A Parametric, Temporal Model of Musical Tension

Morwaread M. Farbood
New York University

Tension in music is a high-level concept that is difficult to formalize due to its complex, multidimensional nature. This paper proposes a quantitative model of musical tension that takes into account the dynamic, temporal aspects of listening. The model is based on data from two experiments. The first was a web-based study that was designed to examine how individual musical parameters contribute directly to a listener's overall perception of tension and how those parameters interact. The second study was an in-lab experiment in which listeners were asked to provide continuous responses to longer, more complex musical stimuli. Both studies took into account a number of musical parameters including harmony, pitch height, melodic expectation, dynamics, onset frequency, tempo, meter, rhythmic regularity, and syncopation. As an initial step, linear and nonlinear models were explored for predicting tension given analytical descriptions of various musical parameters. These models were tested on the continuous-response data from Experiment 2 and shown to be insufficient. An alternate model was proposed based on the notion of a moving perceptual window in time and the concept of trend salience. High correlation with empirical data indicates that this parametric, temporal model accurately predicts tension judgments for complex musical stimuli.

Received May 21, 2009, accepted August 10, 2011.

Key words: musical tension, temporal dynamics, change perception, trend salience, web study

The aggregate experience of listening to music requires the integration of many disparate auditory components. At any given moment in time, a particular musical feature can stand out, attracting and focusing the listener's attention. Musical aspects can also combine to create a powerful effect; a sudden change in harmony might be accompanied by a parallel change in instrumental timbre or dynamics. This flux inherent in music is what defines it as an immediate, temporal experience. It is also a key element in the perception of musical tension, a sensation that arises from the combined interaction of various musical parameters. The phenomenon of tension is evident to listeners and is relatively easy to define in informal, qualitative terms; for example, increasing tension can be described as a feeling of rising intensity or impending climax, while decreasing tension can be described as a feeling of relaxation or resolution. However, formalizing and quantifying such a description is a difficult problem.

Tension as a concept is an emergent phenomenon that previous studies have shown to be judged with considerable consistency by listeners. Fredrickson (1999) found that having extensive familiarity with the music—even to the extent of performing it—does not greatly affect listeners' perception of tension. Nielsen (1987) noted that tension is immanent in the music itself, as opposed to "real" tension (e.g., emotion) experienced by the listener, which may be due to outside causes or reasons particular to the person. Another indicator of the perceptual objectivity of tension is that it does not appear to be dependent on the interest of the listener; it can thus be assessed independently from a listener's aesthetic preferences (Lychner, 1998). This is also indicated empirically by the high within-subject and intersubject correlations for tension judgments in previous studies (Bigand, Parncutt, & Lerdahl, 1996; Krumhansl, 1996). Given the seemingly self-evident nature of tension and the wide range of musical parameters that might contribute to it, most previous studies have opted not to define tension in precise terms to participants (Fredrickson, 1997, 1999; Krumhansl, 1996, 1997; Madsen & Fredrickson, 1993). One exception is Bigand and Parncutt (1999), who defined tension quite specifically as the "feeling that there must be a continuation of the sequence." This particular description, given to their subjects, is likely the reason why their results differed from Lerdahl and Krumhansl's (2007) results; both studies utilized some of the same stimuli but Lerdahl and Krumhansl did not provide subjects with a specific definition of tension.

Music Perception volume 29, issue 4, pp. 387–428. issn 0730-7829, electronic issn 1533-8312. DOI: 10.1525/mp.2012.29.4.387

Aspects of music contributing to tension explored in prior work can be roughly placed into two general categories: the domain-general psychological, and the domain-specific musical. The domain-general category includes studies that explore both low-level perceptual aspects of auditory perception as well as higher-level features that can be linked to general cognitive functions. In the former category are psychoacoustic features such as loudness, timbre, and pitch register. Dynamics has been discussed frequently as a contributor to tension (Burnsed & Sochinski, 1998-2001; Granot & Eitan, 2011; Ilie & Thompson, 2006; Krumhansl, 1996; Misenhelter, 2001; Nielsen 1983, 1987). From a biological perspective, loudness is a potential warning sign of danger and is subject to reflexive response (Granot & Eitan, 2011; Huron, 2006). Unlike the case for loudness, relatively few studies have examined pitch register explicitly (Granot & Eitan 2011; Ilie & Thompson, 2006). Pitch register may also be classified under the category of timbre change. Other prior work focusing on timbral elements have looked at features such as roughness, brightness, and density (Helmholtz, 1877/1954; Hutchinson & Knopoff, 1978; Krumhansl, 1996; Nielsen 1987; Plomp & Levelt, 1965; Pressnitzer, McAdams, Winsberg, & Fineberg, 2000).

Included in the domain-general category of features contributing to tension are factors relating to Gestalt psychology. These principles are arguably not specific to musical style and can be applied to any type of auditory stimuli at different time-span levels ranging from event-level auditory-scene analysis (Bregman, 1990) to phrase-level musical parsing (Lerdahl & Jackendoff, 1983; Meyer, 1956; cf. Tenney & Polansky, 1980). In musical contexts, group boundaries are highly correlated with tension responses curves (Krumhansl, 1996; Nielsen 1987).

Perhaps the broadest domain-general features explored in the context of tension pertain to expectation and emotion (Huron, 2006; Krumhansl, 1997; Madsen, 1998; Margulis, 2005; Meyer, 1956). Expectation is a phenomenon "known to be a basic strategy of the human mind; it underlies the ability to bring past experience to bear on the future" (Margulis, 2005). The way these expectancies are utilized consciously or unconsciously by composers influences the way listeners perceive tension in music. Whether those aspects are deeply schematic expectations or situational expectations, both short- and long-term memories of past auditory experiences are the primary sources for tension judgments. Huron, among others, has proposed that the manner in which expectations are satisfied or denied triggers prediction response mechanisms in the brain. It is the process of evaluating and assessing the success of a prediction that then gives rise to emotions.

Although expectation and emotional response to music have been linked to tension, the precise nature of this relationship has not been determined. Lychner (1998) explored how subjects interpreted the terms "aesthetic response," "felt emotional response," and "musical tension." He found that while subjects did not differentiate between the terms "aesthetic response" and "felt emotional response," they did apply the term "musical tension" in a different way. Tension has often been discussed in the context of emotional response and expectation with little attempt to differentiate it as a separate phenomenon. It is possible that there is an inverse relationship between expectancy and tension; Margulis's model of melodic expectation (2005) describes three kinds of tension responses, one of which is surprise-tension, which correlates inversely with expectancy ratings—that is, highly predictable events generate little surprise-tension. Margulis also links surprise-tension and other types of expectancy-related tension to affective response. In comparing emotion and tension responses, Schubert and Dunsmuir (2004) found some evidence that the resting points in tension judgments corresponded to increasing happiness (valence). Another possible relationship between tension and emotion is that tension response is equivalent to affective arousal (Kruhmansl, 1997; Trolio, 1976). Rozin, Rozin, and Goldberg (2004) measured "affective intensity," which could be equated with tension, although the authors do not explicitly make the connection. Huron's model of expectation (2006) has an arousal-related tension component, although in his case it refers specifically to the physiological response in preparation for an imminent event.

Obtaining continuous judgments in 2D arousal-valence space is a commonly used methodology for collecting data on emotional response to music. Continuous arousal data from a 2D space might be a less precise, noisier measure of 1D tension judgments. Others have proposed that there is a need to differentiate between tension arousal and energy arousal (Ilie & Thompson, 2006; Schimmack & Reisenzein, 2002). However, this could be a function of how the task is specifically described—given that "tension" is under-defined in many experiments, it is likely that listeners instinctively use both percepts to define what tension is. More recently, Eerola and Vuoskoski (2011) explored a 3D space for emotion in music that consisted of valence, tension, and energy. Observing that tension and energy had a strong positive correlation, they concluded that a 2D space was probably sufficient to describe emotion.

One feature that has been little explored, perhaps due to its lack of direct connection with features inherent in the auditory stimulus itself, is semantic meaning. Hackworth and Fredrickson (2010) examined whether knowledge of a text translation in a choral work by Debussy had an effect on tension judgments. Their results indicated that there was no meaningful difference between responses of listeners who understood the text and listeners who did not. Given that a composer setting words to music naturally attempts to compose in a manner that best reflects the meaning of the words, perhaps the only way to determine whether semantic meaning actually has some effect on tension would be to compose a piece that did not have an intentional relationship to the text.

Domain-specific features explored in previous studies are less broad than those described above and primarily focus on the discrete-pitched and rhythmic elements of music, and by definition are exclusive to musical stimuli. Pitched aspects include melodic contour, harmony, and tonal perception. Technically speaking, any of these features can also be linked to expectation of schematic structures, so could in an indirect way fall under the domain-general category. Nonetheless, they are placed in this category because they are features that distinguish music from other auditory stimuli such as speech (cf. Patel, 2008). Pitch height is one such feature explored in tension studies (Bigand et al., 1996; Granot & Eitan, 2011; Krumhansl, 1996; Nielsen, 1983, 1987). Although there are speech analogs to melodic contour, even tonal languages do not have the fixed-interval nature of musical lines. Work that focuses on melodic "attraction" in the voice-leading or tonal sense (Bharucha, 1984; Larson, 2004, Larson & VanHandel, 2005; Margulis, 2005; Narmour, 1990, 1992) is also relevant to tension, although the percept itself is not necessarily discussed explicitly.

Other domain-specific aspects examined in past work pertain to harmonic tension (Bigand & Parncutt, 1999; Bigand et al., 1996; Krumhansl, 1996; Lerdahl & Kruhmhansl, 2007; Nielsen, 1983, 1987; Toivianen & Krumhansl, 2003); all of these studies, with the exception of Toivianen and Krumhansl's and Nielsen's work, explored the empirical validity of Lerdahl's tonal tension model (1996, 2001) in various ways. The general results indicate that the model accurately predicts harmonic tension and that its hierarchical component is an essential element in describing listener judgments of tonal tension. Given the empirically proven descriptive power of Lerdahl's model, it is employed to calculate quantitative values for harmonic tension in the work described in this paper.

Past work that has examined the effect of rhythm and timing on tension has shown that tempo changes affect tension judgments (Ilie & Thompson, 2006; Krumhansl, 1996) while *rubato* does not (Fredrickson & Johnson, 1996). The contributions of rhythm and meter to tension are mostly unexplored perhaps because rhythm is harder to conceive of as a continuous process. A change in loudness or pitch height can be readily defined; the concept of a change in harmonic tension is considerably more complex but still possible to model, as shown in Lerdahl's work. However, an increase in rhythmic tension does not readily lend itself to formalization. There has been some music-theoretic work done in this domain (Hasty, 1997; London, 2004), but it does not directly relate to empirical research on tension.

The musical parameters explored in relation to tension are not the only aspects of prior work that are diverse—experimental methodologies have varied as well. Data collection methods fall into two general categories: asking listeners to make continuous, online ratings or obtaining discrete, retrospective judgments. Previous studies have utilized both hardware and software continuous-input interfaces to record listener judgments of tension in real time. Nielsen (1983, 1987) used a technique he described as a "simenon method," short for "simultaneous" and "nonverbal." He had subjects squeeze a pair of tongs with spring resistance while listening to music. The tongs were instrumented with a potentiometer to measure the compression. In a follow-up study replicating Nielsen's work, Madsen and Fredrickson (1993) used a device called a Continuous Response Digital Interface (CRDI), which consisted of a potentiometer mounted in a dial interfaced with a computer. The CRDI was used in several subsequent experiments by Fredrickson and others and its reliability appears to be high (Capperella, 1989; Fredrickson, 1997, 1999; Fredrickson & Coggiola, 2003; Frego, 1999; Gregory, 1989, 1995; Hackworth & Fredrickson, 2010). Vines, Krumhansl, Wanderley, and Levitin (2006) used a continuously adjustable linear slider on a MIDI controller to collect continuous data for their experiments. Software applications with a slider component have also been implemented (Burnsed & Sochinski, 1998-2001; Krumhansl, 1996). Data collected with Burnsed and Sochinski's interface, called the Tensiometer, appear to correlate highly to those gathered using the CRDI and the original device designed by Nielsen (Fredrickson & Coggiola, 2003). Rozin et al. (2004) used a pressure-sensitive button attached to the arm of a reclining chair. They chose to use this type of interface instead of a dial or slider because they felt pressure-sensitivity was a better metaphor for affective intensity than left-right or up-down motions.

Despite the prevalence of continuous methods of data collection, there have been studies that have employed retrospective judgment measures as well. Bigand et al. (1996) had subjects give single tension ratings to pairs of chords. Bigand & Parncutt (1999) played fragments of chord progressions and asked listeners to rate the tension at the end of each fragment; the fragments started at the beginning of the excerpt and progressively grew longer until the end was reached. This "stop-tension" method was also used in Lerdahl and Krumhansl's tonal tension study (2007). Granot and Eitan (2011) asked listeners to give discrete overall tension ratings for short melodic sequences and select text responses that best described how tension was changing during the course of the sequence.

Granot and Eitan (2011) chose to use discrete judgments instead of online judgments due to the brief nature of their stimuli. Their materials consisted of short, melodic sequences that were composed specifically for the experiment. In regard to the stimuli, they mentioned that although ecological validity of the materials might be questioned, constructing them is the only reliable means to systematically control the musical parameters examined. Two other studies mentioned previously have also employed researcher-composed music: three-chord harmonic sequences (Bigand et al., 1996) and a 31-chord harmonic progression (Bigand & Parncutt, 1999).

Researcher-composed stimuli are uncommon, however. Most studies have used excerpts from the Western classical repertoire ranging from the mid-Baroque era through the late twentieth century (see Table 1). There are only a few examples of non-classical music that have been used in tension studies: "St. Louis Blues" by W. C. Handy sung by Nat King Cole and Ella Fitzgerald (Fredrickson & Coggiola, 2003), "We are the Champions" by Queen, and "I Feel Good" by James Brown (Rozin et al., 2004).

Regarding the participants in tension experiments, many previous studies have examined differences in tension responses between musicians and nonmusicians. Generally speaking, results indicate that both groups respond similarly (Bigand & Parncutt, 1999; Fredrickson, 2000; Fredrickson & Coggiola, 2003; Frego, 1999; Lychner, 1998). Fredrickson (1997) noted that even between two very different groups—second graders and professional musicians—correlation of tension responses was high. Tension judgments also appear to be remarkably consistent for repeated trials; Bigand and Parcutt (1999) noted in one of their experiments that the ratings for the first hearing of an excerpt did not differ much from the fourth hearing.

TABLE 1. Music From the Classical Repertoire Used in Tension Studies

Composer	Study
Stradella	Ilie & Thompson, 2006
Purcell	Rozin et al., 2004
Albinoni	Krumhansl, 1997
Vivaldi	Krumhansl, 1997; Frego, 1999; Ilie & Thompson, 2006
Bach	Lychner, 1998; Toiviainen & Krumhansl, 2003; Lerdahl & Krumhansl, 2007
Handel	Ilie & Thompson, 2006
Alberti	Ilie & Thompson, 2006
Haydn	Nielsen, 1987; Fredrickson, 1995, Fredrickson, 1997; Burnsed & Sochinski, 1998-2001; Ilie & Thompson, 2006
Mozart	Fredrickson & Johnson, 1996; Krumhansl, 1996; Rozin et al., 2004; Ilie & Thompson, 2006
Beethoven	Lychner, 1998; Misenhelter, 2001; Rozin et al., 2004
Bellini	Frego, 1999
Chopin	Bigand & Parncutt, 1999; Rozin et al., 2004; Lerdahl & Krumhansl, 2007
Wagner	Rozin et al., 2004; Lerdahl & Krumhansl, 2007
Brahms	Rozin et al., 2004
Borodin	Rozin et al., 2004
Mussorgsky	Krumhansl, 1997
Sousa	Lychner, 1998
Puccini	Lychner, 1998
Debussy	Hackworth & Fredrickson, 2010
R. Strauss	Nielsen, 1987
Hugo Alfvén	Krumhansl, 1997
Holst	Krumhansl, 1997; Fredrickson, 1999; Rozin et al., 2004
Stravinsky	Vines et al., 2006
Shostakovich	Fredrickson, 2000
Messiaen	Lerdahl & Krumhansl, 2007
Barber	Krumhansl, 1997
Cage	Frego, 1999

Note: Composers are listed in order of birth date.

However, results from studies examining the relative contributions of various musical parameters to tension have indicated some differences between musicians and nonmusicans. One general observation has been that intersubject agreement is higher among musicians than nonmusicians (Bigand & Parncutt, 1999; Krumhansl, 1996). Granot and Eitan (2011) found pitch contour contributed only weakly to tension and that it affected musicians more than nonmusicians. The latter group only appeared to respond to melodic contour in interaction with other parameters. Bigand et al. (1996), on the other hand, had the opposite finding: melodic contour ("horizontal motion") affected nonmusicans considerably more than musicians.

Granot and Eitan (2011) found that with regard to both dynamics and tempo, responses did not differ between musicians and nonmusicians. They further concluded that dynamics was the most influential parameter, followed by pitch register. Granot and Eitan found that lower register is more strongly associated with higher tension values, but only for nonmusicians. They theorized that the "normative" pitch range is extended for musicians due to instrumental training and wider exposure to the expressive capabilities of music. On the other hand, Ilie and Thompson (2006) found that low-pitched music was rated more pleasant and less tense. However, their stimuli were polyphonic and tonal, and the "low" versions were only four semitones below the "high" versions. In contrast, Granot and Eitan used atonal, arrhythmic melodic sequences with far larger changes in pitch register. Another area of difference between musicians and nonmusicians appears to be sensitivity to harmony; musicians tend to be more influenced by tonal hierarchies and key regions than nonmusicians (Bigand & Parncutt, 1999; Bigand et al., 1996).

Despite the fact that previous studies have explored many aspects of music that contribute to the perception of tension, there is no universal framework or theory that describes how these disparate parameters combine to produce a global evaluation of tension. Nielsen (1987) noted that no particular feature was decisive in the experience of tension. Granot and Eitan concluded that neither a model of simple additivity nor "noncongruence"—the theory that parameters at odds with each other elicit tension increases—could explain the results. They suggested a model in which "the various parameters interact, their mutual influences depending, among other variables, on the relative strength or importance of the parameter, on music training, and on specific task demands." They proposed an ecological model in which perceived musical tension is affected by auditory cues for impending threat. While this theory could plausibly account for the parameters they explored in their study (dynamics, pitch register, pitch contour, and tempo), it does not account for features such as rhythm, meter, harmony, and tonality.

The objective of this paper is to propose a *quantitative* model that can predict real-time tension judgments for complex musical stimuli based on the interaction of any variable number of musical parameters. This parametric, temporal model of tension takes into account the real-time dynamics of listening, and more generally, the limitations of short-term memory. The empirical basis for the tension model was derived from two experiments: the first was a web-based study designed to gather data from thousands of subjects from different musical and cultural backgrounds, and the second was a smaller study designed to obtain real-time, continuous responses to complex musical stimuli. The goal of this work was to propose a cognitive framework for the tension model, formalize it mathematically, and then use the empirical data to estimate the value of a number of variables including the weights of the relative contributions of individual musical parameters and the influence of previous events on current tension perception.

Experiment 1

The primary purpose of Experiment 1 was to determine or verify whether several disparate musical parameters individually contributed to the perception of tension and gauge the relative contributions of each parameter. An additional goal was to confirm the *directionality* of each parameter—in other words, what types of changes in the feature would cause a decrease versus an increase in tension. The stimuli composed for the study examined onset frequency, tempo, dynamics (loudness), pitch height, harmony, rhythmic regularity, and meter. Although these were not the only possible features that could be evaluated, they formed an adequately diverse foundation for analyzing score-based music.

Onset frequency, rhythmic regularity, and meter are features that have not been explored systematically in previous work. Onset frequency is not necessarily mutually exclusive with other features; for example, it might overlap with tempo when a ritard or accelerando is rhythmically written out; in other cases, it might be classified as a subtype of texture or density. It was assumed that an increase in onset frequency would result in an increase in tension. On the other hand, it was unclear whether rhythmic regularity and meter would affect tension perception at all. The initial hypotheses were (1) an increase in rhythmic irregularity would result in an increase in tension and (2) a shifting of meter in which perceived strong beats occurred more frequently would also result in an increase in tension. In the latter case, meter was hypothesized to be a hierarchical case of onset frequency.

Method

The format of Experiment 1 was modeled after the Moral Sense Test, a web-based study that explored the nature of human moral judgment (Hauser, Cushman, Young, Jin, & Mikhail, 2005). The design of the Moral Sense Test allowed for a very extensive set of stimuli—

far more than could normally be presented to a single subject in a typical experimental session. Given that the Moral Sense Test had thousands of participants, only a subset of the stimuli needed to be presented to each subject in order to obtain statistically significant results. Likewise, the web-based tension study was brief and utilized a fraction of the total possible stimuli for each participant. The nature of the methodology also allowed for new stimuli to be added at any time. For example, the stimuli exploring meter and rhythmic regularity were added a few months after the study was made available to the public. The web interface was implemented in collaboration with Josh McDermott; the first two parts of the study were designed by McDermott and are not related to the research described here, and the third part was the tension experiment. The interface was written in PHP with Flash for sound playback, and the data was stored in a MySQL database, all of which ran on a dedicated Linux server.

The study employed retrospective, discrete judgments of tension for relatively short stimuli, in a similar manner to other work (Bigand et al., 1996; Granot & Eitan, 2011). The main difference was that graphical shapes were used to depict tension changes instead of verbal descriptions (Figure 1). Two major reasons for this methodology were to (1) provide participants with more intuitive responses to choose from rather than wordy descriptions, and (2) to enable test takers without expertise in English to take part in the study. The changes in tension depicted by the graphical response choices were linear or curvilinear in shape. From a theoretically perspective, the shapes provided an intuitive, visual analog of musical gestures that the tension variations were meant to evoke. On the other hand, this type of visual analog also presented some problems, namely, the possibility that listeners would choose shapes that simply corresponded to change of any sort, not necessarily increase or decrease in tension. However, the instructions for the trials did ask participants to listen specifically for

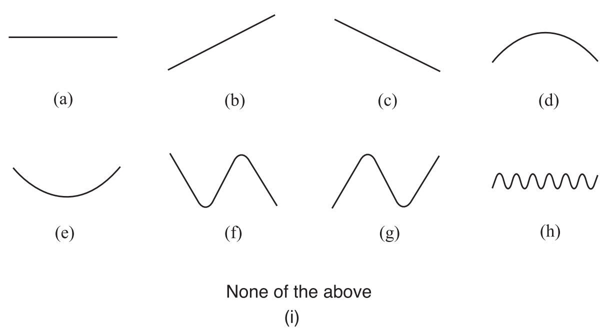


FIGURE 1. Response choices for Experiment 1.

tension. It must be assumed that most subjects were cognizant of the task. Furthermore, as will be discussed in detail later, a comparison of responses to stimuli used in both Experiments 1 and 2 shows that the results of this new methodology correlate highly with results obtained through more typical methods.

Participants. A total of 2,661 subjects from 108 different countries took part in the study. The participants' top five countries of residence were the U.S. (52%), Italy (8%), Canada (6%), Russia (3%), and Japan (3%). While there was substantial variety in the musical backgrounds of the subjects, there were considerably more musically untrained subjects and novices than musicians. Based on self-rated musical experience, 17% of the subjects were categorized as musicians and 83% as nonmusicians. Participants who self-rated a 4 out of 5 or higher on overall level of music training were classified as musicians, and those who self-rated 3 or lower were classified as nonmusicians. The median age was 33 and the average age 38 (SD = 14.7). Subjects were recruited through links on other websites such as the Moral Sense Test as well as the author's personal website. The URL for the study was also announced at conference presentations of other work. The study was publically accessible on the internet and data collection proceeded for two years and seven months.

It is possible that some subjects repeated the study multiple times since there was no way of verifying the identities of the participants. However, given that each subject was presented with a small subset of the stimuli, it was unlikely that there would be multiple ratings for the same stimulus from the same subject even if the person took the study multiple times.

Stimulus materials. The stimuli were composed so that each feature was isolated in at least one case and combined with other parameters moving in the same or opposite direction in various other cases. For example, a single tone repeating in absence of any change except a progressive increase in tempo was a parameter-in-isolation stimulus for tempo (e.g., Figure 2, stimulus A15); the hypothesis was in this case that the listener would perceive an increase in tension. An example of a stimulus constructed to isolate dynamics was a single repeating tone that steadily decreased in loudness (Figure 2, A21). In this case, the hypothesis was that the opposite would occur—the listener would perceive tension decreasing. Other stimuli were designed to have two or more diverging features, for example, a decrease in onset frequency accompanied by an increase in dynamics (Figure 2, A30). This is an example of what Granot and Eitan (2011) describe as "noncongruence." In

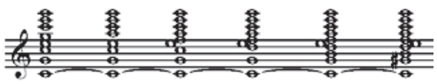
STIM ID	RESPONSE TYPE									Ave Conf	Num Resp
	0	1	2	3	4	5	6	7	NONE 8		
A03											
Responses	15% (3.6)	54% (4.4)	0% (0)	11% (4.3)	9% (3.6)	4% (3)	4% (4)	2% (4)	2% (3)	4.13	54
A04											
Responses	16% (4.2)	41% (4.1)	0% (0)	13% (3.3)	5% (4)	3% (2.5)	10% (4)	2% (3)	10% (4.1)	3.95	61
A05											
Responses	8% (4.1)	3% (4)	1% (2)	50% (4)	7% (3.4)	11% (4)	13% (3.3)	0% (0)	7% (3.4)	3.83	72
A06											
Responses	14% (4.1)	0% (0)	0% (0)	46% (4.4)	7% (4)	6% (3.7)	19% (4.1)	0% (0)	7% (3.8)	4.20	69
A07											
Responses	13% (4.3)	4% (4)	1% (3)	17% (3.7)	18% (4)	18% (3.9)	26% (3.8)	1% (4)	1% (4)	3.96	72
A08											
Responses	20% (3.6)	1% (5)	3% (3.5)	16% (3.8)	25% (3.8)	13% (3.5)	15% (3.9)	3% (4)	4% (4.6)	3.82	79
A11											
Responses	24% (3.5)	2% (5)	0% (0)	17% (3.6)	14% (4.3)	5% (4.3)	21% (4)	14% (4.1)	3% (3)	3.90	58
A12											
Responses	25% (4)	1% (2)	1% (4)	18% (3.6)	6% (4)	10% (3)	27% (3.7)	6% (3.7)	6% (3.2)	3.70	71
A13*											
Responses	29% (3.8)	9% (3.7)	6% (3.1)	15% (3.5)	2% (3.4)	5% (3.1)	7% (3.4)	12% (3.5)	15% (3.5)	3.59	1136
A14											
Responses	5% (3.7)	0% (3.7)	0% (4)	78% (4.6)	0% (3.5)	2% (3.6)	7% (3.6)	3% (3.7)	4% (3.8)	4.42	937
A15+											
Responses	26% (4.1)	28% (3.9)	4% (3.6)	4% (3.2)	1% (2.6)	0% (2)	2% (4.1)	27% (4)	8% (3.6)	3.95	503

FIGURE 2. Stimuli and the response profiles from Experiment 1. All stimuli were rendered in both piano and orchestral strings timbres except those that are starred. A star (*) indicates the stimulus was only rendered in piano timbre. A plus (+) indicates that the stimulus was rendered in unpitched percussion timbre in addition to piano and strings timbres. If no metronome marking is indicated, the tempo for the stimulus is quarter note equals 96 beats per min. (*Continued*)

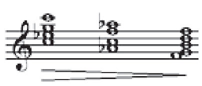
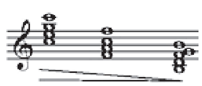
STIM ID	RESPONSE TYPE 0	1	2	3	4	5	6	7	NONE 8	Ave Conf	Num Resp
A16											
Responses	34% (4.1)	28% (3.9)	1% (3.8)	1% (3.2)	2% (3.5)	0% (2.7)	1% (3.8)	23% (3.8)	9% (3.9)	3.96	973
A17+											
Responses	31% (4.1)	14% (4)	1% (2)	3% (4.2)	1% (5)	1% (3)	1% (2.5)	28% (3.9)	21% (4)	3.99	140
A18+											
Responses	29% (4.1)	2% (3.1)	28% (3.9)	5% (3.9)	1% (3.8)	0% (4.5)	1% (3)	26% (4)	6% (3.8)	4.01	529
A19											
Responses	40% (4.1)	1% (3.4)	22% (3.8)	5% (3.3)	2% (3)	1% (3)	2% (2.7)	20% (3.6)	6% (3.3)	3.83	435
A20+											
Responses	32% (4)	2% (3.7)	31% (3.9)	5% (3.5)	3% (3)	1% (3.5)	1% (2.6)	16% (3.6)	9% (3.6)	3.83	645
A21											
Responses	39% (4.3)	1% (3.6)	41% (4.3)	2% (3.7)	0% (3.3)	1% (3.5)	1% (4.3)	12% (4)	4% (3.7)	4.25	1120
A22*											
Responses	5% (3.5)	29% (4)	48% (4.1)	2% (2)	7% (3.6)	0% (0)	7% (3.6)	0% (0)	2% (4)	3.98	42
A23											
Responses	7% (3.3)	6% (4)	33% (4.3)	22% (3.6)	6% (3.6)	9% (3.7)	6% (4.2)	1% (3)	9% (3.5)	3.89	81
A24*											
Responses	5% (3.7)	3% (2)	55% (4)	8% (3.8)	8% (3.6)	8% (3)	5% (3.5)	0% (0)	9% (3.5)	3.77	77
A25											
Responses	6% (4.4)	2% (4)	62% (4.2)	5% (3.3)	5% (3.5)	8% (3.6)	5% (3.8)	1% (4)	6% (4)	4.06	185
A26*											
Responses	7% (4.3)	1% (4)	50% (4.1)	3% (4.3)	12% (3.6)	10% (3.1)	3% (2.6)	1% (1)	12% (3.8)	3.87	90

FIGURE 2. (*Continued*)

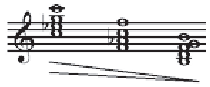
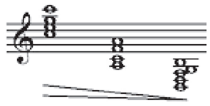
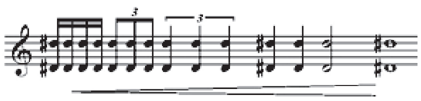
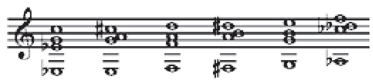
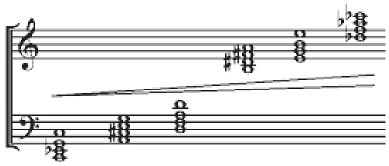
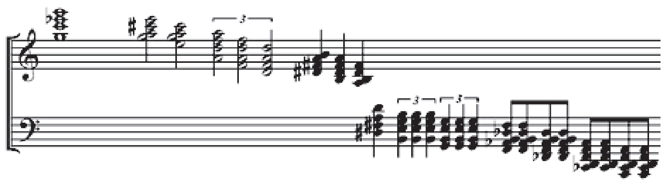
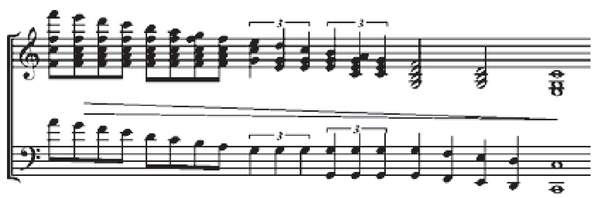
STIM ID	RESPONSE TYPE									Ave Conf	Num Resp
	0	1	2	3	4	5	6	7	NONE 8		
A27											
Responses	5% (3.6)	2% (2.7)	71% (4)	5% (3.3)	3% (3.3)	5% (3.4)	5% (3.3)	1% (3)	4% (4.2)	3.88	194
A28*											
Responses	7% (3.8)	6% (3.7)	47% (4.1)	3% (3.5)	11% (3.3)	10% (3.2)	0% (0)	0% (0)	17% (3.5)	3.79	72
A29											
Responses	4% (4)	1% (4)	69% (4.2)	8% (3.7)	4% (3.3)	4% (3.3)	3% (2.8)	1% (3)	6% (3.6)	4.04	190
A30*											
Responses	39% (4.1)	11% (4)	11% (3.6)	9% (2.7)	3% (3.7)	0% (0)	2% (2.6)	13% (3.8)	11% (3.4)	3.80	122
A46											
Responses	5% (3.8)	83% (4.5)	1% (4.1)	1% (4)	1% (3.6)	1% (3.7)	2% (3.6)	3% (3.6)	2% (3.3)	4.37	1007
A47											
Responses	3% (4)	80% (4.3)	0% (0)	1% (3)	2% (3.5)	2% (2.3)	8% (4)	1% (5)	3% (4)	4.27	172
A50											
Responses	2% (3.5)	15% (3.8)	26% (4)	16% (3.7)	12% (3.5)	10% (3.5)	8% (3.3)	3% (3.4)	9% (3.4)	3.72	1013
A52											
Responses	1% (3.7)	1% (4)	81% (4.6)	7% (3.9)	2% (3.5)	4% (3.7)	1% (3.4)	2% (3.6)	1% (3.9)	4.46	2661

FIGURE 2. *(Continued)*

STIM ID	RESPONSE TYPE 0	1	2	3	4	5	6	7	NONE 8	Ave Conf	Num Resp
A53*											
Responses	48% (4.1)	2% (3.3)	4% (3.6)	3% (2.7)	1% (3.8)	3% (2.8)	1% (3.5)	20% (3.7)	18% (3.8)	3.90	517
A54+											
Responses	29% (3.6)	20% (3.6)	0% (0)	1% (2)	2% (2.6)	1% (3)	1% (1.5)	25% (3.8)	20% (3.9)	3.69	143
A55+											
Responses	36% (4.1)	12% (3.2)	0% (0)	5% (3.1)	1% (2)	0% (0)	2% (2.5)	29% (3.7)	15% (3.6)	3.74	118
A56+											
Responses	44% (4)	10% (3.5)	1% (3.7)	3% (3)	1% (3.3)	1% (2)	1% (2.5)	20% (3.6)	17% (3.5)	3.76	272
A57+											
Responses	38% (4.1)	16% (3.8)	0% (0)	2% (4)	2% (3.5)	1% (2.5)	2% (2.7)	23% (3.8)	18% (3.9)	3.95	262
A58+											
Responses	62% (4.3)	1% (3.2)	2% (3)	2% (3.1)	1% (3.3)	1% (3.6)	2% (3.4)	18% (3.9)	10% (3.6)	4.08	283
A59+											
Responses	57% (4.1)	2% (3.7)	4% (3.4)	2% (3.2)	1% (2.5)	2% (3)	2% (3.3)	22% (4.3)	9% (3.4)	4.04	261
A60											
Responses	63% (4.3)	1% (4)	5% (3)	1% (2)	0% (0)	3% (2.5)	3% (3.5)	17% (3.6)	8% (3.1)	3.94	78
A61+											
Responses	56% (4.3)	2% (3.6)	4% (3.2)	3% (3.7)	2% (2.6)	0% (0)	2% (4)	23% (4.2)	8% (3.5)	4.10	124
A62*											
Responses	55% (4.3)	2% (3.8)	1% (3.3)	3% (3.2)	1% (3.1)	1% (2.8)	3% (3.5)	24% (3.6)	9% (3.3)	3.95	536
A63+											
Responses	44% (4)	1% (5)	7% (3.4)	4% (3.3)	2% (3.6)	1% (3.5)	1% (3)	25% (4)	13% (3.3)	3.87	134
A64+											
Responses	55% (4.3)	2% (3.6)	4% (3.6)	2% (3.6)	1% (2)	3% (3)	0% (0)	22% (3.8)	9% (3.6)	4.05	121
A65+											
Responses	57% (4.2)	12% (3.2)	0% (0)	1% (3.5)	1% (5)	3% (2.5)	4% (3.4)	16% (4.1)	6% (3.2)	3.99	140

FIGURE 2. (*Continued*)

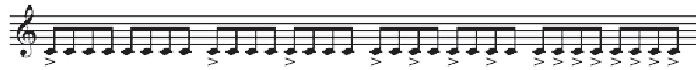
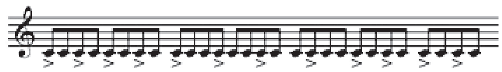
STIM ID	RESPONSE TYPE									Ave Conf	Num Resp
	0	1	2	3	4	5	6	7	NONE 8		
A66+											
Responses	42% (4.1)	15% (3.9)	0% (0)	2% (4)	2% (2.6)	2% (4)	2% (3.3)	24% (4)	11% (3.5)	3.97	123
A67+											
Responses	20% (4.2)	25% (4)	0% (0)	3% (4)	2% (3.6)	1% (2.5)	3% (3.2)	36% (3.9)	9% (4.1)	3.99	268
A68+											
Responses	31% (4.3)	0% (3)	1% (3.5)	4% (3)	10% (3.4)	2% (4.2)	5% (3.5)	39% (3.9)	7% (3.7)	3.93	246
A69+											
Responses	36% (4.1)	2% (3.8)	0% (0)	8% (3.9)	5% (3.2)	2% (3.8)	3% (3.8)	39% (3.8)	5% (3.2)	3.87	259
A70+											
Responses	47% (4.2)	4% (3.4)	0% (0)	5% (3.5)	0% (0)	3% (3.2)	1% (3)	33% (3.9)	8% (3.4)	3.97	276
B26	Bach-Vivaldi excerpt, mm. 97-100										
Responses	4% (3.3)	28% (4.2)	2% (3.5)	0% (0)	12% (3.4)	2% (2.5)	16% (4)	26% (4.2)	9% (3.8)	3.99	85
B27	Bach-Vivaldi excerpt, mm. 97-104										
Responses	3% (4.3)	48% (4.1)	1% (4.2)	1% (3.7)	5% (3.6)	4% (3.6)	18% (3.8)	12% (4.1)	7% (3.9)	4.02	307
B28	Bach-Vivaldi excerpt, mm. 99-100										
Responses	6% (3.8)	40% (4)	1% (3)	2% (3.5)	4% (3.2)	6% (3.1)	15% (4.2)	20% (4)	5% (4.2)	3.96	98
B29	Bach-Vivaldi excerpt, mm. 101-102										
Responses	16% (4.1)	10% (3.8)	0% (0)	1% (1)	5% (3.7)	2% (3.5)	17% (3.8)	35% (3.9)	13% (3.5)	3.84	86
B30	Bach-Vivaldi excerpt, mm. 101-104										
Responses	8% (3.5)	37% (4)	1% (3)	1% (2)	3% (3.2)	4% (3)	14% (4)	29% (4)	3% (4.2)	3.90	119
B31	Bach-Vivaldi excerpt, mm. 103-104										
Responses	5% (4)	38% (4)	2% (3.5)	1% (4)	4% (3.2)	7% (3.7)	12% (4.4)	21% (4)	10% (4)	4.02	107

FIGURE 2. (*Continued*)

these cases, the resulting change in overall tension was difficult to predict and there was no hypothesized "correct" response for this type of stimulus. The intent was to ascertain how the responses differed and how uncertain the listeners were.

Harmony as a parameter was more difficult to define due to its multidimensional nature. In order to determine the harmonic contour of a stimulus, hierarchical tension values were calculated using Lerdahl's tonal tension model (2001) without the melodic attraction component. Melodic attraction was excluded because it overlapped to some extent with pitch height. Figure 3 illustrates a stimulus that was designed to isolate harmony along with its corresponding prolongational analysis and hypothesized response curve. See Appendix A for a detailed description of how harmonic tension was calculated using Lerdahl's model.

The stimuli ranged from 2 to 60 s in duration, and were recorded using MIDI piano, strings, and unpitched

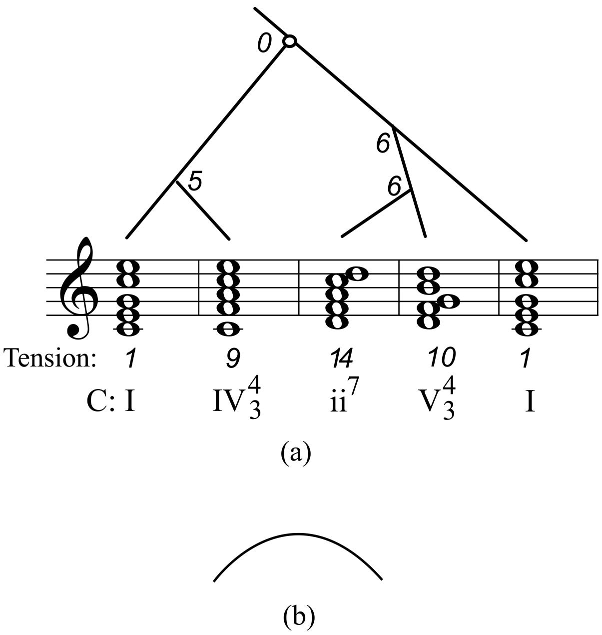


FIGURE 3. (a) An example from Experiment 1 that tested subjects' responses to harmony. The prolongational reduction is shown above the staff; (b) the hypothesized response to the stimulus.

percussion sounds. Each stimulus was entered in Finale (a notation editor) and recorded using two MIDI synthesizers, a Kurzweil K2500 and Roland 5080. The volume scaling for stimuli with changes in loudness was done in an audio editing program. Due to perceptual differences in loudness sensitivity depending on the frequency range, some stimuli were adjusted by hand to sound more perceptually consistent. There were a total of 207 unique audio files created for the experiment. In some cases, several versions of the same stimulus were recorded, varying only in the choice of timbre (piano, strings, or percussion) or modality (major or minor).[1]

Each stimulus was categorized by group type (Table 2). Stimuli within the same group were similar and explored the same musical parameters in slightly different ways. For example, there were multiple versions of a simple I-V^7-I harmonic progression in which the contour of the melodic line differed slightly or the mode was altered (e.g., Figure 2, A03-A08). In addition to stimuli composed specifically for the study, there were fragments from a J. S. Bach organ transcription (BWV 594) of the Vivaldi D Major concerto (RV 208) shown in Figure 4. This particular excerpt was chosen because it was a relatively obscure work and contained clear changes in harmony and key. Furthermore, it was a richer and more ecologically valid stimulus than the other materials. Segments selected from the piece were taken from mm. 81-104 of the first movement and included the entire 24 measures as well as various smaller sections that encompassed whole and partial phrases, transitions, and portions with altered tempo and dynamics.

Task. Subjects navigated to the homepage of the study (a Massachusetts Institute of Technology website), where they were initially asked to choose between an English and Chinese version of the test. From the main page, subjects navigated to different tabs describing the research, the researchers involved, and the study itself (labeled "Start Now!"). Upon clicking on the last tab, a popup window appeared that contained the test materials. A series of six sine tones ranging from 200 Hz to 3,500 Hz were played repeatedly at loud and soft volume levels to help subjects calibrate their speakers properly. This was necessary since observers were not present to ensure listeners heard the stimuli with minimal distortion. Subjects were then asked to listen to audio files and choose graphical shapes that best corresponded to how they perceived tension changing throughout each given audio excerpt. The graphical choices were always ordered as shown in Figure 1. Each stimulus was played twice and all questions were multiple choice. Two hearings ensured that the subject had another chance to listen to the stimulus if they were distracted or interrupted during the course of study—something that cannot be controlled for in an unproctored experimental setting. It is possible that hearing the stimuli a second time might have affected the subjects' responses in some way, but prior work on the consistency of tension responses to multiple hearings discussed above indicates this was not a problem. Finally, subjects were asked to rate the confidence level of their response on a scale from "1" ("least confident") to "5" ("most confident").

Each subject was presented with 11 trials, only the first of which was not chosen randomly: stimulus A52. This first trial was considered to have an "obvious" solution. It served as a practice trial and helped familiarize the participants with the task. Trials 2 through 9 were selected by randomly choosing a group number excluding 111 and 112 (the Bach-Vivaldi excerpts) and then randomly selecting a stimulus from within that group. If the stimulus chosen had been recorded with different instrumental sounds, the timbre was selected randomly as well. Only one stimulus per group could be chosen for any given subject. Previously selected groups could not be chosen again in subsequent trials; in other words, if a stimulus from Group 2 (onset frequency) was chosen, no subsequent trial presented another stimulus from Group 2 again. A stimulus from Group 111 (Bach-Vivaldi) was always chosen for trial 10 and a

[1] All of the stimuli for Experiment 1 and complete response profiles are available at http://www.media.mit.edu/~mary/tension_exp1.

TABLE 2. Classification of Stimuli in Experiment 1

Group	Number of stimuli*	Description
1	7	Simple harmonic progressions with subtle changes in voicing
2	1	Changes in onset frequency
3	1	Changes in pitch height
4	1	Changes in harmony and pitch height corresponding
5	4	Changes in harmony against changes in onset frequency
6	1	Changes in harmony and onset frequency against loudness
7	4	Changes in harmony with differing degrees of changes pitch height
8	4	Different combination of changes in onset frequency, harmony, and pitch height
9	1	Changes in onset frequency, harmony, and pitch height against loudness
10	1	Changes in onset frequency, pitch height, and loudness against harmony
11	2	Changes in onset frequency, harmony, pitch height, and loudness all increasing in intensity
12	1	Changes in harmony, pitch height, and loudness against onset frequency
13	1	Onset frequency, harmonic tension, pitch height, and loudness all decreasing in intensity
14	1	Changes in onset frequency, harmony, and loudness against pitch height
15	1	Increasing dissonance
16	2	Changes in rhythmic regularity
17	1	Changes in loudness
20	2	Changes in tempo
21	2	Changes in tempo, some examples having notes with shorter articulation
22	9	Changes in rhythmic patterns (or rhythmic regularity) coupled with slight changes in onset frequency
23	4	Simpler changes in rhythm (than in group 22) coupled with slight changes in onset frequency
24	4	Changes in accent placement or meter (related to onset frequency)
111	26	Excerpts from Bach-Vivaldi concerto
112	5	Excerpts from Bach-Vivaldi concerto with tempo and dynamics changes

*Note: Does not include versions that are different only in timbre or modality.

stimulus from Group 112 (Bach-Vivaldi with altered tempo and dynamics) was always chosen for trial 11.

One result of having varying numbers of stimuli for each group was an uneven distribution of responses per trial. With so many subjects, however, this again was not a significant problem. Another reason for the unbalanced response distribution was due to the addition of some stimuli midway through the data collection process. In particular, these included stimuli in Groups 20 and higher, which mostly covered stimuli that featured rhythmic irregularity, meter, and Bach-Vivaldi excerpts.

Results

The strategy employed to analyze the data focused on determining whether subjects chose a response curve that corresponded to an individual feature. For all parameters, subjects were considered to be responsive to that feature *if they selected a tension curve that matched the tension contour of that particular feature*. The curves in Figures 1f and 1g were rarely selected either because there were very few stimuli that had more than two directional changes or because those two graphs depicted more changes in tension than subjects could recall with certainty in a single retrospective judgment. This is further supported by the fact that subjects chose the curve shown in Figure 1h, a sinusoid shape depicting a large number of changes, far more often. Statistical significance was determined by a chi-square goodness-of-fit test on two or more categories. An equal distribution across all nine response choices was assumed and responses to curve types that did not correspond to the feature or features being examined were combined. For example, if dynamics was being examined for a particular stimulus, the expected null hypothesis response rate for the curve type matching dynamics was 1/9 the total number of responses, and the expected number of responses for all other choices were combined into one category, or 8/9 of the total number of responses. If two conflicting features were compared, each feature had expected response rates of 1/9 the total responses, and all the other responses were combined for an expected response rate of 7/9 the total responses. When multiple stimuli were examined with respect to a particular

FIGURE 4. Measures 81-104 of J. S. Bach's organ transcription, BWV 594, of the Antonio Vivaldi's D Major Concerto, RV 208. Segments of this excerpt were used in Experiment 1 (stimuli B01-B31) and in Experiment 2 as Excerpt 5.

feature, all of the responses that corresponded to the tension curve of that feature for each stimulus were put into one category, and the combined responses for all other response types into another category for each stimulus. Each category was then combined *across* the different stimuli before a chi-square test was performed on the two categories. In addition to the goodness-of-fit tests, chi-square tests of independence were used for comparisons between factors such as differences in responses between musicians and nonmusicians.

There are several primary conclusions that can be drawn from the data. First, all parameters with the exception of the rhythmic features had an impact on listeners' perception of tension. Second, the parameters that had the clearest effect on tension in isolation were dynamics and pitch height. Musicians responded more strongly than nonmusicians to all features with the exception of pitch height. In tonal contexts, musicians were more sensitive to harmonic changes while nonmusicians responded more to pitch height changes. In the case of more complex stimuli where features counteracted each other, the results tended to be ambiguous; when multiple features were combined in parallel, they considerably strengthened the feeling of tension changing in a particular direction. In general, responses selected more frequently by listeners had higher confidence ratings. There was a positive correlation between the average confidence value for each response choice for each stimulus and the corresponding frequency of that choice, $r(868) = .41, p < .001$ (Pearson's correlation).

Onset frequency and tempo. Stimuli A18 and A20 are two examples of stimuli designed to isolate onset frequency and tempo, respectively (for consistency, A19 is not included here because it did not have an unpitched percussion version unlike A18 and A20). Results indicated that a decrease in both features resulted in a decrease in tension. A chi-square goodness-of-fit test with regard to tempo for A18 and onset frequency for A20 are significant: $\chi^2(1, N = 529) = 157.98, p < .001$ for A18, $\chi^2(1, N = 645) = 248.21, p < .001$ for A20. The response profiles for the two stimuli are strikingly similar; however, this is not particularly surprising given that A20 is effectively a written-out ritard. Likewise, A15 and A16 are designed similarly except for a reversal in direction of onset frequency and tempo change. The response profiles correspond closely to their directional opposites with "increasing" selected about as frequently for A15 an A16 as "decreasing" for A18, A19, and A20.

Other than "increasing" or "decreasing," the most frequently chosen response was "no change." A closer look at pitch-related effects reveals that the response to onset frequency and tempo is more evident when the timbre of the stimulus is unpitched percussion. The response profiles for pitched versus unpitched versions of the A18 and A20 are given in Figure 5. A chi-square test of independence for the differences between the two conditions are significant, $\chi^2(2, N = 1174) = 37.78, p < .001$. These results indicate the pitched versions of the stimuli did not isolate tempo and onset frequently as effectively when the note events were pitched. As a result, some subjects responded to the lack of change in pitch more than the change in tempo or onset frequency.

Pitch height and harmony. Perhaps more than for any other feature, visual curves as analogs for tension were most directly applicable in the case of pitch height, particular if envisioned in staff notation. Results indicated that an increase in pitch height corresponded to an increase in tension; stimulus A14 was designed to isolate pitch height, and the responses highly correlated with the rising then falling curve of the pitch contour, $\chi^2(1, N = 937) = 4197.95, p < .001$. Unlike pitch height, harmony proved more difficult to relate directly to individual curves due to its abstract nature, as well as the fact that it was impossible to isolate changes in harmony from changes in pitch height. In order to address the latter issue, the results were divided into two categories: stimuli where the harmonic movement coincided with the change in pitch height of the melodic line (A05, A06, A46) and stimuli where change in harmony diverged from the pitch contour (A03, A04, A07, A08). Stimuli A03-A08 all contain the same simple three-chord harmonic progression, I-V^7-I in major or minor mode. However, the response to harmony was stronger when the melodic contour corresponded with the harmonic tension curve, $\chi^2(1, N = 266) = 3.04, p = .08$ for A03, A04, A07, and A08 combined (diverging melodic contour), $\chi^2(1, N = 141) = 190.63, p < .001$ for A05 and A06 combined (parallel contour).

Although changes in pitch height interacted with harmony, increasing the amount of change in pitch height

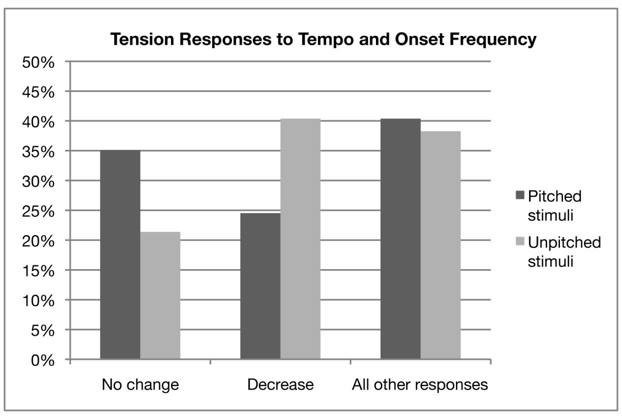


FIGURE 5. A comparison of combined responses to pitched and unpitched versions of stimuli A18 and A20.

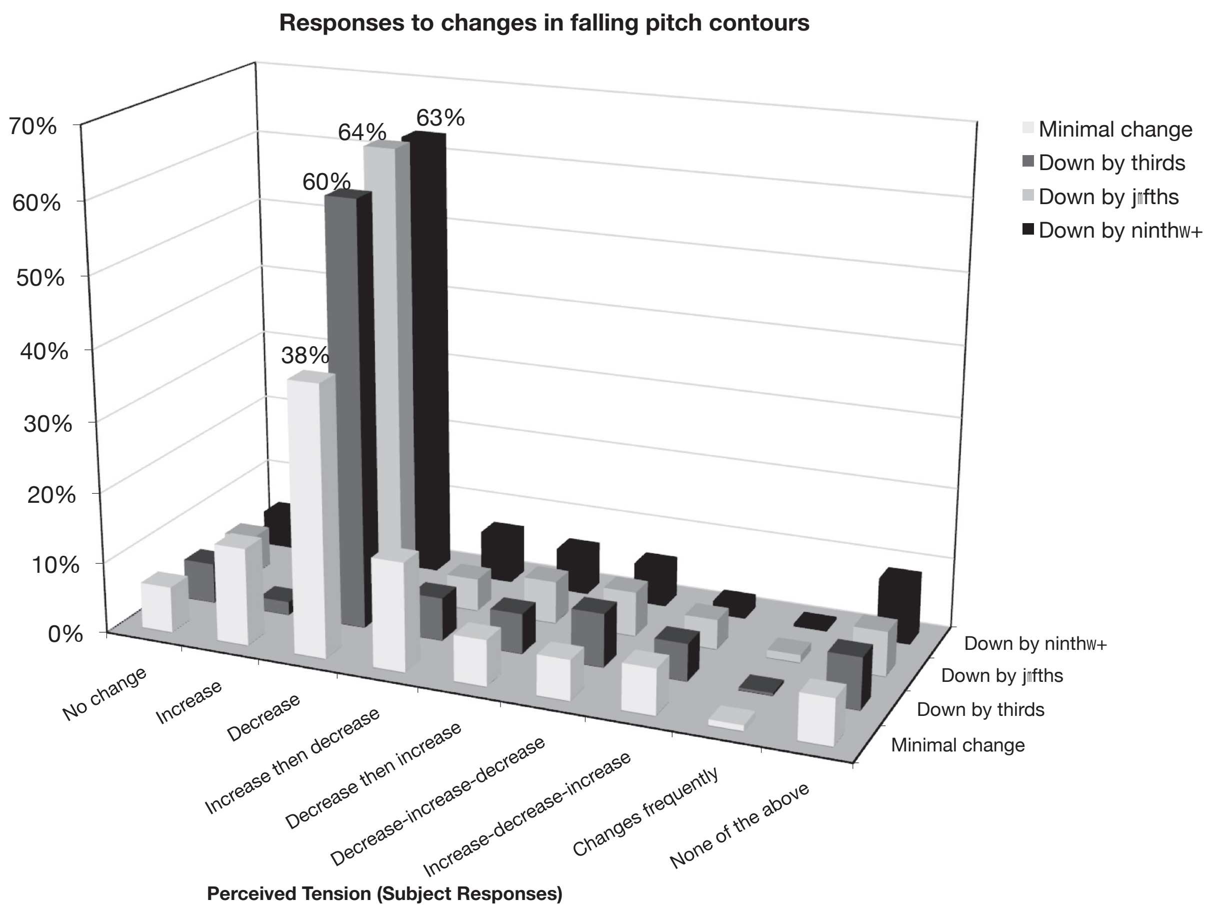


FIGURE 6. Response profiles for stimuli A22-A29 combined by size of melodic interval change.

only increased its effect up to a certain point. Stimuli A22-A29 all have the same harmonic progression and same change in loudness but different degrees of change in pitch height. Given the responses to all four melodic contours (Figure 6), it appears that an increase in interval size between the second and last note corresponded to an increase in tension only for intervals of a third or smaller, $\chi^2(3, N = 931) = 69.25, p < .001$). The response rate for interval descents of a third, fifth, and ninth are fairly close.

Dynamics. As expected, an increase in loudness corresponded to an increase in tension. Stimulus A21 was designed to isolate the effect of loudness, and its response profile with dynamics as the target feature is statistically significant: $\chi^2(1, N = 1120) = 987.56, p < .001$. Much like the case for onset frequency, there were almost as many "no change" responses (39%) as "decreasing" responses (41%), most likely due to the rendering of the stimuli with pitched strings and piano sounds instead of unpitched percussion.

Rhythm and meter. Rhythmic regularity, syncopation, and meter were the only features examined that did not appear to influence the perception of tension. Stimuli A53 and A62 explored the idea that unpredictability of rhythmic onsets contributed to tension. They were designed to either start or end with predictable isochronous onset events and morph to or from highly irregular rhythmic patterns. The response profiles for both A53 and A62 have the highest response frequencies for either no change or highly variable change—there does not appear to be any indication of directionality. Stimuli A54-A61 and A63-A66 were designed to further explore rhythmic regularity by utilizing syncopation and simple alterations in repeated rhythmic patterns. At first glance, there seems to be some sensitivity to rhythmic changes for A54 and A57, indicating higher than chance selection of increasing tension. If only unpitched versions of these stimuli are examined, it becomes more evident, with 31% of subjects selecting "increasing" for A54 and 22% selecting "increasing" for A57, $\chi^2(1, N = 136) = 24.28$, $p < .001$ for a target response of "increasing" with responses to both stimuli folded together. However, these results are complicated by the fact that rhythmic alterations cannot be completely isolated from changes in onset frequency much in the same way harmony cannot be completely separate from pitch-contour changes. Therefore, it is possible that what has been observed is listener response to onset frequency rather than rhythmic regularity.

Stimuli designed to test tension changes resulting from accent placements (A67-A70), some of which attempted to simulate metrical changes, yielded weak results as well. The most frequently chosen responses for all of them were either "no change" or "changes frequently." There were a statistically significant number of responses indicating increasing tension for A67, $\chi^2(1, N = 268) = 48.65, p < .001$. However, this might have been a result of perceived increase in onset frequency of the accented notes, similar to the rhythmic regularity examples discussed above. It is worth noting though that for A67, there was no significant difference in sensitivity between pitched and unpitched versions, unlike the previous cases discussed.

Musicians vs. nonmusicians. Comparisons between responses from musicians and nonmusicians indicated that musicians had greater sensitivity to harmony, tempo, onset frequency, and dynamics—all features with the exception of pitch height. In order to differentiate between musician and nonmusician responses to pitch height versus harmony, all responses to stimuli containing harmonic progressions with pitch height profiles diverging from the harmonic profiles (A03, A04, A07, A08, A11, and A12) were folded together by response type: all responses that matched the tension curve of the harmonic progression were then put into one category, all responses that corresponded to the tension curve of the pitch-height contour into the second category, and all other responses into a third category (Figure 7). Musicians responded more strongly to harmony while nonmusicians responded to pitch height; differences between musicians' and nonmusicans' responses were significant, $\chi^2(2, N = 395) = 17.86, p < .001$.

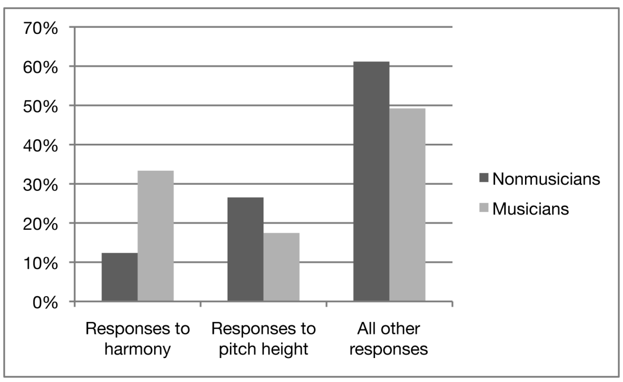


FIGURE 7. A comparison between musicians and nonmusicians showing responses to pitch height versus responses to harmony.

A comparison between the response profiles of musicians and nonmusicans for onset frequency, tempo, and dynamics indicated that musicians had greater sensitivity to all three features. A16, A17, A19, and A20 were used to compare responses to onset frequency, $\chi^2(1, N = 2193) = 24.30, p < .001$ (Figure 8a), A15 and A18 to compare responses to tempo, $\chi^2(1, N = 1032) = 9.34, p = .002$ (Figure 8b), and A21 to compare results for dynamics, $\chi^2(1, N = 1120) = 15.13, p < .001$ (Figure 8c).

General conclusions for isolated features. Regardless of the musical backgrounds of the participants, the two features that listeners responded to most strongly in isolation appeared to be dynamics and pitch height, followed by onset frequency and tempo. Given the stimuli used in the experiment, it was not possible to determine the difference in sensitivity between tempo and onset frequency. The type of onset frequency changes explored in this study resembled written-out tempo changes that increased or decreased linearly. In most musical passages, onset frequency does not tend to vary linearly in such a clear manner. From an ecological perspective, onset frequency is probably more analogous in effect to rubato-like tempo fluctuations rather than ritards and accelerandos. It was also not clear how the influence of harmony compared with the effect of other features since pitch height was always an added factor. However, even in the cases where small changes in pitch height contour matched harmonic tension (e.g., A05-A06), listeners did not respond to harmony as strongly as pitch-height alone (A14) or dynamics changes alone (A21).

Conflicts and convergences. For stimuli with two or more conflicting parameters, the presence of conflicting features appears to have created ambiguity. The opposite effect occurred when two or more features were changing in the same direction—multiple features in parallel resulted in stronger effects than single features in isolation. The ambiguous/reinforcing effects stemming from diverging or converging parameters is evident in the harmony versus melodic contour analysis discussed above. Another example of this can be observed for stimuli A47 and A50. In A47, harmonic tension, pitch height, and loudness are all increasing in parallel, yielding a response frequency for "increasing" of 80% across all subjects. In the case of A50, harmonic tension and onset frequency are increasing while pitch height is decreasing, yielding highly mixed results: 15% for "increasing" and 26% for "decreasing." Responses to A50 also reveal that descending pitch contour contributed more to perceived tension in this case than both harmony and onset frequency combined. This supports the previous observation that pitch height is a particularly influential component.

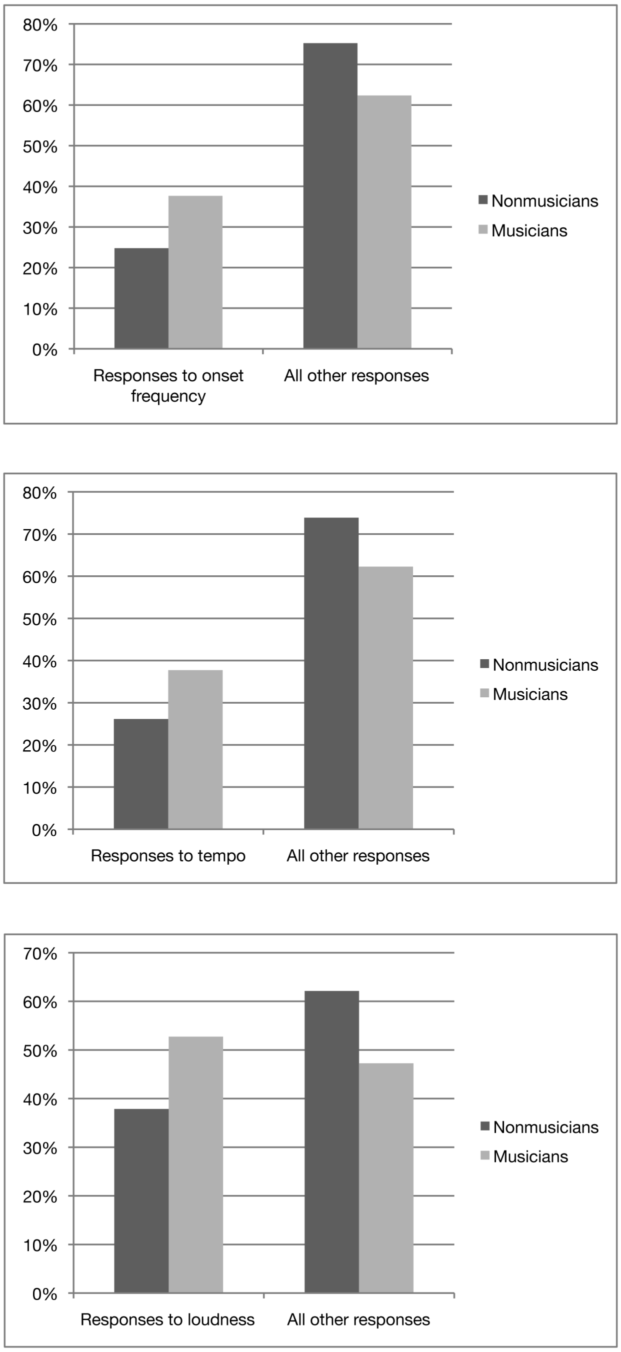


FIGURE 8. A comparison between musicians and nonmusicians for (a) sensitivity to onset frequency for stimuli A16, A17, A19, and A20, (b) sensitivity to tempo for A15 and A18, and (c) sensitivity to dynamics for A21.

In general, it was difficult to evaluate relative contributions of each feature in interactive situations because in most cases there was no way to measure the impact of salience. Most of the stimuli were designed to have changes that were as obvious as possible for the various parameters involved. Salience of any feature is dependent on the number of changes in a given time span. It can be assumed that this time span is restricted to some extent by short-term memory, but it cannot be assumed that there is a straightforward mapping between increase in change of a particular feature and increase in salience. The nonlinear response to change in pitch height is one example of this. That example also might reflect a ceiling effect: increase in pitch height only resulted in a corresponding increase in tension up to some maximal point.

Furthermore, salience is not just a matter of amount of change over time; it is also a function of whether there is an evident trend. This can be observed in the Bach-Vivaldi examples, where tempo and dynamics have been added to some passages. For the most part, it was difficult to draw conclusions from the Bach-Vivaldi excerpts due to the complex nature of the stimuli. In most cases, only added changes in dynamics or tempo enabled listeners to converge on a response other than "changes frequently." The exceptions were excerpts taken from mm. 97-104, in which there were no changes in dynamics, tempo, or onset frequency, but a clear increase in harmonic tension accompanied by a rise in pitch height (Figure 2, B26-B31; Figure 4, mm. 97-104). Made further salient by the lack of change in onset frequency, the harmonic motion strengthened by increasing pitch height appears to have induced a feeling of rising tension. This concept of salience, particularly in the context of complex musical textures, was an important consideration in the development of the quantitative, parametric tension model to be discussed later.

Experiment 2

One advantage of collecting data in the form of retrospective judgments is that it allows for a relatively simple experimental interface, useful in the case of a web-based experiment. However, it does have some limitations; judgments made by listeners after an excerpt has ended may not reflect the experience while the music is playing. It is also difficult to use long excerpts that change over time since these dynamic qualities are not well represented by a single retrospective judgment. In Experiment 2, continuous, real-time judgments of tension were collected through a software interface. The data collected in this study was essential to realizing a model that could predict a tension response based on descriptions of the way each musical feature changed over time. Assuming these feature descriptions were accurate, the goal was to build a model that could predict overall tension by taking into account how individual features influenced or contributed to changes in perceived tension at any point in a musical passage.

Method

Participants. Thirty-five subjects, recruited from the faculty and student body at the Massachusetts Institute of Technology, participated in the experiment on a voluntary basis. Ages ranged from 19 to 59, with a mean age of 30. Participants who self-rated themselves a "3" or less on a music training scale of "1" ("no training") to "5" ("professional") or who had less than 10 years of training in instrumental or vocal performance were categorized as "nonmusicians." The rest of the subjects (45%), who self-rated a "4" or higher on the music training scale and had at least 10 years of formal training were categorized as "musicians." Two of the 35 data sets were thrown out due to confusion or admitted error on the part of the subject.

Stimulus materials. Ten musical stimuli were used for the experiment. Six of these were short (10 s or less); two were carried over from Experiment 1 (A03 and A13, although the latter was rendered with strings timbre for Experiment 2 as opposed to only piano timbre for Experiment 1), and four of them were composed specifically for this study (Figure 9). The members of the latter group were similar in design to those from Experiment 1 and were intended to clarify certain parametric interactions that were not entirely clear from the results of the previous study. In addition to these shorter stimuli, there were four excerpts taken from the classical repertoire: the Bach-Vivaldi passage from Experiment 1 (Figure 4), Schoenberg Klavierstück, Op. 11 No. 12, mm. 1-8 (Figure 9), Beethoven Symphony No. 1, first movement, mm. 1-4 (Figure 9), and Brahms Piano Concerto No. 2, first movement, mm. 146-162 (Figure 10). The longer stimuli were 20 s to 1 min in length. All of the stimuli, both long and short, were rendered in MIDI and converted to audio (wav) format. Dynamics were added after the conversion to audio and were not added by changing MIDI velocity values.

Task. Test subjects were presented with a computer interface written in C++ with the Win32 API. Moving the mouse caused a slider bar to move up and down without requiring the subject to press the mouse button. This was done so that participants would not tire from constantly holding the mouse down or worry about an extra action that might distract from the listening experience. Subjects were instructed to raise the slider if they felt tension increasing, and to lower it if they felt tension lessening. The slider values were collected at a sample rate of 60 Hz. Each stimulus was played four times, and tension responses were recorded for all iterations. After listening and responding to an excerpt, subjects were asked to select a confidence value for their response. The playback of each trial was preceded by visual cues (numbers counting down from 3) that appeared on the interface to prepare the subject.

At the beginning of each trial, participants were asked to position the slider at a spot approximately one quarter of the way from the bottom (marked with arrows) and to try to use the space above that point for their judgments. This position was determined through a series of pilot studies that explored various initial slider positions. These tests revealed that subjects tended to run out of space at the bottom of the slider. Although this precaution was taken, recent work has shown that the initial position of a continuous input device ultimately does not make a difference (Hackworth & Fredrickson, 2010).

Results

Quantifying musical parameters. Since the ultimate goal was to define a tension model in terms of multiple musical parameters, all of these parameters had to be adequately described and quantified first before their contribution to the subject responses could be assessed. All of the parameters confirmed in Experiment 1 as well as one additional parameter, melodic expectation, were quantified for each stimulus in Experiment 2. None of the excerpts required all of the features to be quantified. For example, if there was no change in tempo throughout an excerpt, the graph representing it (a flat line) was not used in the subsequent analyses.

Some features were easy to quantify; for example, tempo, a one-dimensional parameter, was described in terms of beats per minute with respect to time. Harmony and melodic expectation, on the other hand, were complex enough to require more sophisticated methods of quantification. Even though there were some descriptive overlaps, tempo and onset frequency were treated separately and given individual graphs. Most stimuli did not require more than one or two pitch height graphs—melody and bass lines for the most part—although if there existed a prominent inner line, that was included as well. The individual pitch-height values were not connected by linear interpolation; since the x-axis of the graphs spanned the time of the excerpt, the values were extended for their respective durations, resulting in step functions. This graph format was the same for other features with values that changed at discrete time intervals.

As shown in Experiment 1, pitch height is an important factor in how listeners perceive tension. However, it is also a somewhat crude metric that does not take into

FIGURE 9. Stimuli from Experiment 2.

FIGURE 10. Excerpt 10 from Experiment 2: Brahms Piano Concerto No. 2, first movement, mm. 146-162.

account the tonal context or the deeply schematic expectations of melodic contour described in Narmour's Implication-Realization (I-R) model (Narmour, 1990, 1992). So in addition to pitch height, a graph of melodic expectation values was generated using Margulis' (2005) melodic expectation model, which takes into account the schematic expectations discussed in the I-R model and measures of proximity and distance between intervals. Stability, an additional factor in Margulis's model, is derived from Lerdahl's tonal pitch space and melodic attraction models, thus addressing tonal relationships as well as melodic interval relationships. For more details on how these values were calculated, see Appendix B.

Harmonic tension values for the tonal stimuli were quantified using Lerdahl's (2001) tonal tension model in the same manner as Experiment 1 (see Appendix A for details). The harmonic tension graph for the Schoenberg excerpt was calculated in a different way because it was atonal. Each chord was given a tension value based on the interval classes it contained; this was done by assigning each interval class (six total) a rank value corresponding to its relative psychoacoustic dissonance. The final tension value consisted of the sum of the values associated with the interval classes found in that chord.

The graphs for loudness were derived directly from the audio files. The values were produced by Jehan's (2005) psychoacoustic loudness model, which takes into account outer and inner ear filtering (more or less the equivalent of the Fletcher-Munson curves at an average pressure level), frequency warping into a cochlear-like frequency distribution, frequency masking, and temporal masking (Glasberg & Moore, 2002; Moore & Glasberg, 1995; Zwicker & Fastl, 1999). The graph values are measured in dB with a reference silence at -60 dB. A point was computed every 256 samples at 44,100 Hz with a window length of 4096 samples. The results were then filtered to obtain a smoother curve.

Data preprocessing. The subject data and feature graphs were resampled to 50Hz and then normalized to zero mean and unit standard deviation (*Z*-score). The feature graphs were slightly smoothed using a raised-cosine, moving-average filter to remove the sharp edges and discontinuities. For correlation purposes, an additional step was taken: the subject data and feature graphs were down-sampled again so that each point selected occurred in the middle of every beat or note event. Since these sampled points were evenly distributed throughout each excerpt, they did not correspond precisely to note events in cases where tempo was altered (Excerpts 2, 6, and 9). This downsampling was done to avoid a problem inherent in slider values sampled at a high rate: the points are not independent of each other given that any two adjacent values (at time $t = n$ and $t = n+1$) change very little, and independence is a requirement for statistics such as Pearson's *r*. A graph containing all of the feature analyses along with the mean tension response for Excerpt 3 (Beethoven) is shown in Figure 11 to provide an idea of what the processed data looked like.

Correlation and multiple regression analysis. The first step in the analysis process was to evaluate how each feature graph for each excerpt correlated with the subject data. Given that the correlation between beat-sampled subject responses was high, mean $r(1006) = .47$, mean $p = .003$ (Pearson's correlation), the mean of the responses across subjects was used for all subsequent analyses. Within-subject correlation for repeated listenings indicated that there was a marked difference between the first listening and the later listenings and high consistency between the last three (see Table 3). This is possibly due to the initial uncertainty of listening to something unfamiliar; by the second listening, subjects have some veridical knowledge of the music and can predict events and react with greater deliberation.

The next step was to understand how the individual musical features contributed to the tension judgments. Results of correlation analysis for the feature graphs and mean tension responses are shown in Tables 4 and 5, and the multiple regression results are shown in Table 6. Although the data was downsampled to minimize serial dependence, Spearman's ρ was computed in addition to Pearson's *r* as a comparative measure. Spearman's ρ was utilized in this case because it has been shown to be more accurate than Pearson's *r* for serial correlation (Schubert, 2002; Vines et al., 2006).

These correlation and regression results give an indication of how much each musical feature contributed to the perception of tension for individual excerpts. For Excerpts 1, 2, and 4, the feature with the highest correlation to tension was pitch height. In Excerpt 1, there were just two competing features: pitch height and harmony.

TABLE 3. Average Within-Subject Correlation Values for Repeated Listenings

	Trial 2		Trial 3		Trial 4	
Trial	Mean *r* (Pearson's)	Mean *p*	Mean *r* (Pearson's)	Mean *p*	Mean *r* (Pearson's)	Mean *p*
1	.68	.09	.62	.12	.61	.13
2	1	0	.82	.04	.79	.05
3	–	–	1	0	.85	.04
4					1	0

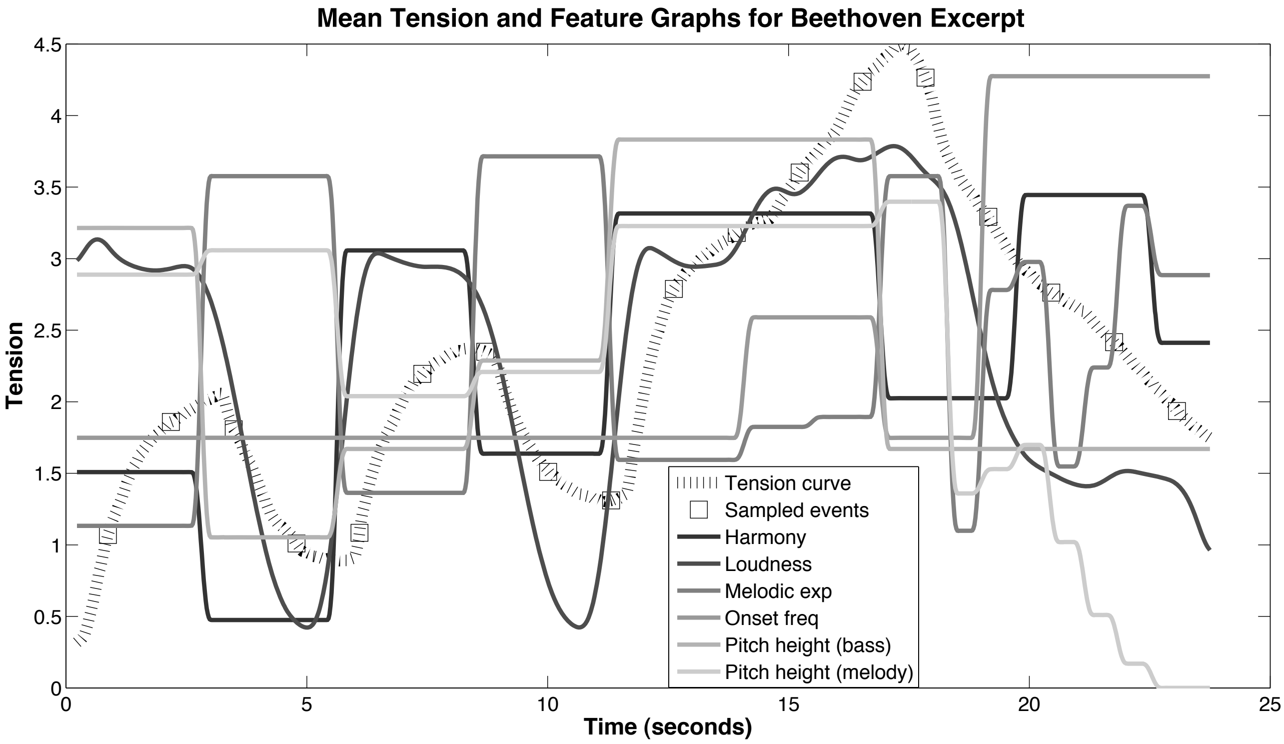


FIGURE 11. Graphs showing the musical features and mean tension response for Excerpt 3 (Beethoven) from Experiment 2.

Excerpt 1 was also used in Experiment 1 (A03), and the response profile from the previous study matches the results from Experiment 2. This appears to be the case for musicians versus nonmusicians as well; while pitch height was an important contributor to tension for both groups, musicians appeared to be more sensitive to harmony than nonmusicians. Excerpt 2 was also designed to examine two competing features, in this case pitch height and tempo, with tempo appearing to have little effect. In Excerpt 4, the effect of pitch height predominated over the decrease in dynamics. The results might have been the opposite if the directions were reversed: if dynamics were increasing while pitch height was decreasing, the tension response might have been increasing given that looming sounds elicit more attention than fading sounds (Granot & Eitan, 2011; Neuhoff, 1998, 2001). The same argument could be made for Excerpt 6, where increasing tempo correlated more highly with tension than decreasing dynamics.

Dynamics was the most significant contributor to tension in Excerpts 3 (Beethoven), 9 (cadence with ritard and crescendo), and 10 (Brahms). In the case of the Beethoven, dynamics was followed by harmony as the feature with the highest correlation. However, the beta values from the regression analysis indicate something rather different: all features except pitch height of the bass contributed roughly about the same, while harmony was slightly less influential. Excerpt 9 consisted of three features—harmony, pitch height of melody, and tempo—moving in a decreasing direction against dynamics moving in an increasing direction; the dynamics in this case "won," despite the fact that it was in opposition to three features combined. The effect of looming sounds on attention, mentioned previously in the context of Excerpts 4 and 6, might be the reason for this. For Excerpt 10 (Brahms), onset frequency was the only other positive contributor, although it was a far second to dynamics.

Harmony was the strongest contributing feature in Excerpts 5 (Bach-Vivaldi) and 7. In the Bach-Vivaldi, harmony was followed by pitch height of the inner line, distinctive particularly in the first part of the excerpt, and onset frequency, which increases suddenly to sixteenth notes in the latter half of the excerpt. In the case of Excerpt 7, there was only one real changing feature—harmonic dissonance—and not surprisingly, it was highly correlated with tension judgments.

Excerpt 8 (Schoenberg) had several features with high correlations: harmony, dynamics, onset frequency, pitch height of the inner voice, and pitch height of the bass. This can be explained by the fact that those features are highly correlated with each other, thus combining to

TABLE 4. Correlations Between Mean Tension Data and Musical Features For Each Excerpt in Experiment 2.

Excerpt 1: I-V^7-I progression with rising melody, *df* = 4				
	Nonmusicians		Musicians	
	Pearson's *r*	Spearman's *ρ*	Pearson's *r*	Spearman's *ρ*
Harmony	.28	.43	.53	.71
Pitch height - Melody	.94**	.89*	.79	.71

Excerpt 2: Chromatic scale up with ritard, *df* = 10				
	Nonmusicians		Musicians	
	Pearson's *r*	Spearman's *ρ*	Pearson's *r*	Spearman's *ρ*
Pitch height - Melody	1.00**	1.00**	.99**	1.00**
Tempo	-.99**	-1.00**	-.98**	-1.00**

Excerpt 3: Beethoven Symphony No. 1, *df* = 16				
	Nonmusicians		Musicians	
	Pearson's *r*	Spearman's *ρ*	Pearson's *r*	Spearman's *ρ*
Harmony	.43	.51*	.48*	.57*
Dynamics	.62**	.59*	.55*	.53*
Melodic expectation	.01	.05	-.001	.04
Onset frequency	.19	.36	.35	.43
Pitch height - Bass	.34	.46	.24	.35
Pitch height - Melody	.24	.42	.10	.26

Excerpt 4: Chromatic scale up with decrescendo, *df* = 10				
	Nonmusicians		Musicians	
	Pearson's *r*	Spearman's *ρ*	Pearson's *r*	Spearman's *ρ*
Dynamics	-.55	-.80**	-.24	-.36
Pitch height - Melody	.94**	1.00**	.77**	.64*

Excerpt 5: Bach-Vivaldi concerto, *df* = 93				
	Nonmusicians		Musicians	
	Pearson's *r*	Spearman's *ρ*	Pearson's *r*	Spearman's *ρ*
Harmony	.63**	.68**	.45**	.36**
Dynamics	-.0002	.10	.02	.06
Melodic expectation	-.29**	-.25*	-.17	-.10
Onset frequency	.36**	.33**	.27**	.21*
Pitch height - Bass	-.30**	-.38**	-.06	.02
Pitch height - Inner	.54**	.05	.45**	.25*
Pitch height - Melody	-.28**	-.28**	-.11	-.03

(Continued)

TABLE 4. Continued.

Excerpt 6: Accelerando plus diminuendo, *df* = 4				
	Nonmusicians		Musicians	
	Pearson's *r*	Spearman's *ρ*	Pearson's *r*	Spearman's *ρ*
Dynamics	-.09	-.37	-.04	.03
Tempo	.78	.83	.76	.60

Excerpt 7: Increasingly dissonant harmony, *df* = 4				
	Nonmusicians		Musicians	
	Pearson's *r*	Spearman's *ρ*	Pearson's *r*	Spearman's *ρ*
Harmony	.95**	1.00**	.96**	1.00**

Excerpt 8: Schoenberg, *df* = 21				
	Nonmusicians		Musicians	
	Pearson's *r*	Spearman's *ρ*	Pearson's *r*	Spearman's *ρ*
Harmony	.55**	.16	.46*	.19
Dynamics	.69**	.73**	.64**	.62**
Onset frequency	.69**	.69**	.71**	.73**
Pitch height - Bass	.76**	.37	.65**	.41
Pitch height - Inner	.74**	.41*	.63**	.36
Pitch height - Melody	-.41	-.14	-.29	-.10

Excerpt 9: Resolving progression with ritard and crescendo, *df* = 6				
	Nonmusicians		Musicians	
	Pearson's *r*	Spearman's *ρ*	Pearson's *r*	Spearman's *ρ*
Harmony	-.56	-.55	-.73*	-.71
Dynamics	.96**	.79*	.93**	.62
Pitch height - Melody	-.66	-.55	-.79*	-.71
Tempo	-.69	-.40	-.82*	-.62

Excerpt 10: Brahms, *df* = 64				
	Nonmusicians		Musicians	
	Pearson's *r*	Spearman's *ρ*	Pearson's *r*	Spearman's *ρ*
Harmony	-.23	-.23	-.32**	-.33**
Dynamics	.77**	.73**	.82**	.73**
Melodic expectation	.02	.01	.06	.01
Onset frequency	.23	.37**	.25*	.35**
Pitch height - Bass	.29*	.22	.22	.11
Pitch height - Melody	-.17	-.27*	-.29*	-.35**

Note: $^{*}p \le .05$ $^{**}p \le .01$

create strong cues for tension changes. On the other hand, beta values from the regression analysis are largest for pitch height of the bass and are considerably smaller for the other features.

Although we can get an idea of the relative importance of musical features for each excerpt, ascertaining a more global perspective requires some understanding of context. In excerpts where a certain feature had a

TABLE 5. Correlations Between Musical Features For All Excerpts in Experiment 2.

		Excerpt 1: I-V7-I progression with rising melody, *df* = 4						
		D	M	O	PH-B	PH-I	PH-M	TM
H	Pearson	–	–	–	–	–	.13	–
	Spearman	–	–	–	–	–	.54	–
PH-M	Pearson	–	–	–	–	–	–	–
	Spearman	–	–	–	–	–	–	–

		Excerpt 2: Chromatic scale up with ritard, *df* = 10						
		D	M	O	PH-B	PH-I	PH-M	TM
PH-M	Pearson	–	–	–	–	–	–	-1.00**
	Spearman	–	–	–	–	–	–	–
TM	Pearson	–	–	–	–	–	–	–
	Spearman	–	–	–	–	–	–	–

		Excerpt 3: Beethoven Symphony No. 1, *df* = 16						
		D	M	O	PH-B	PH-I	PH-M	TM
H	Pearson	.34	-.56*	.35	.44	–	-.22	–
	Spearman	.25	-.30	.52*	.44	–	-.06	–
D	Pearson	–	-.51*	-.29	.54*	–	.48*	–
	Spearman	–	-.52*	-.09	.57*	–	.64**	–
M	Pearson	–	–	.05	-.56*	–	-.11	–
	Spearman	–	–	.08	-.37	–	-.05	–
O	Pearson	–	–	–	-.28	–	-.80**	–
	Spearman	–	–	–	.11	–	-.23	–
PH-B	Pearson	–	–	–	–	–	.50*	–
	Spearman	–	–	–	–	–	.48*	–
PH-S	Pearson	–	–	–	–	–	–	–
	Spearman	–	–	–	–	–	–	–

		Excerpt 4: Chromatic scale up with decrescendo, *df* = 10						
		D	M	O	PH-B	PH-I	PH-M	TM
D	Pearson	–	–	–	–	–	-.79**	–
	Spearman	–	–	–	–	–	-.79**	–
PH-M	Pearson	–	–	–	–	–	–	–
	Spearman	–	–	–	–	–	–	–

		Excerpt 5: Bach-Vivaldi concerto, *df* = 93						
		D	M	O	PH-B	PH-I	PH-M	TM
H	Pearson	-.13	-.50**	.54**	-.49**	.36**	-.62**	–
	Spearman	-.08	-.54**	.50**	-.53**	-.04	-.65**	–
D	Pearson	–	.31**	-.31**	-.05	.19	.28**	–
	Spearman	–	.28**	-.23*	-.12	.24*	.31**	–
M	Pearson	–	–	-.26**	.30**	-.23*	-.52**	–
	Spearman	–	–	-.20	.28**	.002	-.53**	–
O	Pearson	–	–	–	-.23*	.20*	-.67**	–
	Spearman	–	–	–	-.17	-.02	-.56**	–
PH-B	Pearson	–	–	–	–	-.46**	.48**	–
	Spearman	–	–	–	–	-.01	.48**	–

(Continued)

TABLE 5. Continued.

		D	M	O	PH-B	PH-I	PH-M	TM
PH-I	Pearson	–	–	–	–	–	-.43**	–
	Spearman	–	–	–	–	–	-.08	–
PH-M	Pearson	–	–	–	–	–	–	–
	Spearman	–	–	–	–	–	–	–
		Excerpt 6: Accelerando plus diminuendo, *df* = 4						
		D	M	O	PH-B	PH-I	PH-M	TM
D	Pearson	–	–	–	–	–	–	-.67
	Spearman	–	–	–	–	–	–	-.60
TM	Pearson	–	–	–	–	–	–	–
	Spearman	–	–	–	–	–	–	–
		Excerpt 7: Increasingly dissonant harmony, *df* = 4						
		D	M	O	PH-B	PH-I	PH-M	TM
H	Pearson	–	–	–	–	–	–	–
	Spearman	–	–	–	–	–	–	–
		Excerpt 8: Schoenberg, *df* = 21						
		D	M	O	PH-B	PH-I	PH-M	TM
H	Pearson	.14	–	.18	.85**	.86**	-.42*	–
	Spearman	-.09	–	.03	.14	.35	-.07	–
D	Pearson	–	–	.74**	.37	.37	-.09	–
	Spearman	–	–	.53**	.18	.31	-.11	–
O	Pearson	–	–	–	.38	.36	-.01	–
	Spearman	–	–	–	.14	.06	.05	–
PH-B	Pearson	–	–	–	–	.99**	-.70**	–
	Spearman	–	–	–	–	.37	-.62**	–
PH-I	Pearson	–	–	–	–	–	-.68**	–
	Spearman	–	–	–	–	–	-.44*	–
PH-M	Pearson	–	–	–	–	–	–	–
	Spearman	–	–	–	–	–	–	–
		Excerpt 9: Resolving chord progression with ritard and crescendo, *df* = 6						
		D	M	O	PH-B	PH-I	PH-M	TM
H	Pearson	-.47	–	–	–	–	.98**	.89**
	Spearman	-.48	–	–	–	–	1.00**	.86*
D	Pearson	–	–	–	–	–	-.57	-.70*
	Spearman	–	–	–	–	–	-.48	-.62
PH-M	Pearson	–	–	–	–	–	–	.91**
	Spearman	–	–	–	–	–	–	.86*
TM	Pearson	–	–	–	–	–	–	–
	Spearman	–	–	–	–	–	–	–
		Excerpt 10: Brahms, *df* = 64						
		D	M	O	PH-B	PH-I	PH-M	TM
H	Pearson	-.23	-.11	-.08	.16	–	.37**	–
	Spearman	-.24*	-.09	-.17	.12	–	.39**	–
D	Pearson	–	.33**	.25*	.26**	–	-.35**	–
	Spearman	–	.33**	.21	.19	–	-.42**	–

(Continued)

TABLE 5. Continued.

M	Pearson	–	–	.10	.11	–	-.18	–
	Spearman	–	–	.04	.08	–	-.20	–
O	Pearson	–	–	–	.34**	–	-.20	–
	Spearman	–	–	–	.08	–	-.33**	–
PH-B	Pearson	–	–	–	–	–	.10	–
	Spearman	–	–	–	–	–	.03	–
PH-M	Pearson	–	–	–	–	–	–	–
	Spearman	–	–	–	–	–	–	–

Note: H = harmony, D = dynamics, M = melodic expectancy, O = onset frequency, PH-B = pitch height of bass, PH-I = pitch height of inner voice, PH-M = pitch height of melody, TM = tempo. $*p \leq .05$ $**p \leq .01$

TABLE 6. Multiple Regression Results For All Excerpts in Experiment 2.

MODEL		β	F	df	R^2	R^2_{adj}
Excerpt 1		–	10.47*	4	.88	.79
(I-V^7-I with rising melody)	Harmony	0.27				
	Pitch height - Melody	0.83*				
Excerpt 2		–	587.6**	10	.99	.99
(Chromatic scale up with ritard)	Pitch height - Melody	1.30**				
	Tempo	0.31				
Excerpt 3		–	20.65**	16	.92	.87
(Beethoven)	Harmony	0.47**				
	Dynamics	0.72**				
	Melodic expectation	0.68**				
	Onset frequency	0.76**				
	Pitch height - Bass	-0.02				
	Pitch height - Melody	0.65**				
Excerpt 4		–	188.03**	10	.98	.97
(Chromatic scale up with decrescendo)	Dynamics	0.73**				
	Pitch height - Melody	1.49**				
Excerpt 5		–	15.95**	93	.56	.53
(Bach-Vivaldi)	Harmony	0.62**				
	Dynamics	-0.07				
	Melodic expectation	-0.08				
	Onset frequency	0.26*				
	Pitch height - Bass	0.20*				
	Pitch height - Inner	0.55**				
	Pitch height - Melody	0.52**				
Excerpt 6		–	50.68**	4	.97	.95
(Accelerando with diminuendo)	Harmony					
	Dynamics	0.77**				
	Tempo	1.26**				
Excerpt 7		–	44.16**	4	.92	.90
(Increasingly dissonant harmony)	Harmony	0.85**				
Excerpt 8		–	7.79**	21	.75	.65
(Schoenberg)	Harmony	-0.11				
	Dynamics	0.22				
	Onset frequency	0.32				
	Pitch height - Bass	1.03				
	Pitch height - Inner	-0.35				
	Pitch height - Melody	0.12				

(Continued)

TABLE 6. Continued.

MODEL		β	F	df	R^2	R^2_{adj}
Excerpt 9		–	92.24**	6	.99	.98
(Resolving chord progression with ritard and crescendo)	Harmony	-0.52				
	Dynamics	1.07**				
	Pitch height - Melody	-0.08				
	Tempo	0.50				
Excerpt 10		–	24.31**	64	.71	.68
(Brahms)	Harmony	-0.14				
	Dynamics	0.86**				
	Melodic expectation	-0.41**				
	Onset frequency	0.04				
	Pitch height - Bass	0.07				
	Pitch height - Melody	0.08				

Note: $*p \leq .05$ $**p \leq .01$

clear trend rather than subtle fluctuations, the correlation values were higher, regardless of the inherent listener sensitivity to that feature. For example, the loudness graph does not correlate at all with the subject data for the Bach-Vivaldi while it does so significantly for the Beethoven. This coincides with the qualitative observation that there are almost unnoticeable fluctuations in dynamics in the Bach-Vivaldi but very obvious changes in the Beethoven. In general, the results reflect the salient features of each excerpt; likewise, in the case of the Brahms, loudness was the strongest factor contributing to tension. This can be explained by the fact that the melodic, harmonic, and rhythmic aspects are quite complex and change frequently in an irregular manner.

Linear and nonlinear predictive models. The next step was to explore the process of building a predictive model of tension by attempting to fit linear and nonlinear models to a portion of the empirical data given the musical features, and then trying to predict the remaining data using the resulting model. The assumption was that tension could be expressed as time-varying function of a set of musical parameters. The goal was to approximate this function so that it matched the subject data as accurately as possible. Each model was trained on the first portion of the data for each excerpt and then tested on the remaining part in order to evaluate the resulting model's analytical usefulness.

In preparation, the feature graphs were further smoothed in order to eliminate any sharp edges and discontinuities. This was necessary because the empirical data itself was smooth due to the relatively gradual slider motions used by the listeners to indicate tension changes. It would be difficult for a model—particularly a linear model—to estimate smooth data with graphs containing abrupt changes.

The prediction results for both the linear and quadratic models were quite poor, even when more than three-quarters of the empirical data was used to train the models. In most cases, the linear models performed better than the quadratic models. While the quadratic models were able to fit the training data more accurately, the prediction results were often highly inaccurate due to overfitting. The results for the Brahms excerpt, shown in Figure 12, is a typical example. The problem here is a common one: given enough degrees of freedom (or in this case, enough features), it is possible to model anything, as is indicated by how well the quadratic model fit the training data for a complex excerpt like the Brahms. The question remains whether what has been modeled is the relevant data or the noise. In this case, the answer appears to be that little of what was actually captured in those models was musically relevant. Another problem, especially for the short excerpts, was that the training data were insufficient to capture the musical dynamics necessary to make accurate tension predictions. It is possible that given more diverse and longer excerpts, predictions might be more accurate. Nonetheless, a very different approach was subsequently adopted due to the poor results.

A parametric, temporal model for predicting tension. The primary problem with the method just described was the fact that the models were unable to capture the moment-to-moment experience of music listening. Listening is a dynamic process, and trying to build a successful model that did not take this into account was perhaps, in hindsight, an impossible task. The next method took into account the real-time nature of music perception. This new approach included three major components:

1. An *attentional window* that models a perceptual moving window in time and extracts a current tension trend.

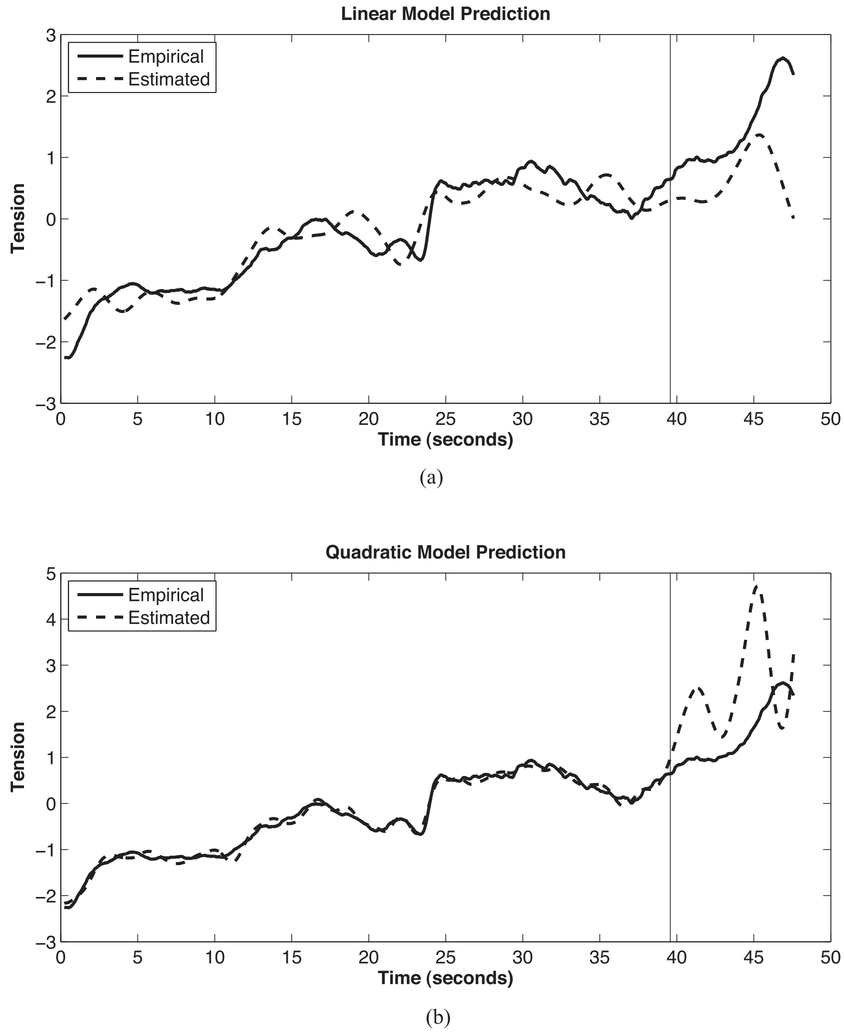


FIGURE 12. The results of (a) linear regression and (b) polynomial regression for Excerpt 10 (Brahms) from Experiment 2. The vertical line indicates the division between training data for the model and out of sample data used for prediction.

2. A *memory window*, defined simply as the direction of the tension trend in a window of time immediately preceding the attentional window.
3. Differing weights for the influence of various musical parameters.

The model essentially defines tension as a function of musical features by estimating the combined directional change of all of the features in each attentional window and using this information to generate tension predictions. The combined directional change, or *tension trend,* for an attentional window is weighted both by what immediately precedes it in the memory window and the perceptual weights of the individual musical parameters. The influence of the memory window is dependent on whether its trend matches the directionality of the attentional window's trend or not; if they match, the magnitude of the tension trend is additionally increased. Tension trends for each attentional window are then integrated over time to generate a final tension prediction.

The musical feature graphs serving as inputs to the prediction model are identical to the ones used for the correlation and regression analysis; the only feature that was not included was melodic expectation, because it did not correlate in a clear directional way with tension. The slope of an individual feature over a particular time interval is defined as the slope of the best linear fit of that section of the feature graph. This function is defined as $s_f(t)$, the slope of the interval from t to $t + d$ of feature f, where t is time, and d is the window duration; d is a multiple of the sampling interval (in this case, 0.02 s for

50 Hz). The slope of a tension trend is determined by summing all of the weighted slopes of the individual feature graphs over that discrete time window. If all features have positive slopes, the sum of those slopes would indicate a clear increase in tension for that window. However, if the slopes conflict, they would to some degree cancel each other out. The relative weights for each feature were estimated from the results of Experiment 1: dynamics and pitch height of the melody were given the highest weight (weight = 3) followed by tempo and onset frequency (weight = 2), and finally harmony, pitch height of bass, and pitch height of inner lines (weight = 1). These values were then normalized to sum to one.

Qualitative observation of the data from Experiments 1 and 2 indicated that listeners were drawn to salient trends in tension, and that the magnitude of the tension judgments increased as the trends progressed. In other words, a clear trend generated the expectancy that the trend would continue in the current direction. The model implements this observation by multiplying the magnitude of the slope of a tension trend by the memory constant β, if the cumulative slope of the immediately preceding memory window matches the direction (increase/decrease) of the trend in the attentional window. The cumulative slope of a tension trend at time t is thus defined as

$$s'(t) = \beta \sum_f w_f s_f(t) \tag{1}$$

where $s_f(t)$ is the slope of best linear fit of feature f at time t, as described above; $\beta = 1$ if the sign of $s(t-d)$ does not equal the sign of $s(t)$; β is some positive value, empirically determined, if the sign of $s(t-d)$ is the same as the sign of $s(t)$; w_f is the weight of feature f with $\sum_f w_f = 1$. The ideal β value for the case where the attentional window trend continues in the same direction as the memory window trend was determined to be 5, given the empirical data from Experiment 2. Significantly larger values than 5, up to 100, resulted in predictions that were also highly correlated to the data, but lacking in detail; values less than 5 resulted in much poorer results.

In order to model moment-to-moment integration of auditory perception, the tension trends for attentional windows are evaluated every step size h = 0.25 s. Although in reality cognitive processing is continuous, discrete evaluation of these windows at the resolution of 250 ms was deemed sufficient given that it allowed for sub-beat time resolution; h also had to divide d evenly. The attentional window trends evaluated every 250 ms are averaged with previous trends, resulting in recent windows weighted more strongly to simulate memory decay. Thus, the slope of the tension curve at time t is defined as

$$S(t) = \sum_{\tau=0}^{\frac{d}{h}-1} s'(t - \tau h) k_\tau \tag{2}$$

where d is the attentional window duration and k is a decay constant for a moving average filter with $\sum_i k_i = 1$.

Given this equation for the slope of the predicted tension curve, the actual tension value F_{ten} at time t is defined as

$$F_{ten}(t) = S(T)(t - hT) + h \sum_{i=0}^{T-1} S(i) \tag{3}$$

where h is the step size of the moving window increment (0.25 s) and $T = \left\lfloor \frac{t}{h} \right\rfloor$. Note that this equation simplifies to

$$F_{ten}(t) = h \sum_{i=0}^{T-1} S(i) \tag{4}$$

when t is a multiple of h. This falls out of the fact that time is continuous, but the model calculates the tension values in discrete segments.

One final addition to the model was not a function of the features. Observation of the data made it clear that the initial tension change for all responses, regardless of the excerpt, was an upward sweep of the slider. This might be due in part to the nature of the interface or to the fact that going from silence to some sound always results in a natural, sudden increase in tension. This slider movement was accounted for by adding a uniform upward tension motion at the beginning of each prediction curve.

Predictions generated with attentional windows ranging from d = 1-8 s and memory windows ranging from 0-8 s were tested with the data from Experiment 2 in order to determine which values provided the most accurate predictions. Accuracy of predictions was evaluated by correlating the empirical judgments with the predicted tension curve. As shown in Figure 13, a 3 s attentional window resulted in the most accurate tension predictions overall, although some of the excerpts were too short to be included in the data for the larger window sizes. To get a better picture of the effect of the memory windows, the results were broken down by individual memory window sizes for the Brahms and Bach-Vivaldi excerpts, the only two excerpts long enough to test window durations up to 8 s. These results, shown in Figure 14, indicate that the

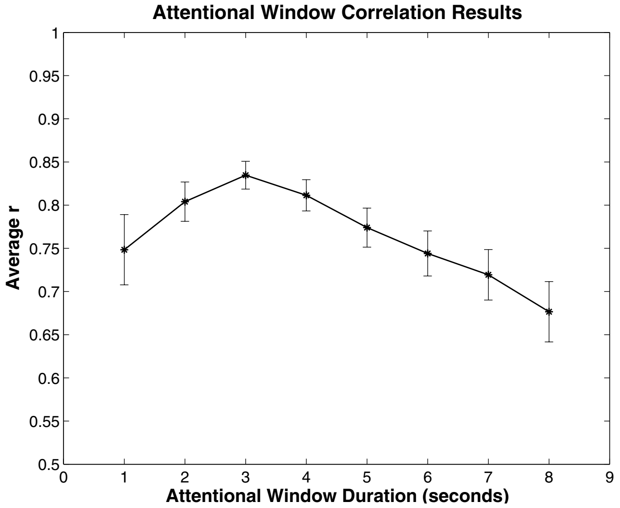


FIGURE 13. Average correlation values (Pearson's *r*) for tension prediction across all excerpts for different attentional window sizes. Error bars indicate standard error.

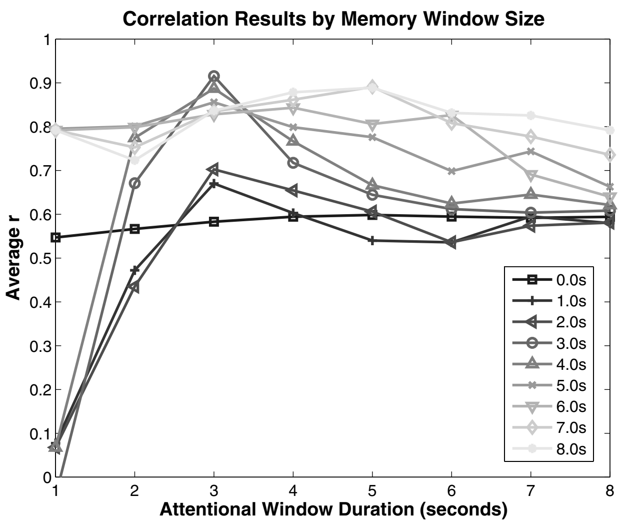


FIGURE 14. Correlation results (Pearson's *r*) by memory window size for Excerpts 5 (Bach-Vivaldi) and 10 (Brahms) combined.

optimal memory window size is also 3 s. Note the sharp difference in results between having no memory window (graph depicting 0.0 s) and a memory window of any size. When there is no memory window at all, the accuracy of the model differs very little regardless of the size of the attentional window. The prediction results for the Bach-Vivaldi and Brahms excerpts with attentional and memory window sizes of 3 s are shown in Figure 15. Results of correlation between the model outputs and the empirical data are shown for all stimuli in Table 7. Attentional and memory window durations of 3 s were used for eight out of the ten excerpts that could accommodate them (i.e., have total durations of 6 s or longer), and slightly shorter 2 s window sizes were used for the two shortest excerpts (Excerpts 4 and 6). All of the correlations are high; the only result that is not significant is Spearman's ρ for Excerpt 4.

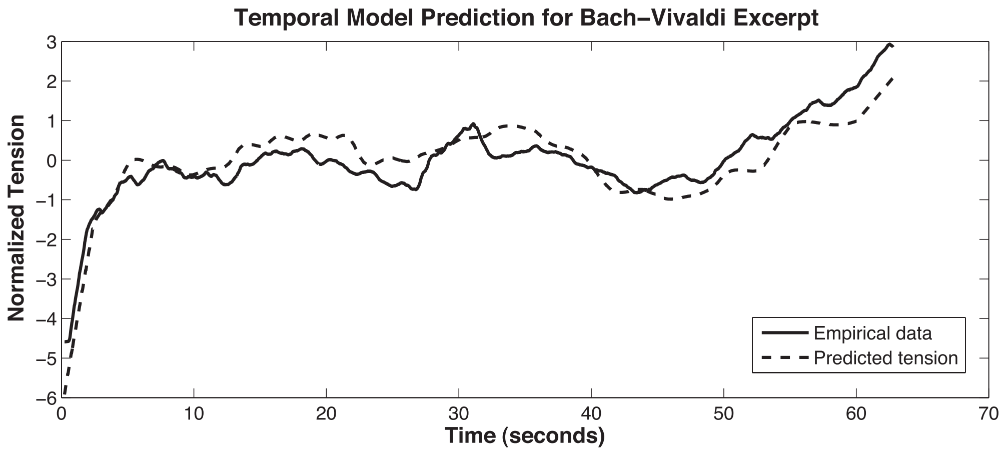
Discussion

The first part of this paper described a web-based experiment that presented listeners with short, simply constructed stimuli that attempted to isolate and combine changes in different musical features that hypothetically contributed to tension. These features included dynamics, tempo, onset frequency, pitch height, harmony, rhythmic regularity, and meter. There were several conclusions reached from the results of Experiment 1:

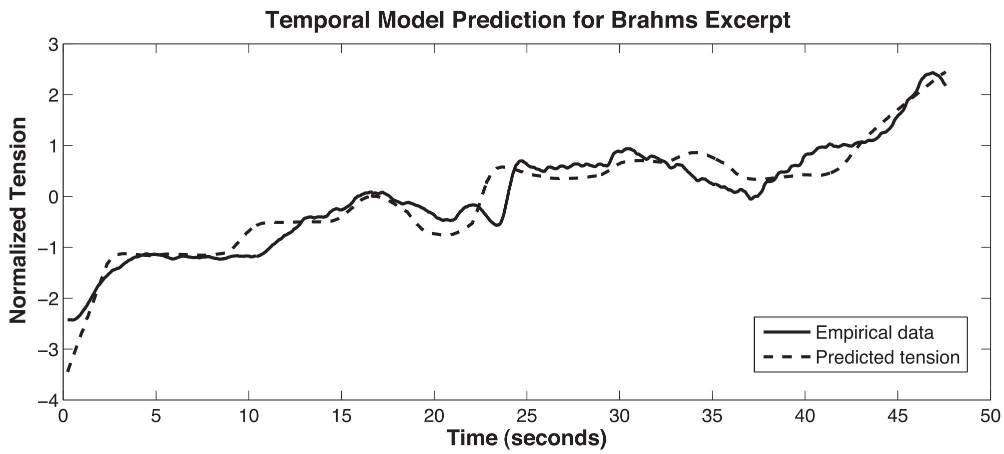
1. All parameters examined with the exception of rhythmic features had an impact on listeners' perception of tension.
2. In isolation, dynamics and pitch height had a clearer effect on tension than other parameters.
3. Musicians responded more strongly than nonmusicians to all features with the exception of pitch height in tonal contexts, where musicians responded more to harmony changes while nonmusicians responded more to pitch height changes.
4. In the case of more complex stimuli, where features counteracted each other, the results were often ambiguous.
5. When multiple features were combined in parallel, they considerably strengthened the feeling of tension changing in a particular direction.

Conclusion 1 is a new finding, since rhythmic regularity, meter, syncopation, and onset frequency have not been explored before. High sensitivity to dynamics (Conclusion 2) is consistent with previous findings (Burnsed & Sochinski, 1998-2001; Granot & Eitan, 2011; Ilie & Thompson, 2006; Krumhansl, 1996; Misenhelter, 2001; Nielsen 1987), but the considerable influence of pitch height is more consistent with results of Bigand et al. (1996) than Eitan and Granot (2011). Conclusion 3 also confirms Bigand et al. (1996) and Bigand and Parncutt's (1999) findings that musicians are more sensitive to tonal context than nonmusicians. Conclusions 4 and 5 align with the observations of Eitan and Granot that "noncongruence" does not result in an increase of tension and that there are interactions among parameters.

The web-based, curve selection methodology employed in Experiment 1 does not have a precedent in previous



(a)



(b)

FIGURE 15. Predicted tension for (a) Excerpt 5, Bach-Vivaldi and (b) Excerpt 10, Brahms produced using a 3 s attentional window and a 3 s memory window.

TABLE 7. Correlations For the Empirical Data From Experiment 2 and Tension Curves Predicted by the Temporal Model

	Pearson's r		Spearman's ρ	
	r	p	ρ	p
Excerpt 1: I-V^7-I progression with rising melody (A03), $df = 4$	.93	.008	.83	.058
Excerpt 2: Chromatic scale up with ritard, $df = 10$	.86	< .001	1	< .001
Excerpt 3: Beethoven Symphony No. 1, $df = 16$	.67	.002	.65	.004
Excerpt 4: Chromatic scale up with decrescendo, $df = 10$	.99	< .001	.89	< .001
Excerpt 5: Bach-Vivaldi concerto, $df = 93$	.88	< .001	.81	< .001
Excerpt 6: Accelerando plus diminuendo, $df = 4$	.90	.015	.60	.24
Excerpt 7: Increasingly dissonant harmony (A13), $df = 4$	.92	.009	1	.003
Excerpt 8: Schoenberg, $df = 21$	.93	< .001	.85	< .001
Excerpt 9: Resolving progression with ritard and crescendo, $df = 6$	.98	< .001	.93	.002
Excerpt 10: Brahms, $df = 64$	.95	< .001	.92	< .001

studies on musical tension. The graphical nature of the response choices was in part intended to present the user with an intuitive depiction of changing tension within a short time span. In addition to the artificially constructed stimuli, short segments taken from an excerpt from a Bach-Vivaldi concerto were added to include more complex, ecologically valid stimuli. The use of the Bach-Vivaldi excerpt in both Experiments 1 and 2 provided a means to compare the web-based methodology of Experiment 1 with the continuous response methodology from Experiment 2. This comparison entailed the reconstruction of a continuous curve from the discrete responses to the short, two-bar Bach-Vivaldi stimuli from Experiment 1, then comparing this reconstructed curve to the analogous continuous response from Experiment 2. Since the Bach-Vivaldi excerpt in Experiment 2 was rendered in a MIDI orchestral strings timbre, only responses to the strings version of the Bach-Vivaldi in Experiment 1 were considered in the analysis.

The largest connected series of two-bar excerpts from Experiment 1 encompassed mm. 91-104 of the excerpt, thus this particular section of the excerpt was utilized for the comparison. There was a mean of 52 responses per two-bar segment rendered in strings timbre ($SD = 9.8$). Three versions of the reconstructed curve were pieced together: the first consisted of the sum of all of the curve choices (excluding the nonspecific "changes frequently" and "none of the above" responses), each weighted by response frequency; the second reconstruction was identical to the first *except* flat responses were not included; the third reconstruction consisted of *only* the most frequently chosen curve for each two-bar segment, also weighted by response frequency.

In the first method, each two-measure segment was reconstructed as a sum of the response curves,

$$\frac{1}{100 - w_1} \sum_{c=2}^{7} w_c f_c \qquad (5)$$

where f_c is one of seven possible response choices (shown in Figure 1a-g) and w_c is the percentage of subjects who selected that particular curve. The multiplier $\frac{1}{100 - w_1}$ represents the "no change" responses by essentially smoothing the summed curves by attenuating the amplitude of the contours. The formula in the second reconstruction method was identical to the first with the exception of the excluded multiplier term for the "no change" responses:

$$\sum_{c=2}^{7} w_c f_c \qquad (6)$$

The third method entailed choosing only the curve with the highest response rate for each two-bar segment, represented by a single term $w_{max} f_{max}$. The final reconstituted curves were created by joining together all of the two-bar reconstructions in the correct order; the curves generated with the three methods are shown in Figure 16 along with the mean continuous response from Experiment 2. The three reconstructed graphs were sampled twice every measure.

Correlations between the continuous graph from Experiment 2 and the reconstructions were high: for the graph consisting of weighted curves with the "no change" responses included, $r(26) = .66$, $p < .001$ (Pearson's correlation); for weighted curves without the "no change" responses, $r(26) = .73$, $p < .001$; for only the most frequently chosen curves, $r(26) = .64$, $p < .001$. While the correlation results indicate that the data from Experiment 1 are similar to the data from Experiment 2, despite significant differences in methodology, there do appear to be some discrepancies between the reconstructed graphs (all three of which are quite similar), and the Experiment 2 continuous data. One prominent example is the sharp drop in tension in m. 93 in the Experiment 2 data that is not evident in the Experiment 1 data. This difference can be explained by the fact that the A minor chord on the downbeat of m. 93 is the resolution of a cadential figure. Since the stimuli are only presented in two-bar segments in Experiment 1, there can be no connection made between the preceding chords in m. 92 and the downbeat of m. 93. This reveals a natural limitation of the methodology used in Experiment 1: the fact that musical context cannot be adequately represented by joining temporally unrelated segments together. However, for judging short, independent musical excerpts, this problem does not apply. The problem inherent in stringing together responses to individual segments taken out of musical context is directly related to the importance of the memory window in the temporal, predictive model derived from the results of Experiment 2. Encoding past events that are no longer present in an attentional window, but still exist in some abstract form in working memory, is a crucial aspect of the listening experience. In summary, although the results of Experiment 1 provided a means for directly comparing the influence of important musical parameters on the perception of tension, the methodology did not inherently allow for a real-time accounting of parameter interaction. The continuous data from Experiment 2, on the other hand, did provide the means for this type of analysis.

Standard correlation and regression analysis of the data from Experiment 2 did not yield any new insights into parametric interaction. They did offer some idea of the

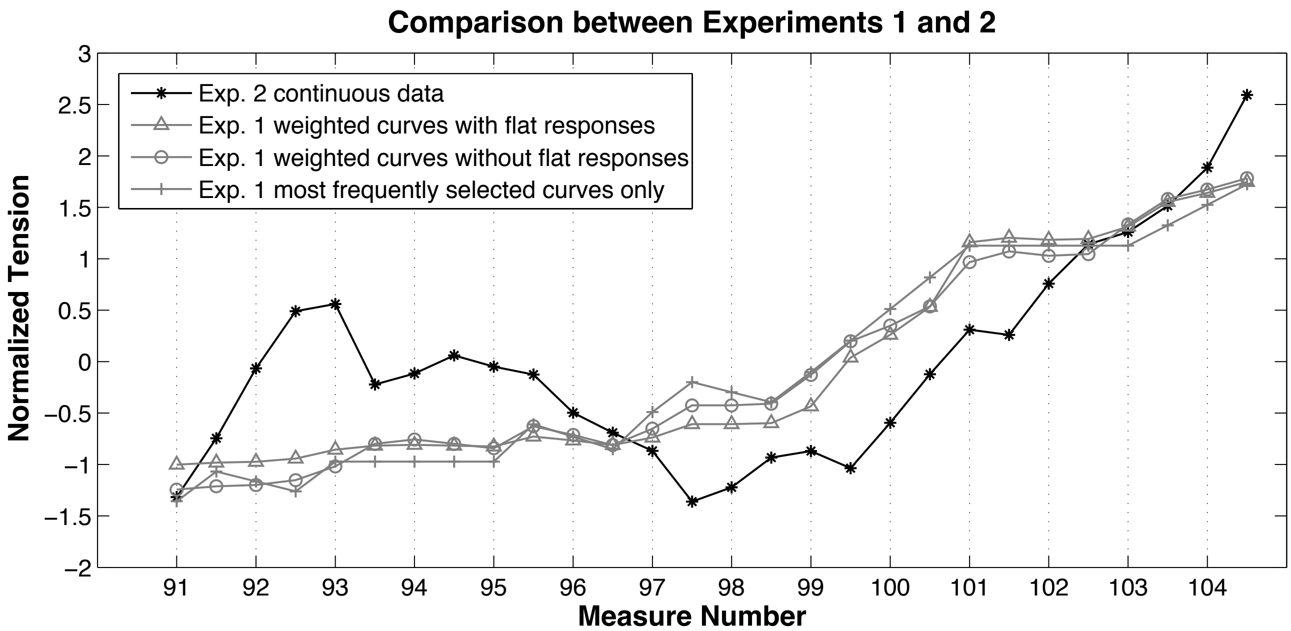


FIGURE 16. Curves reconstructed from Experiment 1 responses to mm. 91-104 of the Bach-Vivaldi excerpt graphed alongside the mean response from Experiment 2.

contribution of individual parameters in a given excerpt, but did not provide a way to determine how tension is derived from changes in those parameters. The first step taken toward building a predictive model involved using part of the empirical data to train linear and quadratic models that were then tested on the remaining portion of the data. The results were poor due to several reasons: there were not enough training data, particularly in the short excerpts, to come up with a viable predictive model; the polynomial models resulted in overfitting of the training data; and most significantly, the whole idea of building a model without regard to the cognitive implications of real-time listening was problematic to begin with.

Given these observations, a temporal, parametric model was conceived based on the concept of a perceptual moving window in time. The model consisted of a moving attentional window in which a tension trend was extracted over the course of a fixed time span. This trend, generally speaking, was defined as the sum of the weighted slopes of all the concurrent musical features. The magnitude of the trend was further influenced by the slope of the immediately preceding memory window; the incline or decline of the slope was increased if the direction of the current trend continued the previous trend of the memory window. Both attentional and memory window durations of 3 s were found to best fit the empirical data from Experiment 2.

Although the two window types have distinct designations, they are in a sense *both* memory windows. The "memory" window represents the contents of echoic memory while the "attentional" window represents encoding of previous musical processing in working memory. While it is difficult to determine a fixed length for short-term memory store, the combined attentional and memory window durations do fall in that range.

The optimal window sizes and the value of β, the trend multiplier, were determined empirically by finding the highest correlation to the Experiment 2 data. However, correlation as a measure of accuracy has certain problems. For example, high correlations can be obtained for predictions that are very smooth and lack detail. This was the case for predictions generated through the use of larger window sizes. Although the correlation values for these larger windows were on average lower than those for the 3 s windows, they were still quite high. Very short window sizes appeared to reflect the local details of the tension changes more accurately in some cases, but failed on the longer excerpts since they were unable to predict the larger trends adequately. One such example is the Beethoven excerpt, which was perhaps just short enough (24 s) for large-scale trends to be less of an issue. The 3 s attentional and memory windows resulted in high correlations with the empirical data, but the predicted curves were lacking in contour detail. The 0.5 s attentional and memory windows (Figure 17) yielded a tension prediction that better matched the detailed changes in the empirical data, but might not have tracked large-scale changes in tension had the excerpt been any

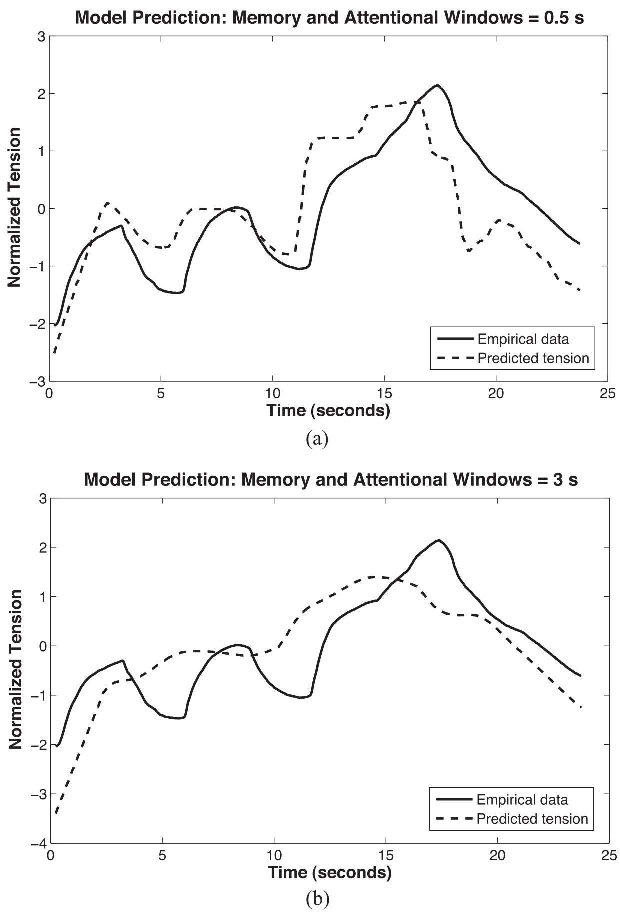


FIGURE 17. Predicted tension for Excerpt 3 (Beethoven) produced using (a) 0.5 s attentional and memory windows and (b) 3 s attentional and memory windows.

longer. It is also possible that shorter window durations worked better in this case because the lengthy rests in the first part of the excerpt caused listeners to interpret the musical surface in a discontinuous manner.

These observations suggest that several improvements to the model can be made. In particular, combining a contour generated at a finer time resolution with the results of the optimal 3 s windows might provide better predictions at a local level while maintaining accuracy at a global level. Another way the model can be improved, especially in the case of shorter excerpts, is to find a better way to predict the initial upsweep of the slider at the start of each excerpt. The current model simply uses a fixed linear increase over a 2 s duration. This slider movement is probably best modeled as a function of the loudness contour of the initial note onset. Alternatively, a simpler way to deal with this issue would be to trim the first two seconds of the data.

The model's success in predicting the two longest and most complex excerpts, the Bach-Vivaldi and Brahms, indicates that the basic cognitive concepts underlying the model are sound; it presents a first systematic attempt to predict tension given descriptions of disparate musical features. Elements of the model also touch upon the concepts of *end significance* and *slope significance* that Rozin et al. (2004) discuss in the context of recall of affective intensity in music. Rozin et al. found that recall of overall affective intensity in music had little to do with the sum of all the moments of intensity; rather, it was the perceived intensity at the end of a listening excerpt and the clear trends or slopes in intensity that contributed most. While arguably "affective intensity" is not precisely musical tension, there is a fair amount of overlap between these two perceptual phenomena. Although Rozin et al. focused on predicting *recalled* affect and do not offer real-time predictions of how musical features contribute to affective intensity, these concepts can be applied to the memory window aspect of the predictive model presented here, as it relates to the general recall of previously heard music in time. Whether the listening experience is continuing or has concluded, it can be assumed that the cognitive mechanisms that encode previous processing of musical information are still at work. Rozin et al. supports this idea of an evident trend or slope resulting in better encoding by citing work on the evaluation of past experiences (Hsee & Abelson, 1991), preference of commercials (Baumgartner, Sujan, & Padgett, 1997), and pain perception (Ariely, 1998; Loewenstein & Prelec, 1993). The results of the predictive model described here appear to confirm these findings.

The success of the model also suggests that judgment of "absolute" tension is malleable and partially dependent on context. The salience of detectable trends is by definition a function of relative changes in perceived tension. Two separate components can be distinguished from one another: the percept of change itself and the absolute perception of tension. The model attempts to account for absolute tension through the weighting of individual parameters while treating the combined, relative perception of tension change as a key factor. While a *sforzando* or an unexpectedly dissonant chord might result in a sudden spike in tension, such changes constitute isolated cases and are perhaps not as fundamental to the understanding tension as the concept of trends. This is further exemplified in the model by the influence of the memory window in determining the slope of the trend in the attentional window. When there is no memory window, an important part of how tension trends are perceived by the listener is absent. The memory window in essence helps determine the salience of a trend, and it is this concept of salience, whether determined by relative or absolute means, that is a critical factor in the perception of tension.

Author Note

Thanks to Fred Lerdahl, who provided feedback in the early stages of this work and on the prolongational reductions. The contents of this paper were also improved by helpful comments from David Temperley and three anonymous referees.

A preliminary analysis of the empirical data discussed in this article was presented at the 10th International Conference on Music Perception & Cognition (ICMPC10), Sapporo, Japan in August 2008, and reported in the conference proceedings.

Correspondence concerning this article should be addressed to Mary Farbood, Department of Music and Performing Arts Professions, 35 W. 4th St., Suite 777, New York University, New York, NY 10012. E-MAIL: mfarbood@nyu.edu

Appendix A

Calculating Harmonic Tension Using Lerdahl's (2001) Model

Lerdahl and Jackendoff's (1983) generative theory of tonal music (hereafter GTTM) formalizes the way listeners perceive hierarchical structures in tonal music. These structures are intended to model musical intuition and take the form of explicit rules that assign or "generate" structures that listeners unconsciously infer from the musical surface of a piece. There are four components to their theory: grouping structure, which segments music into motives, phrases, and sections; metrical structure, a hierarchy of alternating strong and weak beats; time-span reduction, a hierarchy of structural importance of pitches with respect to their position in the grouping and metrical structures; and prolongational reduction, a hierarchy that expresses harmonic and melodic tension and relaxation. Lerdahl (2001) has significantly extended GTTM by developing a precise model of how a piece is heard as it unfolds harmonically at multiple hierarchical levels. His theories stem from empirical evidence that listeners of varying musical backgrounds and different cultures hear pitches, chords, and regions as relatively close or distant from a given tonic in an orderly way.

Lerdahl approaches tonal tension in a systematic way by defining a formula for computing quantitative predictions of tension and attraction for events in any passage of tonal music. In order to calculate these values, the following four components are required:

1. A representation of hierarchical event structure
2. A model of tonal pitch space and all distances within it
3. A treatment of surface dissonance
4. A model of voice-leading attractions

The first component is equivalent to GTTM's prolongational reduction and can be represented in tree notation. The second component describes the internalized knowledge of listeners concerning distances of pitches, chords, and tonal regions from one another, beyond the pattern of any particular piece. It consists of three embedded spaces, the first two representing within-key hierarchies, and the third one between-key distances. The diatonic chord distance between chords x and y is defined as

$$\delta(x, y) = i + j + k \qquad \text{(A1)}$$

where i is the number of steps between two regions on the chromatic fifths circle (i.e., distance between two chords with regard to key), j is the number of steps between two chords on the diatonic fifths circle (distance with regard to chord function), and k is the number of distinctive pitch classes in the basic space of y compared to those in the basic space of x. The basic space for a chord consists of its pitch classes at the chromatic, diatonic, triadic, fifths, and root (Figure A1).

The third component, treatment of surface dissonance, is largely psychoacoustic. For example, nonharmonic tones are less stable, therefore more tense. A chord is more stable in root position than in inversion, and more stable with the root note in the melody. The surface tension associated with a chord is defined by

$$T_{diss} = f + g + h \qquad \text{(A2)}$$

where f is the chord voicing (1 if the melody is not the chord root, 0 otherwise), g is the inversion (2 if the chord is not in root position, 0 if it is), and h is the sum of all nonharmonic tones (sevenths = 1, diatonic nonharmonic tones = 3, and chromatic nonharmonic tones = 4).

For the purposes of quantifying harmonic tension for both Experiments 1 and 2, the fourth component,

(a) octave (root) level:	0											
(b) fifths level:	0							7				
(c) triadic level:	0				4			7				
(d) diatonic level:	0		2		4	5		7		9		11
(e) chromatic level:	0	1	2	3	4	5	6	7	8	9	10	11

FIGURE A1. Diatonic basic space, set to I/C (C =0, C# =1,... B =11). Taken from Lerdahl & Krumhansl (2007).

voice-leading attractions, was excluded. Only the first three components were included in the calculations, resulting in the following formula:

$$T_{hier} = \delta(x,y) + h + T_{diss} \tag{A3}$$

where h is the inherited tension value derived from a GTTM prolongational analysis. The prolongational reduction for the first several measures of the Bach-Vivaldi excerpt is shown in Figure A2. Table A1 lists the numerical values for each component and the calculated harmonic tension values for the excerpt.

Appendix B

Calculating Melodic Expectation Using Margulis' (2005) Model

Margulis' (2005) melodic expectation model is in part an extension of Narmour's Implication-Realization (I-R) theory (1990 1992). As in the case of the I-R model, the concept of purely schematic expectations as well as a systematic way of accounting for the effect of proximity between melodic events and direction of melodic intervals are of central importance. An additional factor in Margulis' model, stability, is derived from Lerdahl's tonal pitch space and melodic attraction models, thus addressing tonal relationships as well as melodic interval relationships.

The core formula defining the expectancy of a melodic event is defined as follows:

$$z = (smp) + d \tag{B1}$$

where z is the amount by which pitch x is expected to follow pitch y, s is the stability rating of x (see Table B1), m is the mobility rating of x (2/3 if x repeats y and 1 in all other cases), p is the proximity rating of x (see Table B2), and d is the direction rating of x (see Table B3).

When calculating stability ratings, the context is I (tonic) in the current key unless the following apply:

1. The current chord is a secondary chord, in which case the new context is I in the new tonicized key.
2. The current melody note is a not a chord tone, in which case the context shifts to the current chord.
3. The previous melody note was the seventh of its chord, in which case the current note is promoted to the highest stability rating if it is the lower diatonic neighbor of the previous note.

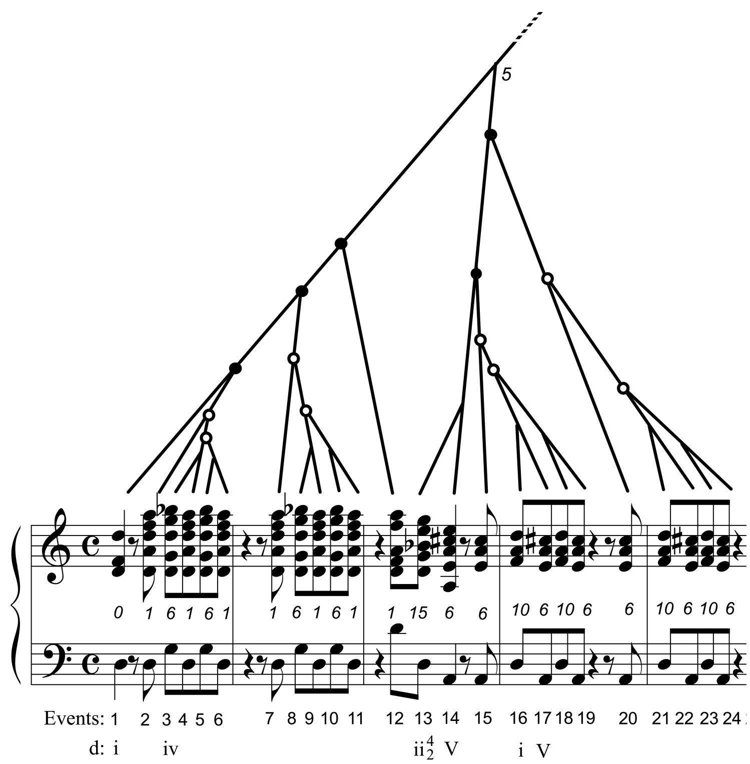


FIGURE A2. Prolongational reduction of the first 24 events in the Bach-Vivaldi excerpt.

TABLE A1. Values Summed to Calculate the Hierarchical Tension Values For the Bach-Vivaldi Excerpt Shown in Figure A2.

Event	Chord pair	*i*	*J*	*k*	Total chord distance (*i* + *j* + *k*)	Inherited values	Scale degree value	Inversion value	Non-harmonic tones	Total harmonic tension
1	δ(1)	0	0	0	0	0	0	0	0	0
2	δ(2,1)	0	0	0	0	0	1	0	0	1
3	δ(3,4)	0	1	4	5	0	1	0	0	6
4	δ(4,2)	0	0	0	0	0	1	0	0	1
5	δ(5,6)	0	1	4	5	0	1	0	0	6
6	δ(6,4)	0	0	0	0	0	1	0	0	1
7	δ(7,1)	0	0	0	0	0	1	0	0	1
8	δ(8,9)	0	1	4	5	0	1	0	0	6
9	δ(9,7)	0	0	0	0	0	1	0	0	1
10	δ(10,11)	0	1	4	5	0	1	0	0	6
11	δ(11,9)	0	0	0	0	0	1	0	0	1
12	δ(12,1)	0	0	0	0	0	1	0	0	1
13	δ(13,14)	0	1	5	6	5	1	2	1	15
14	δ(14,1)	0	1	4	5	5	1	0	0	11
15	δ(15,14)	0	0	0	0	5	1	0	0	6
16	δ(16,17)	0	1	4	5	5	0	0	0	10
17	δ(17,15)	0	0	0	0	5	1	0	0	6
18	δ(18,19)	0	1	4	5	5	0	0	0	10
19	δ(19.17)	0	0	0	0	5	1	0	0	6
20	δ(20,14)	0	0	0	0	5	1	0	0	6
21	δ(21,22)	0	1	4	5	5	0	0	0	10
22	δ(22,20)	0	0	0	0	5	1	0	0	6
23	δ(23,24)	0	1	4	5	5	0	0	0	10
24	δ(24,22)	0	0	0	0	5	1	0	0	6

4. The current chord is a predominant chord with strong voice-leading tendencies such as an augmented sixth or Neapolitan, in which case the context shifts to V in the current key.

The total expectation value is calculated by averaging the weighted values of expectations at each hierarchical level defined by Lerdahl and Jackendoff's time-span reduction, described in GTTM (1983). The weight given to each level depends on the length of time between the notes. At the note-to-note level (lowest level) the expectation ratings receive a weight of 15. Ratings at levels beyond the note-to-note level, up to and including time spans 2 s in duration, receive a weight of 5. Ratings at levels with time-span durations from 2 to 6 s receive a weight of 2. No levels

TABLE B1. Stability Ratings For Melodic Events in Margulis' (2005) Melodic Expectation Model

Stability rating	Key and chord context
6	Chord root (and, after a seventh in the melody, the pitch one diatonic step down from it
5	Chord third and fifth
4	Other diatonic pitches
2	Chromatic pitches

TABLE B2. Proximity Ratings For Melodic Events in Margulis' (2005) Melodic Expectation Model

Pitch distance in semitones	Proximity rating
1 (m2)	36
2 (M2)	32
3 (m3)	25
4 (M3)	20
5 (P4)	16
6 (d5)	12
7 (P5)	9
8 (m6)	6
9 (M6)	4
10 (m7)	2
11 (M7)	1
12 (P8)	0.25
13 (m9)	0.02
≥14 (M9)	0.01

TABLE B3. Direction Ratings For Melodic Events in Margulis' (2005) Melodic Expectation Model

Interval in semitones	Direction rating
0 (P1)	6 for continuation
1 (m2)	20 for continuation
2 (M2)	12 for continuation
3 (m3)	6 for continuation
4 (M3)	0
5 (P4)	6 for reversal
6 (d5)	12 for reversal
7 (P5)	25 for reversal
8 (m6)	36 for reversal
9 (M6)	52 for reversal

with durations longer than 6 s are considered. Thus the formula for overall expectedness of a note is defined as

$$\frac{\sum_{i=1}^{n} w_i z_i}{\sum w_i}, \tag{B2}$$

where i is the level under consideration, n is the highest level, w_i is the weight of the level under consideration, and z_i is the expectancy rating for the pitch at that level.

For the purposes of this study, only the lowest two levels of expectation values were calculated, resulting in the following formula:

$$\frac{15i+5j}{20}, \tag{B3}$$

where i is the expectation value at the lowest level, and j is the expectation value at the second-lowest level. The melodic expectation calculations for the first two measures of the Bach-Vivaldi excerpt are shown in Figure B1. The numbers above the staves show the four main parameters (stability, mobility, proximity, and distance) and the total expectation values for the first-level, note-to-note events displayed on the upper staff. The numbers below the staves show the calculations for the second hierarchical level displayed in the lower staff. The numbers in between the two staves are the final expectation values, consisting of the weighted sum of the two levels.

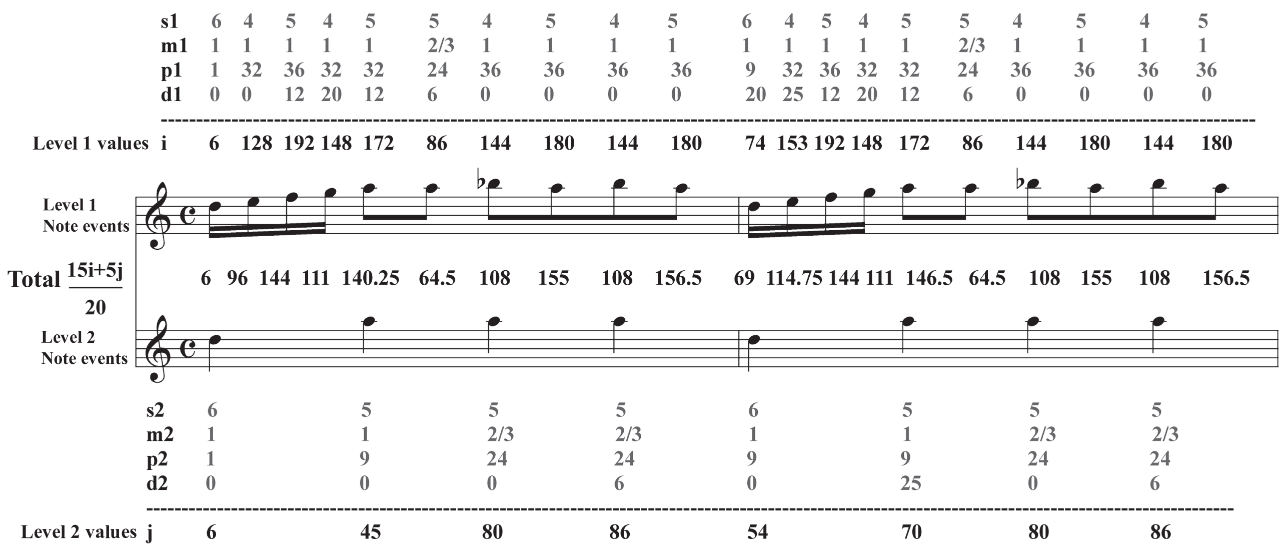


FIGURE B1. Melodic expectation values for the first two measures of the Bach-Vivaldi excerpt.

References

Ariely, D. (1998). Combining experiences over time: The effects of duration, intensity changes and on-line measurements on retrospective pain evaluations. *Journal of Behavioral Decision Making, 11,* 19–45.

Baumgartner, H., Sujan, M., & Padgett, D. (1997). Patterns of affective reactions to advertisements: The integration of moment-to-moment responses into overall judgments. *Journal of Marketing Research, 34*, 219–232.

Bharucha, J. J. (1984). Anchoring effects in music: The resolution of dissonance. *Cognitive Psychology, 16*, 485-518.

Bigand, E., & Parncutt, R. (1999). Perception of musical tension in long chord sequences. *Psychological Research, 62*, 237-254.

Bigand, E., Parncutt, R., & Lerdahl, F. (1996). Perception of musical tension in short chord sequences: The influence of harmonic function, sensory dissonance, horizontal motion, and musical training. *Perception and Psychophysics, 58*, 125-141.

Bregman, A. S. (1990). *Auditory scene analysis*. Cambridge, MA: MIT Press.

Burnsed, V., & Sochinski, J. (1998-2001). The relationship between dynamics and tension in Haydn's Symphony No. 104: A developmental study. *Psychomusicology, 17*, 19-35.

Capperella, D. A. (1989). Reliability of the Continuous Response Digital Interface for data collection in a study of auditory perception. *Southeastern Journal of Research in Music Education, 45*, 245-258.

Eerola, T., & Vuoskoski, J. (2011). A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music, 39*, 18-49.

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

Fredrickson, W. E. (1997). Elementary, middle, and high school students' perception of tension in music. *Journal of Research in Music Education, 45*, 626-635.

Fredrickson, W. E. (1999). The effect of musical performance on perception of tension in Gustav Holst's First Suite in E-flat. *Journal of Research in Music Education, 47*, 44-52.

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

Fredrickson, W. E., & Coggiola, J. C. (2003). A comparison of music majors' and nonmajors' perception of tension for two selections of jazz music. *Journal of Research in Music Education, 51*, 259–270.

Fredrickson, W. E., & Johnson, C. M. (1996). The effect of performer use of rubato on listener perception of tension in Mozart. *Psychomusicology, 15*, 78-86.

Frego, R. J. D. (1999). Effects of aural and visual conditions on response to perceived artistic tension in music and dance. *Journal of Research in Music Education, 47*, 31-43.

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

Granot, R. Y., & Eitan, Z. (2011). Tension and dynamic auditory parameters. *Music Perception, 28*, 219-246.

Gregory, D. (1989). Using computers to measure continuous music responses. *Psychomusicology, 8*, 127-134.

Gregory, D. (1995). The Continuous Response Digital Interface: An analysis of reliability issues. *Psychomusicology, 14*, 197-208.

Hackworth, R. S., & Fredrickson, W. E. (2010). Effect of text translation on perceived musical tension in Debussy's *Nöel des enfants qui n'ont plus de maisons*. *Journal of Research in Music Education, 58*, 184–195.

Hasty, C. (1997). *Meter as rhythm*. Oxford, UK: Oxford University Press.

Hauser, M., Cushman, F., Young, L., Jin, R. K.-X., & Mikhail, J. (2007). A dissociation between moral judgments and justifications. *Mind and Language*, *22*, 1-21.

Helmholtz, H. L. F. (1954). *On the sensations of tone as a physiological basis for the theory of music* (A. J. Ellis, Trans.). New York: Dover. (Original work published 1877)

Hsee, C. K., & Abelson, R. P. (1991). Velocity relation: Satisfaction as a function of the first derivative of outcome over time. *Journal of Personality and Social Psychology, 60*, 341–347.

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

Hutchinson, W., & Knopoff, L. (1978). The acoustical component of western consonance. *Interface, 7*, 1-29.

Ilie, G., & Thompson, W. F. (2006). A comparison of acoustic cues in music and speech for three dimensions of affect. *Music Perception, 23*, 319–329.

Jehan, T. (2005). *Creating music by listening* (Unpublished doctoral dissertation). Massachusetts Institute of Technology, Cambridge, MA.

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. (1997). An exploratory study of musical emotions and psychophysiology. *Canadian Journal of Experimental Psychology, 51*, 336-353.

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

Krumhansl, C. L., & Schenk, D. L. (1997). Can dance reflect the structural and expressive qualities of music? *Musicae Scientiae, 1*, 63-83.

Larson, S. (2004). Musical forces and melodic expectations: Comparing computer models with experimental results. *Music Perception. 21*, 457-498.

LARSON, S., & VANHANDEL, L. (2005). Measuring musical forces. *Music Perception, 23*, 119-136.

LERDAHL, F. (1996). Calculating tonal tension. *Music Perception, 13*, 319-363.

LERDAHL, F. (2001). *Tonal pitch space*. New York: Oxford University Press.

LERDAHL, F., & JACKENDOFF, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.

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

LOEWENSTEIN, G. F., & PRELEC, D. (1993). Preferences for sequences of outcomes. *Psychological Review, 100*, 91–108.

LONDON, J. (2004). *Hearing in time: Psychological aspects of musical meter*. New York: Oxford University Press.

LYCHNER, J. A. (1998). An empirical study concerning terminology relating to aesthetic response to music. *Journal of Research in Music Education, 46*, 303-319.

MADSEN, C. K. (1998). Emotion versus tension in Haydn's Symphony No. 104 as measure by the two-dimensional continuous response digital interface. *Journal of Research in Music Education, 46*, 546-554.

MADSEN, C. K., & FREDRICKSON, W. E. (1993). The experience of musical tension: A replication of Nielsen's research using the continuous response digital interface. *Journal of Music Therapy, 30*, 46-63.

MARGULIS, E. H. (2005). A model of melodic expectation. *Music Perception, 22*, 663-714.

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

MISENHELTER, D. (2001). An investigation of music and nonmusic majors' responses to musical tension and dynamics in Beethoven's Symphony No. 7. *Missouri Journal of Research in Music Education, 38*, 56-67.

MOORE, B. C. J., & GLASBERG, B. R. (1995). A revision of Zwicker's loudness model. *Acta Acustica, 82*, 335–345.

NARMOUR, E. (1990). *The analysis and cognition of basic melodic structures*. Chicago, IL: University of Chicago Press.

NARMOUR, E. (1992). *The analysis and cognition of melodic complexity: The implication-realization model*. Chicago, IL: University of Chicago Press.

NIELSEN, F. V. (1983). *Oplevelse af musikalsk spaending* [The experience of musical tension]. Copenhagen: Akademisk Forlag.

NIELSEN, F. V. (1987). Musical "tension" and related concepts. In T. A. Sebeok & J. Umiker-Sebeok (Eds.), *The semiotic web '86: An international yearbook* (pp. 491-514). Berlin: Mouton de Gruyter.

NEUHOFF, J. G. (1998). Perceptual bias for rising tones. *Nature, 395*, 123-124.

NEUHOFF, J. G. (2001). An adaptive bias in the perception of looming auditory motion. *Ecological Psychology, 13*, 87-110.

PATEL, A. D. (2008). *Music, language, and the brain*. Oxford, UK: Oxford University Press.

PLOMP, R., & LEVELT, W. J. M. (1965). Tonal consonance and critical bandwidth. *Journal of the Acoustical Society of America, 38*, 548-560.

PRESSNITZER, D., MCADAMS, S., WINSBERG, S., & FINEBERG, J. (2000). Perception of music tension for nontonal orchestral timbres and its relation to psychoacoustic roughness. *Perception and Psychophysics, 62*, 66-80.

ROZIN, A., ROZIN, P., & GOLDBERG, E. (2004). The feeling of music past: How listeners remember musical affect. *Music Perception, 22*, 15–39.

SCHIMMACK, U., & REISENZEIN, R. (2002). Experiencing activation: Energetic arousal and tense arousal are not mixtures of valence and activation. *Emotion, 2*, 412–417.

SCHUBERT, E. (2002). Correlation analysis of continuous emotional response to music: Correcting for the effects of serial correlation. *Musicae Scientiae, Special Issue*, 213–236.

SCHUBERT, E., & DUNSMUIR, W. T. M. (2004). Introduction to interrupted time series analysis of emotion in music. In S. Lipscomb, R. Ashley, R. O. Gjerdingen, & P. Webster (Eds.), *Proceedings of the 8th International Conference on Music Perception and Cognition* (pp. 445-448). Evanston, IL.

TENNEY, J., & POLANSKY, L. (1980). Temporal gestalt perception in music. *Journal of Music Theory, 24*, 205-241.

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

TROLIO, M. (1976). Theories of affective response to music. *Contributions to Music Education, 4*, 1-20.

VINES, B. W., KRUMHANSL, C. L., WANDERLEY, M. M., & LEVITIN, D. J. (2006). Cross-modal interactions in the perception of music performance. *Cognition, 101*, 80–113.

ZWICKER, E., & FASTL, H. (1999). *Psychoacoustics: Facts and models*. Berlin: Springer Verlag.