilistic approaches have been proposed to model the cognitive processes involved in acquiring and generating predictions in these domains (e.g., Friston, 2010; Levy, 2008; Perruchet & Pacton, 2006). Taking language as an example, the present findings are consistent with existing results showing that the information-theoretic structure of a sentence predicts aspects of performance such as word reading times (see e.g., Levy 2008), and suggest that these, in turn, may predict emotional responses, at least for certain kinds of linguistic production such as poetry. The same may be true of visual domains involving structured sequences of human movement, such as dance or silent film. These hypotheses should be tested in future research. In doing so, the framework we have used here provides a rigorous methodology for identifying quantitative dynamic relationships between perception, performance and the structure of the stimulus.

ACKNOWLEDGMENTS

REFERENCES

Bailes, F., & Dean, R. T. (2012). Comparative time series analysis of perceptual responses to electroacoustic music. *Music Perception*, *29*, 359–375.

Bar, M. (2007). The proactive brain: using analogies and associations to generate predictions. *Trends in Cognitive Science,* 11, 280–289. doi: 10.1016/j.tics.2007.05.005

Bharucha, J. J., & Stoeckig, K. (1986). Reaction time and musical expectancy: Priming of chords. *Journal of Experimental Psychology: Human Perception and Performance*, *12*, 403–410.

Bharucha, J. J., & Stoeckig, K. (1987). Priming of chords: Spreading activation or overlapping frequency spectra? *Perception & Psychophysics*, *41*, 519–524.

Bubic, A., von Cramon, D. Y., and Schubotz, R. I. (2010). Prediction, cognition and the brain. *Frontiers in Human Neuroscience,* 4, 25. doi: 10.3389/fnhum.2010.00025

Carlsen, J. C. (1981). Some factors which influence melodic expectancy. *Psychomusicology: Music, Mind & Brain*, *1*, 12–29.

Castellano, M. A., Bharucha, J. J., & Krumhansl, C. L. (1984). Tonal hierarchies in the music of North India. *Journal of Experimental Psychology: General*, *113*, 394–412.

Clarke, E. F. (1985). Structure and expression in rhythmic performance. In P. Howell, I. Cross, & R. West (Eds.), *Musical structure and cognition* (pp. 209–236). London: Academic Press.

Clarke, E. F. (1988). Generative principles in music performance. In J. Sloboda (Ed.), *Generative processes in music: The psychology of performance, improvisation and composition* (pp. 1–26). Oxford: Oxford University Press.

Clarke, E. F. (1989). The perception of expressive timing in music. *Psychological Research-Psychologische Forschung*, *51*, 2–9.

Conklin, D., & Witten, I. H. (1995). Multiple viewpoint systems for music prediction. *Journal of New Music Research*, *24*, 51–73.

Cooper, W. E., & Danly, M. (1981). Segmental and temporal aspects of utterance-final lengthening. *Phonetica*, *38*, 106-115.

Cristià, A., McGuire, G. L., Seidl, A., and Francis, A. L. (2011). Effects of the distribution of acoustic cues on infants' perception of sibilants. *Journal of Phonetics,* 39, 388–402. doi: 10.1016/j.wocn.2011.02.004

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. (2014). Cognitive processes in musical improvisation. In G. E. Lewis & B. Piekut (Eds.), *The Oxford Handbook of Critical Improvisation Studies*, vol. 1. Oxford: Oxford University Press. doi: http://dx.doi.org/10.1093/oxfordhb/9780195370935.013.007

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. doi: 10.1080/09298215.2013.859710

Dean, R. T., Bailes, F., & Dunsmuir, W. T. M. (2014a). 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., & Dunsmuir, W. T. M. (2014b). Shared and distinct mechanisms of individual and expertise-group perception of expressed arousal in four works. *Journal of Mathematics and Music*, 8, 207-223.

DeLong, K. A., Urbach, T. P., and Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience,* 8, 1117–1121. doi: 10.1038/nn1504

Eerola, T. (2004). Data-driven influences on melodic expectancy: Continuations in North Sami Yoiks rated by South African traditional healers. In *Proceedings of the eighth*

international conference on music perception and cognition, Evanston, IL. Causal Productions, Adelaide (pp. 83–87).

Egermann, H., Pearce, M. T., Wiggins, G. A., & McAdams, S. (2013). Probabilistic models of expectation violation predict psychophysiological emotional responses to live concert music. *Cognitive, Affective, & Behavioral Neuroscience*, 1–21. doi:10.3758/s13415-013-0161-y

Egner, T., Monti, J. M., and Summerfield, C. (2010). Expectation and surprise determine neural population responses in the ventral visual stream. *Journal of Neuroscience*, 30, 16601–16608. doi: 10.1523/jneurosci.2770-10.2010

Farbood, M. M. (2012). A parametric, temporal model of musical tension. *Music Perception*, *29*, 387–428.

Fletcher, N. H. (1977). Analysis of the design and performance of harpsichords. *Acustica*, *37*, 139–147.

Fredrickson, W. E. (1995). A comparison of perceived musical tension and aesthetic response. *Psychology of Music*, *23*, 81–87.

Fredrickson, W. E. (2000). Perception of tension in music: Musicians versus nonmusicians. *Journal of Music Therapy*, *37*, 40–50.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience,* 11, 127–138. doi: 10.1038/nrn2787

Gabrielsson, A., & Juslin, P. N. (1996). Emotional expression in music performance: Between the performer's intention and the listener's experience. *Psychology of Music*, *24*(1), 68–91.

Gingras, B. (2008). *Expressive strategies and performer-listener communication in organ performance* (Unpublished doctoral dissertation). McGill University, Montreal (Canada).

Gingras, B., Asselin, P.-Y., & McAdams, S. (2013). Individuality in harpsichord performance: disentangling performer- and piece-specific influences on interpretive choices. *Frontiers in Psychology*, *4*. doi:10.3389/fpsyg.2013.00895

Gingras, B., & McAdams, S. (2011). Improved score-performance matching using both structural and temporal information from MIDI recordings. *Journal of New Music Research*, *40*, 43–57.

Goebl, W. (2001). Melody lead in piano performance: Expressive device or artifact? *Journal of the Acoustical Society of America*, *110*, 563–572.

Goodchild, M., Gingras, B., & McAdams, S. (2015). Analysis, performance, and perception of an unmeasured prelude for harpsichord. Manuscript submitted for publication.

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

Granger, C. W. J. (2004). Time series analysis, cointegration, and applications. *American Economic Review*, *94*, 421-425.

Hale, J. (2006). Uncertainty about the rest of the sentence. *Cognitive Science*, *30*, 643–672. doi: 10.1207/s15516709cog0000_64

Hanslick, E., & Cohen, G. (1854). Vom musikalisch-Schönen [On the Beautiful in Music]. *Darmstadt: Wissenschaftliche Buchgesellschaft*.

Hansen, N. C. & Pearce M. T. (2014). Predictive uncertainty in auditory sequence processing. *Frontiers in Psychology*, 5, 1052. doi: 10.3389/fpsyg.2014.01052

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

Ilie, G., & Thompson, W. F. (2006). A Comparison of Acoustic Cues in Music and Speech for Three Dimensions of Affect. *Music Perception*, *23*(4), 319–330. doi:10.1525/mp.2006.23.4.319

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

Juslin, P. N., & Timmers, R. (2010). Expression and communication of emotion in music performance.

Juslin, P. N., & Vastfjall, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences*, *31*(5), 559.

Kendall, R. A., & Carterette, E. C. (1990). The communication of musical expression. *Music Perception*, *8*, 129–164.

Kessler, E. J., Hansen, C., & Shepard, R. N. (1984). Tonal schemata in the perception of music in Bali and in the West. *Music Perception*, *2*, 131–165.

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., & Friederici, A. D. (2004). Music, language and meaning: brain signatures of semantic processing. *Nature Neuroscience*, *7*, 302–307.

Krumhansl, C. L. (1990). Tonal Hierarchies and Rare Intervals in Music Cognition. *Music Perception*, *7*, 309–324.

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

Krumhansl, C. L. (2002). Music: A link between cognition and emotion. *Current Directions in Psychological Science*, *11*, 45–50.

Krumhansl, C. L., Louhivuori, J., Toiviainen, P., Järvinen, T., & Eerola, T. (1999). Melodic expectation in Finnish spiritual folk hymns: Convergence of statistical, behavioral, and computational approaches. *Music Perception*, *17*, 151–195.

Krumhansl, C. L., Toivanen, P., Eerola, T., Toiviainen, P., Jarvinen, T., & Louhivuori, J. (2000). Cross-cultural music cognition: cognitive methodology applied to North Sami yolks. *Cognition*, *76*, 13–58.

Large, E. W., Palmer, C., & Pollack, J. B. (1995). Reduced Memory Representations for Music. *Cognitive Science*, *19*, 53–96.

Lerdahl, F. (2001). *Tonal pitch space*. Oxford : Oxford University Press.

Lerdahl, F., & Krumhansl, C. L. (2007). Modeling tonal tension. *Music Perception*, *24*, 329–366.

Levitin, D. J., & Tirovolas, A. K. (2009). Current advances in the cognitive neuroscience of music. *Annals of the New York Academy of Sciences*, *1156*, 211–231.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition,* 106, 1126-1177.

Madsen, C. K. (1997). Emotional response to music as measured by the two-dimensional CRDI. *Journal of Music Therapy*, *34*, 187–199.

Madsen, C. K., Brittin, R. V., & Capperella-Sheldon, D. A. (1993). An empirical method for measuring the aesthetic experience to music. *Journal of Research in Music Education*, *41*, 57–69.

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*, 46–46.

Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. Cambridge, MA: MIT Press.

McAdams, S., Vines, B., Vieillard, S., Smith, B. & 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.

Meyer, L. B. (1956). *Emotion and meaning in music*. Chicago: University of Chicago Press.

Narmour, E. (1990). *The analysis and cognition of basic melodic structures: The implication-realization model*. Chicago: University of Chicago Press.

Omigie, D., Pearce, M. T., & Stewart, L. (2012). Tracking of pitch probabilities in congenital amusia. *Neuropsychologia*, *50*, 1483-1493.

Omigie, D., Pearce, M. T., Williamson, V., & Stewart, L. (2013). Electrophysiological correlates of melodic processing in congenital amusia. *Neuropsychologia*, 51, 1749-1762.

Oram, N., & Cuddy, L. L. (1995). Responsiveness of Western Adults to Pitch-Distributional Information in Melodic Sequences. *Psychological Research-Psychologische Forschung*, *57*, 103–118.

Palmer, C. (1989). Mapping musical thought to musical performance. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 331–346.

Palmer, C. (1996). On the assignment of structure in music performance. *Music Perception*, *14*, 23–56.

Panksepp, J., & Bernatzky, G. (2002). Emotional sounds and the brain: the neuro-affective foundations of musical appreciation. *Behavioural Processes*, *60*, 133–155. doi:10.1016/S0376-6357(02)00080-3

Pearce, M. T. (2005). *The construction and evaluation of statistical models of melodic structure in music perception and composition* (Unpublished doctoral dissertation). City University, London.

Pearce, M. T., Conklin, D., & Wiggins, G. (2005). Methods for combining statistical models of music. In *Computer Music Modeling and Retrieval* (pp. 295–312). Springer.

Pearce, M. T., Ruiz, M. H., Kapasi, S., Wiggins, G. A., & Bhattacharya, J. (2010). Unsupervised statistical learning underpins computational, behavioural, and neural manifestations of musical expectation. *NeuroImage*, *50*, 302–313.

Pearce, M. T., & Wiggins, G. (2004). Improved methods for statistical modelling of monophonic music. *Journal of New Music Research*, *33*, 367–385.

Pearce, M. T., & Wiggins, G. A. (2006). Expectation in melody: The influence of context and learning. *Music Perception*, *23*, 377–405.

Pearce, M. T., & Wiggins, G. A. (2012). Auditory Expectation: The Information Dynamics of Music Perception and Cognition. *Topics in Cognitive Science*, *4*, 625–652. doi:10.1111/j.1756-8765.2012.01214.x

Penttinen, H. (2006). On the dynamics of the harpsichord and its synthesis. In V. Verfaille (Ed.), *Proc. 9th Int. Conf. on Digital Audio Effects (DAFx-06)* (pp. 115–120). Montreal, Canada.

Perruchet, P., and Pacton, S. (2006). Implicit learning and statistical learning: one phenomenon, two approaches. *Trends in Cognitive Science*, 10, 233–238. doi: 10.1016/j.tics.2006.03.006

Povel, D. P. (1984). A theoretical framework for rhythm perception. *Psychological Research*, *45*, 315-337.

Pfordresher, P. Q., Palmer, C., & Jungers, M. K. (2007). Speed, accuracy, and serial order in sequence production. *Cognitive Science, 31*, 63-98.

Repp, B. H. (1990). Patterns of expressive timing in performances of a Beethoven minuet by 19 famous pianists. *Journal of the Acoustical Society of America*, *88*, 622–641.

Repp, B. H. (1992). Diversity and commonality in music performance - an analysis of timing microstructure in Schumann's "Träumerei." *Journal of the Acoustical Society of America*, *92*, 2546–2568.

Repp, B. H. (1995). Expressive timing in Schumann's "Träumerei" - an analysis of performances by graduate student pianists. *Journal of the Acoustical Society of America*, *98*, 2413–2427.

Repp, B. H. (1996). The art of inaccuracy: why pianists' errors are difficult to hear. *Music Perception*, *14*, 161–183.

Saffran, J. R. (2003a). Statistical language learning. *Current Directions in Psychological Science*, 12, 110–114. doi: 10.1111/1467-8721.01243

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52.

Schellenberg, E. G. (1997). Simplifying the implication-realization model of melodic expectancy. *Music Perception*, *14*, 295–318.

Schellenberg, E. G., Adachi, M., Purdy, K. T., & McKinnon, M. C. (2002). Expectancy in melody: Tests of children and adults. *Journal of Experimental Psychology:General*, *131*, 511–536.

Shaffer, L. H., & Todd, N. P. (1987). The interpretive component in musical performance. In A. Gabrielsson (Ed.), *Action and perception in rhythm and music* (Vol. Royal Swedish Academy of Music No. 55, pp. 139–152).

Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics, 6,* 461-464

Shepard, R. N. (1982). Structural representations of musical pitch. In D. Deutsch (Ed.), *The Psychology of Music*, (pp. 343–390). Orlando, FL: Academic Press.

Sloboda, J. A. (2010). Music in everyday life: The role of emotions. In P. N. Juslin & J. Sloboda (Eds.), *Handbook of Music and Emotion: Theory, Research, Applications* (pp. 493–514). Oxford: Oxford University Press.

Smith, B. K. (1995). PsiExp: An environment for psychoacoustic experimentation using the IRCAM musical workstation. Presented at the Society for Music Perception and Cognition, University of California, Berkeley, USA.

Steinbeis, N., Koelsch, S., & Sloboda, J. A. (2006). The role of harmonic expectancy violations in musical emotions: Evidence from subjective, physiological, and neural responses. *Journal of Cognitive Neuroscience*, *18*, 1380–1393.

Tillmann, B., Bharucha, J. J., & Bigand, E. (2000). Implicit learning of tonality: A self-organizing approach. *Psychological Review*, *107*, 885–913.

Tillmann, B., Bigand, E., & Pineau, M. (1998). Effects of global and local contexts on harmonic expectancy. *Music Perception*, *16*, 99–117.

Tilney, C. (1991). *The art of the French unmeasured prelude* (Vol. 1–2). London: Schott.

Todd, N. (1985). A model of expressive timing in tonal music. *Music Perception*, *3*, 33–58.

Toiviainen, P., & Krumhansl, C. L. (2003). Measuring and modeling real-time responses to music: The dynamics of tonality induction. *Perception*, *32*, 741–766.

Wolpert, D. M., and Flanagan, J. R. (2001). Motor prediction. *Current Biology*, 11, R729–R732.

Table 1. Time-series analysis of performance timing (dIOI and CV of IOI) using information content and entropy (from two IDyOM models, using default and optimized viewpoints) as predictors.

Modeled parameter	Predictors considered (IDyOM model in parentheses)	Components of the optimal model (coefficients for the external predictors in parentheses)	BIC	Correl. (modeled vs. observed values)
dIOI	IC and entropy (default)	F1.dIC (+0.01), AR(1-5)	70.1	0.65
dIOI	IC and entropy (optimized)	F1.dIC (+0.02), AR(1-5)	71.8	0.64
CV of IOI	IC and entropy (optimized)	F1.IC (+0.02), F1-2.Entropy (+0.06, +0.05), AR(1-2), MA(25)	-150.6	0.52

The chosen data series were stationary, and all models listed showed satisfactory white noise residuals. AR(n): autoregressive error terms; Ln: lag n; Fn, lead n; MA(n): moving average error term lag n, where n represents the number of events. For example, 'F1.dIC (+0.02)' (fourth row) means: lead one of the (first) differenced Information Content data series (derived from the optimized IDyOM model in this case), had a coefficient of +0.02. See the Methods section in Experiment 1 for further details of the time-series analysis.

Table 2. Time-series analysis of tension ratings (dTension) using information content and entropy (from two IDyOM models, using default and optimized viewpoints) and performance timing (dIOI) as predictors.

Modeled parameter	Predictors considered (IDyOM model in parentheses)	Components of the optimal model (coefficients for the external predictors in parentheses)	BIC	Correl. (modeled vs. observed values)
dTension Musicians	IC and entropy (default), dIOI	L1.dIOI (+0.75), AR(1,5,6)	521.0	0.65
dTension Musicians	IC and entropy (optimized), dIOI	L1.dIOI (+0.75), AR(1,5,6)	521.0	0.65
d2Tension Nonmusicians	IC and entropy (optimized), dIOI	L1.d2IOI (+0.04), F1.d2IOI(-0.13), AR(1-5,8)	513.4	0.61
dTension Harpsichordists	IC and entropy (optimized), dIOI	F(1,2).dIOI(-0.65, +0.31), AR(1,2)	540.4	0.54

The chosen data series were stationary, and all models listed showed satisfactory white noise residuals. AR(n): autoregressive error terms; Ln: lag n; Fn: lead n, where n represents the number of events. For example, 'L1.dIOI (0.75) AR(1,5,6)' (first row) means: lead one of the (first) differenced inter-onset interval series, derived from the optimized model, had a coefficient of +0.75. See the Methods section in Experiment 1 for further details of the time-series analysis. Note that although a range of parameters are considered as possible predictors (column 2), a smaller selection remains in the optimal model (as shown in column 3). Thus, for example, the models of rows 1-2 are identical.

FIGURE CAPTIONS

Figure 1. Excerpt from the score of the *Prélude non mesuré No. 7*, by Louis Couperin (1626-1661). The excerpt corresponds to the first of five musical systems. The opening arpeggio pattern (with its 6+6 subdivision) is indicated by brackets. Used with permission from Tilney, C. (1991), *The art of the French unmeasured prelude* (Vol. 1–2), London: Schott.

Figure 2. Information content and entropy predicted by the default and optimized versions of the IDyOM model.

Figure 3. Mean duration (in seconds) of the inter-onset interval for each event of the *Prélude non mesuré No. 7*. The shaded area corresponds to one standard deviation below and above the mean for each event. Note the discontinuity in the vertical axis.

Figure 4. Observed and modeled differenced inter-onset intervals for each event of the *Prélude non mesuré No. 7*. Note the discontinuity in the vertical axis.

Figure 5. Observed and modeled coefficients of variation of inter-onset interval for each event of the *Prélude non mesuré No. 7*.

Figure 6. Standardized (*z*-scored) perceived tension for the harpsichordists, nonharpsichordist musicians, and nonmusicians, for each event of the *Prélude non mesuré No. 7*.

Figure 7. Observed and modeled differenced perceived tension for each event of the *Prélude non mesuré No. 7* for the nonharpsichordist musicians.

Figure 8. Observed and modeled second difference of perceived tension for each event of the *Prélude non mesuré No. 7* for the nonmusicians.

APPENDIX

In this research, we used two versions of IDyOM (Pearce, 2005), one using the *default* viewpoints and the other using the *optimized* viewpoints. The default model uses a single viewpoint linking pitch interval (the pitch difference in semitones between a note and its immediate predecessor; denoted below by *cpint*) with scale degree (the difference in semitones of a note from the tonic of the key; denoted below by *cpintfref*). The optimized version uses a set of viewpoints that has been optimized through hill-climbing to maximize the predictability of the musical piece. See the main text for a detailed description of IDyOM.

Table A1 gives the full set of viewpoints submitted to the optimization procedure, which also considered all possible pairwise links between these individual viewpoints. The viewpoints selected for the optimized model are shown in Table A2.

Table A1: The viewpoints submitted to the selection procedure for the optimized model.

Viewpoint	Description
cpitch	chromatic pitch (MIDI pitch number)
cpitch-class	octave equivalent pitch class (chroma)
tessitura	three values: whether a note is between 66 (G#4) and 74 (D5), above, or below this range
cpint	pitch interval in semitones
cpint-size	absolute size of pitch interval
cpcpint	pitch interval class (cpint mod 12)
cpcint-size	absolute size of cpcint
contour	pitch contour (-1, 0, 1)
newcontour	Boolean: whether or not contour is the same as the previous contour
cpintfip	pitch interval from the first note in the piece
cpintfref	chromatic scale degree
inscale	Boolean: whether or not the note is in the scale

Table A2: The viewpoints selected for the optimized model.

Viewpoint added	Mean Information Content
cpint-size	4.849
tessitura × inscale	4.430
cpcint-size × cpintfref	4.296
cpitch × cpint-size	4.149
cpitch-class × cpintfref	4.080
cpitch × cpintfip	4.055
cpintfref × inscale	4.054
tessitura × cpcint	4.044
cpcint	4.037