Proceedings of the 10th International Conference on Music Perception and Cognition (ICMPC 10). Sapporo, Japan.
Ken'ichi Miyazaki, Yuzuru Hiraga, Mayumi Adachi, Yoshitaka Nakajima, and Minoru Tsuzaki (Editors)

A Global Model of Musical Tension

Morwaread Farbood

Dept. of Music and Performing Arts Professions, New York University, USA
mfarbood@nyu.edu

ABSTRACT

Musical tension is a high-level concept that is difficult to formalize due to its subjective and multi-dimensional nature. This paper presents a quantitative, parametric model of tension based on empirical data gathered in two experiments. The first experiment is an online test with short musical excerpts and multiple choice answers. The format of the test makes it possible to gather large amounts of data. The second study requires fewer subjects and collects real-time responses to musical stimuli. Both studies present test subjects with examples that take into account a number of musical parameters including harmony, pitch height, melodic expectation, dynamics, onset frequency, tempo, and rhythmic regularity. The goal of the first experiment is to confirm that the individual musical parameters contribute directly to the listener's overall perception of tension. The goal of the second experiment is to explore linear and nonlinear models for predicting tension given descriptions of the musical parameters for each excerpt. The data from these two experiments are then correlated to musical features and finally used to train and test linear and nonlinear predictive models of tension.

I. INTRODUCTION

Music is structured sound. Through parsing and interpreting these structures, listeners arrive at a musical experience that is highly personal. How these constituent parts translate into something as subjective as emotion is multilayered and complex. One key to gaining insight into this process is the concept of musical tension. The perception of tension is an important intermediate step between the recognition of musical structures and the affective response.

While tension is a fundamental concept in theories of Western music, there exists no universal framework that describes how disparate musical features combine to produce a general feeling of tension. In most types of music throughout the world, sound dimensions such as pitch, duration, loudness, and timbre are categorized and organized into ordered relationships. Musical structures built from these categories, depending on how they are arranged, create expectancies. 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). Both the expectancies themselves and how these they are resolved (or not) influence the way people perceive tension in music.

Previous empirical studies (Krumhansl, 1996) have indicated that tension judgments appeared to be influenced by melodic contour, harmony, tonality, note density, and segmentation, as well as expressive features such as dynamics and tempo variation. Likewise, the model presented here describes an approach to modeling musical tension that takes into account multiple structural and expressive features in music. The objective of this work is to define and quantify the effect of these individual parameters on the overall perception of tension and to describe how these features reinforce and counteract each other.

The difference between the approach described here and previous work is (1) the pursuit of a rigorous framework that tries to analyze and describe tension *globally*, rather than focusing on a subset of features (in most cases, harmonic and melodic aspects) (2) using the empirical results to systematically develop a new theory based on the *interaction* of musical features (Krumhansl 1996; Palmer, 1996; Bigand & Parncutt, 1999; Smith & Cuddy, 2003; Lerdahl & Krumhansl, 2007, to name a few).

II. METHODOLOGY

The global tension model is based on a significant amount of empirical data gathered in two experiments. The first experiment is a web-based study designed to gather data from thousands of subjects from different musical and cultural backgrounds. The second experiment is a smaller study designed to obtain real-time, continuous responses to stimuli. In these experiments, subjects were asked to listen to musical examples and describe how they felt the tension was changing. The musical excerpts were composed or selected with six parameters in mind: harmony, melodic expectation, pitch height, tempo, onset frequency, and dynamics. By reducing degrees of musical freedom, the examples isolated or combined these features in order to effectively gauge how they affected subjects' overall perception of tension. Some examples consisted of a single feature changing over time, while others included two or more features either in concert or opposition to one another.

A. The Role of Prior Work

Since the model is defined in terms of multiple musical parameters, all of these parameters must be adequately described before their contribution to the whole can be analyzed. In other words, if features such as harmony and melodic expectation contribute in some way to the overall feeling of tension, their contributions in isolation of other features must be quantified first before they can be combined and compared with other parameters that might strengthen or weaken their contribution.

Some of these features are easy to quantify—for example, tempo is a one-dimensional feature that can be described in terms of beats per minute with respect to time. Harmony and melodic expectation, on the other hand, are complex multidimensional features. The prior work presented here, in particular, Lerdahl's tonal tension model (Lerdahl, 2001; Lerdahl, 2007) and Margulis' melodic expectation model (Margulis, 2005), are utilized to quantitatively describe the individual contributions of harmony and melodic expectation to tension; given that these two models are already quantitative, they are ideal for this purpose. The resulting descriptions produced by them are then used to inform the analysis required to define a new global tension model. The

ISBN: 978-4-9904208-0-2

goal of this thesis is not to find new models for describing harmonic tension, melodic tension, or any other individual parameter, but to apply available theoretical descriptions of them (assumed to be reasonable approximations of listeners' perceptions) when necessary, and then determine how they combine and interact with other parameters like tempo and dynamics to produce a global feeling of tension.

B. Experiment 1

Experiment 1 collected data from nearly 3000 subjects, from 108 different countries. The musical examples were 2 to 60 seconds long and recorded using piano, strings, and unpitched percussion sounds. Each example was entered in Finale (a notation editor) and played back and recorded with two MIDI synthesizers, a Kurzweil K2500 and Roland 5080. There were a total of 207 audio files used in the experiment, but each subject was tested on only 11 examples, 10 of which were randomly chosen from descriptive categories.

Subjects chose from a selection of curves graphically depicting changes in tension (Figure 1). The graphical choices were used to enable test-takers with limited English skills to participate. After test takers selected an answer that best fit how they perceived the tension in a musical example, they rated the confidence of their response on a scale of 1 to 5. This confidence value provided additional information on how the listener judged an example and how clear or unclear the change in tension was.

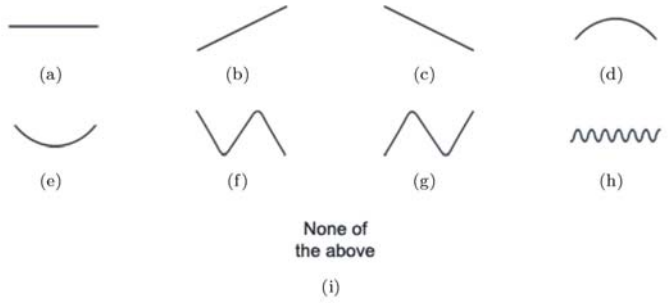

Figure 1: Response choices for Experiment 1.

At the beginning of the test, subjects were presented with a sample question that was answered for them (Figure 4). This question was assumed to have an obvious answer (or at least as obvious an answer as possible given the nature of the subject mater). Certainly not all of the questions had a clear answer; the sample question was meant as a guide to show the subject what a feasible response might be.

The main hypothesis being tested was the assumption that changes in each parameter would correlate to changes in tension. For the case of loudness and tempo/onset frequency, common sense dictated that an increase in those features would result in an increase in tension. Likewise, increase in pitch height was assumed to correspond to an increase in tension. Defining harmonic tension was more complex—Lerdahl's tonal tension model (Lerdahl, 2001; Lerdahl 2007) was used to provide a quantitative assessment of tension for each chord (see Figure 2 and Table 1). Rhythmic irregularity, unlike the other features, was not a parameter that would obviously affect tension. The hypothesis was that an increase in rhythmic irregularity (i.e. lack of consistency in onset frequency) would result in an increase in tension.

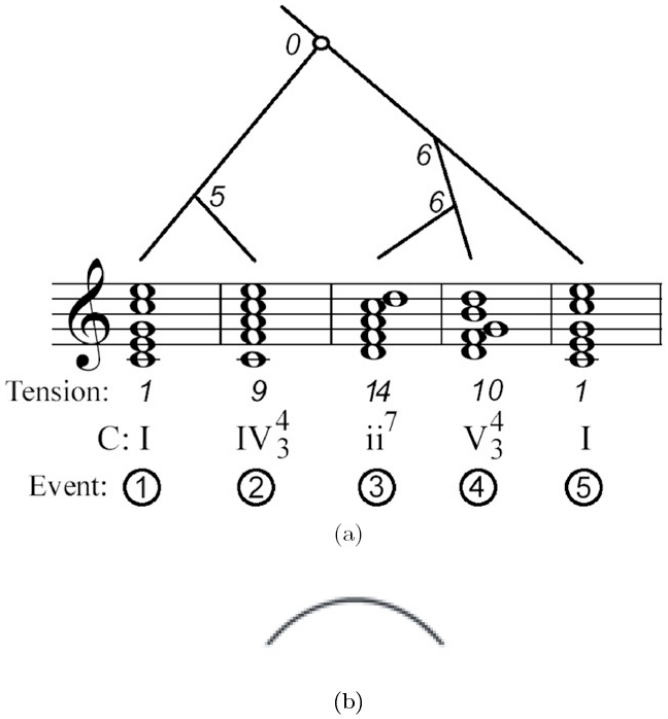

Figure 2: (a) An example from Experiment 1 used to test subjects' responses to harmony. The prolongational reduction is shown. (b) The answer corresponding to harmony for example shown in (a).

The approach used in composing the examples was simple: each feature was isolated in at least one example and combined with other parameters moving in the same or opposite *direction*. Direction refers to the increase or decrease in tension caused by changes in the feature. For example, if something is speeding up, the general feeling will most likely be that the tension is increasing. Thus a single note repeating without any change except for tempo would be an example of that parameter being featured in isolation. For example, in the example in Figure 3, the tempo is slowing down yet the harmony is intensifying. This is naturally confusing to the listener. If the listener hears that the net effect is an increase in tension, it indicates that the changes in harmony are having a stronger overall effect on the perception of tension than the decrease in tempo. If the listener feels that the overall the tension is decreasing, it indicates that the changes in tempo are having a stronger effect than the harmonic movement.

Clearly, there is no correct response for any of these questions. The interest lies not in finding a single right answer, but in determining the degree to which participants are uncertain.

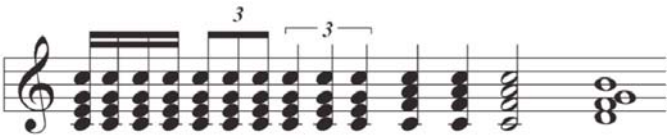

Figure 3: Example where two parameters, harmony and onset frequency, appear to be conflicting in their contribution to tension.

Table 1: Chart showing values required to calculate harmonic tension values using Lerdahl's tonal tension model. Values correspond to the example in Figure 4-11.

Event	Branched pair	Chord distance values	Inherited values	Scale degree value	Inversion value	Nonharmonic tones	TOTAL TENSION
1	δ(1,5)	0	0	1	0	0	1
2	δ(2,1)	5	0	1	2	1	9
3	δ(3,4)	6	6	1	0	1	14
4	δ(4,5)	6	0	1	2	1	10
5	-	0	0	1	0	0	1

C. Experiment 2: Real-Time Response

Experiment 2, in contrast, had far fewer subjects but more varied, complex examples. Unlike the first experiment, real-time responses to stimuli were recorded. Collecting data with retrospective judgments has the advantage that it allows a relatively simple experimental setting (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. Also, it is difficult to use long examples that change over time since it would require that the responses change over time; these dynamic qualities are not well represented by a single retrospective judgment. This method provides a relatively efficient way of capturing a richer response to the data (Toiviainen and Krumhansl, 2003).

Ten musical examples were used as stimuli in Experiment 2. Six of these examples were short (10 seconds or less) and were composed specifically for the study. They were similar to the questions found in Experiment 1 and were designed to clarify some points that were not entirely clear from the results of the previous study. In addition to these shorter questions, there were four excerpts taken from the classical repertoire: Schoenberg Klavierstück, Op. 11 No. 12, a Bach organ transcription of a Vivaldi concerto, Beethoven Symphony No. 1, and Brahms Piano Concerto No. 2. The longer examples were 20 seconds to one minute in length. They were also considerably more complex than any of the examples composed for the study.

Thirty-five subjects, drawn from the faculty and student bodies at MIT, participated in the experiment. Their ages ranged from 19 to 59, with a mean age of 30. Approximately half of the participants were experienced musicians; the median and mean number of years of combined musical training and active performance for all subjects were 10 and 12 respectively.

Test subjects were presented with a computer interface written in C++ for Windows. Moving the mouse up and down caused a large vertical slider bar to move up and down without the subject having to press the mouse button. This was done so that subjects would not tire of holding the mouse button down or worry about an extra action that might distract from the listening experience. Subjects were instructed to raise the slider if they felt a general feeling of musical tension increasing, and to lower it when they felt it lessening. Each musical excerpt was played four times. After listening and responding to an excerpt, subjects were asked to select a confidence value. The playback of each iteration was preceded by visual cues that would appear on the interface to prepare the subject.

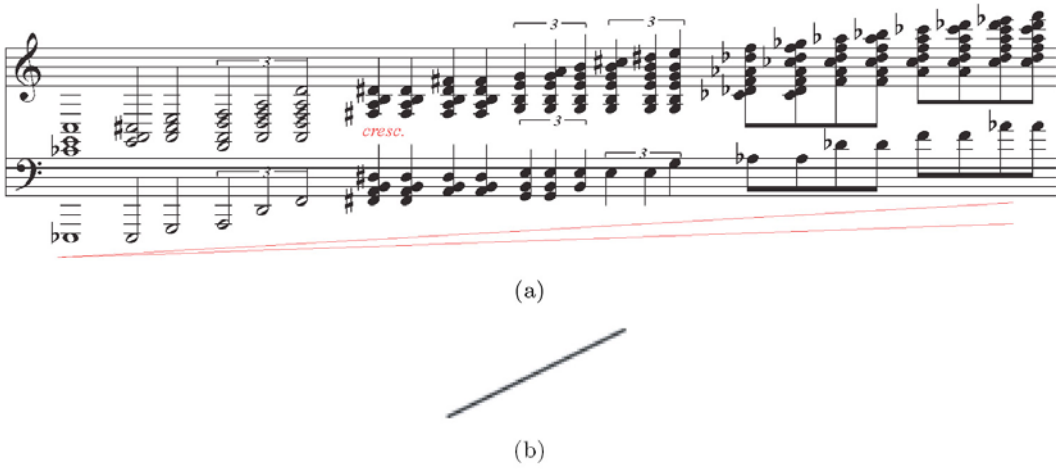

Figure 4: (a) Initial sample question in Experiment 1. All parameters are clearly increasing in tension. (b) Assumed "correct" answer: the graphical response indicating tension is increasing.

III. RESULTS

The results of Experiments 1 and 2 clearly demonstrated that the musical features explored in both studies have a significant and calculable impact on how listeners perceive musical tension. Results from both experiments contributed to a more global picture of how changes in individual musical parameters affect changes in tension.

A. Experiment 1 Results

As discussed mentioned in the previous section, there were nine possible graphical responses subjects could choose from in Experiment 1 (Figure 1). A perusal of the data indicated that subjects tended to select answers that reflected what they heard at the very end of excerpts. Curves more complex than the first four response choices were rarely selected even if they corresponded more closely to a salient musical feature. It is possible that response choices such as those illustrated in Figures 1(f) and 1(g) depict more changes in tension than subjects could track and recall with certainty.

Analysis of Experiment 1 data clearly showed that all features tested with the exception of rhythmic irregularity had a significant effect on subjects' perception of changing tension. In more complex examples where features were counteracting each other, the relative importance of each feature appeared to depend on its salience. When multiple features were combined in parallel, they considerably strengthened the feeling of changing tension.

When two features were paired so that one intensified while the other relaxed, the results showed that they often counteracted one another. In the case of loudness versus onset frequency, the initial results indicated that loudness had a considerably stronger effect than onset frequency.

Overall, pitch height appeared to have the clearest effect (possibly because of its more obvious mapping to the graphical curve), while onset frequency seemed to have the weakest, particularly when opposed to other features. One thing lacking was a quantitative way to compare the differences resulting from the amount of change of each feature and how this amount might have affected the result. This was only effectively shown for changes in pitch height.

The results of comparing responses of musically inexperienced and musically experienced subjects indicate that musicians have a greater sensitivity to harmony and onset frequency. While it appears that non-musicians were slightly more responsive to changes in pitch height when comparing examples featuring simple harmonic progressions and small changes in pitch, this might be the result of sensitivity (or lack of it) to harmony. In other words, given a non-tonal context, all subjects, regardless of musical background, might respond similarly to changes in pitch, but in a tonal context, musicians are drawn more to harmonic motion, thus dampening the effect of pitch change if it's in opposition to harmonic direction.

B. Experiment 2 Analysis

The goal of Experiment 2 was to define a model that could quantitatively describe and predict the subject data (slider values corresponding to perceived tension at any given point in time in an excerpt) given descriptions of the way each musical feature changed and contributed to tension over time in the excerpt. Assuming these descriptions to be accurate, a new, global model of tension could be implemented—a new model that could predict overall tension by taking into account how all the individual features detracted or contributed to increases or decreases in perceived tension at any point in the excerpt.

1) Feature graphs

All of the musical parameters confirmed in Experiment 1 as well as one additional parameter, melodic expectation, were quantitatively described for each excerpt in Experiment 2. These descriptions or *feature graphs* included the following parameters:

- Harmonic tension
- Melodic Expectation
- Pitch height for soprano, bass, and inner voices
- Dynamics (loudness)
- Onset frequency
- Tempo

None of the excerpts required all of the possible feature graphs. For example, if there was no change in tempo throughout an excerpt, the graph representing it (all zeros) was not required.

While pitch height is an important factor in how listeners perceive tension, it is also somewhat crude. It does not take into account some of deeply schematic expectations of melodic contour described in Narmour's theories (Narmour, 1990; Narmour, 1992) as well as the tonal implications. So in addition to pitch height, a graph was added that described melodic expectation. Margulis' model of melodic expectation, with a few minor adjustments, was used to analyze the examples.

The graphs for loudness were derived directly from the audio files used in the experiment. The values were produced with Jehan's psychoacoustic loudness model (Jehan, 2005), 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). Frequency warping models how the inner ear (cochlea) filters sound. Frequency masking is a phenomenon that occurs when frequencies are close to one another—so close that listeners have difficulty perceiving them as unique. Temporal masking is based on time rather than on frequency. Humans have trouble hearing distinct sounds that are close to one another in time; for example, if a loud sound and a quiet sound are played simultaneously, the quiet sound will be inaudible. However, if there is enough of a delay between the two sounds, the quieter sound would be heard.

In addition to the feature graphs, a vector was generated for each example consisting of all zero values except at points in time where note onsets, beats, and downbeats were present (Figure 5). Binary values were assigned to the three labels: onset (1), beat (2), downbeat (4). These values were summed as needed.

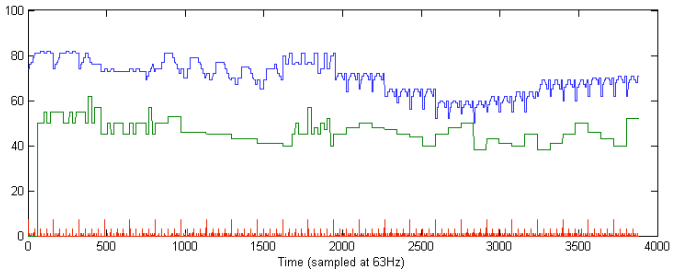

Figure 5: Graph showing beat and onset markings in red for Q05 (Bach-Vivaldi). The pitch height graphs for the melody and bass line have been added in blue and green for reference.

Harmonic tension was by far the most complex feature described. Given that Lerdahl's tonal tension model was already supported by empirical evidence and quantitative in description, it was ideal for analyzing each excerpt and producing harmonic tension values. It is used in its entirety except for the melodic attraction rule, which is partially represented in the melodic expectation graph. As in the case of the simple harmonic progressions analyzed in excerpts from Experiment 1, the first step in the analysis process was to produce a prolongational reduction of each example.

2) Linear correlation of features

The first step in the analysis process was to get an idea of how each feature graph for each example correlated with the subject data. Feature graphs and subject data were down-sampled to 50Hz and then normalized. Normalization consisted of first subtracting the mean of each graph from all its points and then making them unit variance by dividing by the standard deviation. The former was done in order to take into account differences in slider offsets at the beginnings and ends of sample sets. The latter was required to level the relative differences in change between subjects without altering the information. For example, two subjects might respond differently on an absolute scale but very similarly on a relative scale—one subject might move the slider on average 2 units for a certain amount of change in tension while another subject would move it 10 units for the same change. The mean of the subject data was used in the subsequent analyses.

In general, it was difficult to correlate the feature graphs with the subject data because of the jagged edges of the the functions were at odds with the smooth curves of the slider movements. The correlation values represent the relative important of each feature fin a given excerpt. In examples where a certain feature had a clear trend rather than subtle fluctuations, the *r* (correlation coefficient) values and *p*-values indicated a clear user response to that feature.

In other words, the importance of the feature is proportional to its salience. However, it must be noted that these results can be misleading if there are nonlinear effects, as a linear correlation is not going to capture them. Nevertheless, it still gives a very good indication of which features are more important than others for each example.

3) A Predictive Model

The final step was the implementation of a model that mathematically described and predicted how listeners perceived tension in an excerpt given how the feature graphs described the changing musical parameters in the excerpt. As noted before, these feature descriptions—three of which were based on other theories (Lerdahl's tonal tension model, Margulis' melodic expectation model, and Jehan's psychoacoustic loudness model)—were assumed to be accurate representations of their respective musical parameters.

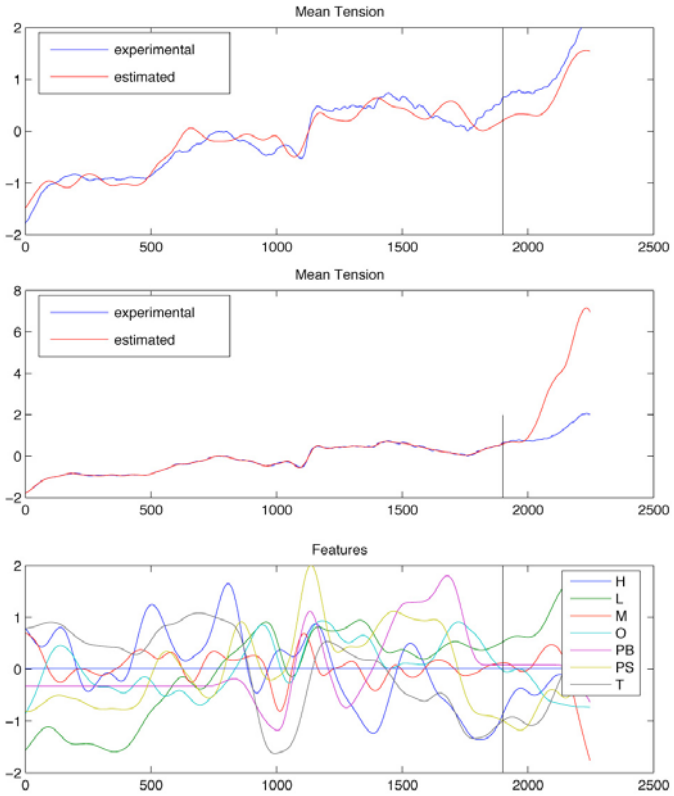

Figure 6: The top window shows the results of linear regression for a Brahms excerpt. The center window shows the results of polynomial (quadratic in this case) regression for the same excerpt. The blue lines represent subject data and the red lines represent the model. The vertical line indicates the division between training data for the model and out of sample data used for prediction. The bottom window shows the smoothed feature graphs for the excerpt.

Linear and nonlinear (polynomial) regression were performed in an attempt to fit the subject data with the feature descriptions and then predict results for new data. It was assumed that tension could be expressed as time-varying function of a set of musical parameters. The goal was to approximate the function so that it matched the subject data as accurately as possible.

Regression analysis was performed on the first part (training data) of each example. The results were then used to predict the second part (the out of sample data). While the linear model worked fairly well for most of the examples, there were cases where the quadratic model performed more successfully. This suggests that a fairly simple nonlinear model can do better than a linear model in some cases and is sufficient to capture the complexity of the problem. However, there were some cases where nonlinear regression resulted in apparent overfitting of the data (Figure 6).

A general issue that needs to be considered is the fact that the training data in some examples was insufficient to produce a robust model. Particularly in the case of short examples, the brief time-span of musical events covered by the training data did not contain the necessary information to adequately predict responses for future situations; the accuracy of predictions are always dependent on the range of events that have already occurred.

IV. DISCUSSION

In the course of examining and analyzing the data from Experiments 1 and 2, there were a number of issues that came to light. Generally speaking, there was a lack of examples that allowed the quantification of features with respect to their salience in Experiment 1. The only feature that had examples addressing this issue was pitch height. Even in that case, additional variables made the comparisons less straightforward. In future experiments, examples should be composed such that different quantities of change in loudness, tempo, or harmony can be assessed. For example, given an example where the tempo increases to twice the original speed, there ought to be at least two more examples that increase at different ratios of the original tempo (e.g. 1.5 and 4). In this way, the thresholds for perceiving significant changes can be evaluated systematically.

On a different note, there was so much data collected in Experiment 1, that not all of it could be analyzed. There still remains much to be discovered given the detailed surveys the subjects filled out. It would be particularly interesting to see if subjects from Western countries responded differently from non-Western ones. Careful sorting would have to be done based on musical background as well as country of origin in order to determine how much a subject has been influenced by Western music.

Perhaps the most successful experimental format would combine the best features of Experiments 1 and 2. The results of Experiment 2 would have been much stronger with more subject data. While having thousands of subjects (as was the case for Experiment 1) for this type of study might seem implausible, it should be possible to collect real-time slider responses to musical stimuli in a web-based setting. The biggest problem would be the lack of an observer to instruct the subject and monitor the test. However, one might argue that with thousands of data sets, it might not matter so much.

Perhaps the most significant feature missing from the list of parameters considered for the tension model was timbre. While there were different instrumental sounds used in the experiments, they were there merely for control purposes. It would be interesting to see if timbral features such as brightness and roughness are as influential as harmony or tempo in determining listeners' perception of tension.

Another feature that ought to be considered is meter. Although there were a few examples dealing with meter in Experiment 1, they were not very successful in gauging the effect on listeners' perception of tension. Furthermore, they were purely experimental examples thrown in considerably after the data collection process began. In any case, the perception of meter and its influence on other musical structures are so intricately intertwined that it might not be possible to isolate it as a parameter in the same way other features were tested in Experiment 1. Nonetheless, the possible avenues to explore in that domain, as well as those related to the parameters discussed for the current model, remain endless.

Regardless of possible future directions, the investigation described here raises interesting questions about the nature of complex musical phenomena like tension and the mathematical models that could be used to describe them. While the possibility of finding a truly global model seems implausible, some reasonable approximation might be possible, given some modularity in the description of the parameters that are dependent on the cultural and musical backgrounds of listeners. Furthermore, even if all listeners had the same history of listening, the perception of a high-level phenomenon like tension would still be subject to individual interpretation.

On the other hand, responses to features such as loudness are biologically wired due to extra-musical necessity, thus universal across all cultures. Huron (2006) explores some theories on how expectation (and thus, tension) might be rooted in basic, innate responses to the environment. If a truly global model can be envisioned, it would need to accurately predict how the innate and acquired components interact.

REFERENCES

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

Farbood, M. (2006). A Quantitative, Parametric Model of Musical Tension. Ph.D. dissertation, Massachusetts Institute of Technology.

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

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

Jehan, T. (2005). *Creating Music by Listening*. Ph.D. dissertation, Massachusetts Institute of Technology.

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

Lerdahl, F. (2001). *Tonal Pitch Space*. Oxford University Press, New York.

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

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

Narmour, E. (1990). *The Analysis and Cognition of Basic Melodic Structures*. University of Chicago Press, Chicago.

Narmour, E. (1992). The Analysis and Cognition of Melodic Complexity: The Implication-Realization Model. University of Chicago Press, Chicago.

Palmer, C. (1996). Anatomy of a performance: Sources of musical expression. *Music Perception*, *13*, 433-453.

Smith, N. A., & Cuddy, L. L. (2003). Perceptions of musical dimensions in Beethoven's Waldstein sonata: An application of tonal pitch space theory. *Musicae Scientiae, 7*, 7-34.

Toiviainen, P. and Krumhansl, C. L. (2003). Measuring and modeling real-time responses to music: The dynamics of tonality induction. *Perception*, *32(6),* 741–766.