A Quantitative, Parametric Model of Musical Tension

by

Morwaread M. Farbood

A.B. Music
Harvard College (1997)

M.S. Media Arts and Sciences
Massachusetts Institute of Technology (2001)

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February, 2006

Author ....
Program in Media Arts and Sciences
January 2006

Certified by ....
Tod Machover
Professor of Music and Media
Thesis Supervisor

Accepted by ....
Andrew B. Lippman
Chair, Departmental Committee on Graduate Students

A Quantitative, Parametric Model of Musical Tension

by Morwaread M. Farbood

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning, on January 2006,
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

Abstract

This thesis presents a quantitative, parametric model for describing musical tension. While the phenomenon of tension is evident to listeners, it is difficult to formalize due to its subjective and multi-dimensional nature. The model is therefore derived from empirical data.

Two experiments with contrasting approaches are described. 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 resulting model is considered for potential incorporation into computer-based applications. Specifically, it could be used as part of a computer-assisted composition environment. One such application, Hyperscore, is described and presented as a possible platform for integration.

Thesis supervisor: Tod Machover
Title: Professor of Music and Media

Thesis Committee

Thesis supervisor ....

Tod Machover

Professor of Music and Media

MIT Program in Media Arts and Sciences

Thesis reader ....

Fred Lerdahl

Fritz Reiner Professor of Music

Columbia University

Thesis reader ....

François Pachet

Research Scientist

Sony Computer Science Laboratory, Paris

Acknowledgments

This work couldn't have been done without the help and support of many friends, colleagues, and mentors. I'd like to thank the following people in particular:

My advisor Tod Machover, who has given me so many great opportunities and supported me all these years.

My thesis committee members Fred Lerdahl and François Pachet, both of whose work has inspired me.

Colleagues and friends who contributed to my thesis work: Tristan Jehan, Josh McDermott, Joe Presbrey, and Bernd Schoner.

Friends: Tristan Jehan, Mike Hawley, Junne Kamihara, Ariane Martins, Anthony Beurivé, J. C., and many others from MIT, Harvard, and beyond.

Fellow students, past and present, from the Opera of the Future group.

Jeanne Bamberger, Marcus Thompson, David Deveau, and the rest of the MIT music department.

Bernd and the Schoner family.

My family: Mom, Dad, and Mina.

To the memory of David Lewin (1933-2003)

Table of Contents

1 Introduction **23**

1.1 Musical parameters 24

1.2 Objective and approach 24

1.3 Motivation 25

1.4 Overview of Thesis 26

2 Prior Work **29**

2.1 Schenkerian analysis 29

2.2 Generative theory of tonal music 30

2.3 Lerdahl's tonal tension model 31

2.3.1 Perception of tonal hierarchies 31

2.3.2 Lerdahl's criteria for tonal tension 33

2.4 Melodic structures 34

2.4.1 Implication-Realization theory 35

2.4.2 Margulis' model of melodic expectation 36

2.4.3 Melodic contour and tension 38

2.5 Some comments on tension outside the tonal context 39

2.6 Modeling expressive characteristics of music 40

2.7 The role of prior work in defining a new model 42

3 Method and Experimentation 43

3.1 Experiment 1: Web-based study 43

3.1.1 Procedure 43

3.1.2 Selection of musical examples 44

3.1.3 Design of musical examples 47

3.1.4 Demographics 48

3.2 Experiment 2 50

3.2.1 Test format 50

4 Model Results 55

4.1 Experiment 1 analysis 55

4.1.1 Pitch height 55

4.1.2 Loudness 58

4.1.3 Onset frequency 59

4.1.4 Harmony 60

4.1.5 Rhythmic irregularity 63

4.1.6 Interaction of features 64

4.1.7 Three corresponding features opposed by a fourth feature 67

4.1.8 Musicians versus non-musicians 73

4.1.9 Summary 75

4.2 Experiment 2 Analysis 76
4.2.1 Feature graphs 76
4.2.2 Linear correlation of individual features with empirical data 83
4.2.3 Musical experience 86
4.2.4 The goal: a predictive model 87
4.3 Note on implementation of analysis tools 91

5 Applications 93

5.1 The original tension line in Hyperscore 94
5.2 Later versions 94
5.3 Final Hyperscore version 97
5.3.1 Harmonic tension line 100
5.4 The future of Hyperscore 104
5.5 A new tension line 104

6 Conclusions 111

6.1 Summary 111
6.1.1 Experimental results 111
6.1.2 The model 112
6.1.3 Musicians vs. non-musicans 113
6.1.4 Applications 113
6.2 Discussion 114
6.3 Contributions 115

6.4 Future directions and final remarks 115

Appendix A Musical excerpts from Experiment 2 **117**

Appendix B Feature analysis from Experiment 2 **121**

Bibliography **125**

List of Figures

2-1 Middleground analysis of Bach Prelude in C Major, *Well-Tempered Clavier* Bk. I [Forte and Gilbert, 1982]. 30

2-2 GTTM analysis of a Bach chorale showing the prolongational tree, metrical grid, and grouping analysis [Lerdahl, 2001]. 32

3-1 Response choices for Experiment 1. 45

3-2 (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. 45

3-3 Three musical examples from group 001. 46

3-4 Example where two parameters, harmony and tempo, appear to be conflicting in their contribution to tension. 49

3-5 The first question in Experiment 1. All parameters are moving in the same direction: the harmony is relaxing, the tempo is slowing down, the volume is decreasing, and the pitch contour is falling consistently. 49

3-6 Map showing the origin of subjects in Experiment 1. Numbers shown include only the people who completed both the first and final surveys (image courtesy of Rich Chan and Google). 49

3-7 Top twelve countries of origin for subjects in Experiment 1. 50

3-8 Experiment 2: countdown to prepare subjects for excerpt playback. 51

3-9 Screen shot of Experiment 2 interface in data recording/subject listening mode. 52

4-1 (a) Example where pitch height is the only feature changing. (b) The “correct” answer corresponding to pitch height for (a). 56

4-2 Graph showing subjects’ responses to changes in pitch height. 56

4-3 Four examples with varying degrees of change in pitch height. 57

4-4 Subjects’ responses to musical examples shown in Figure 4-3. 57

4-5 (a) An example designed to measure subjects’ responses to loudness. (b) Answer corresponding to loudness for example shown in (a). 58

4-6 Graph showing subjects’ responses to all examples that have loudness as the only changing feature. 58

4-7 (a) An example designed to test subjects’ responses to onset frequency. (b) Answer corresponding to onset frequency for example shown in (a). 59

4-8 Graph showing subjects’ responses to all examples where the only changing feature is onset frequency. 59

4-9 An example designed to isolate and test tempo changes (a subcategory of onset frequency). While tempo is the only feature changing, pitched notes are present, which might be affecting the results. Subjects might not only be responding to changes tempo, but lack of pitch change as well, thus dampening the perceived tension change. 59

4-10 Subjects’ responses to two types of examples featuring change in onset frequency. The first set consists of examples recorded with pitched instruments (strings or piano) and the second set consists of examples recorded with unpitched percussion. The presence of pitch in the first set appears to have affected subjects’ responses to onset frequency. 60

4-11 (a) An example used to test subjects’ responses to harmony. The prolongational reduction is shown. (b) Answer corresponding to harmony for example shown in (a) 61

4-12 Chart showing values required to calculate harmonic tension values using Lerdahl’s model. Values correspond to the example in Figure 4-11. 61

4-13 Three examples designed to test subjects’ responses to changes in harmony. All three contain the same chord progression but voiced slightly differently. . 62

4-14 Harmonic changes paralleled by clear changes in pitch height (i.e. harmonic tension increasing while pitch height increasing). 62

4-15 Graph showing subjects' responses to all harmony examples and a subset of those examples where pitch changes parallel harmonic movement. 62

4-16 (a) An example used to gauge subjects' responses to rhythmic irregularity. (b) Answer corresponding to rhythmic irregularity for example shown in (a) 63

4-17 Subjects' responses to examples where the only feature changing is rhythmic regularity. 64

4-18 An example showing changes in rhythmic regularity as well as slight changes in average onset frequency. 64

4-19 Subjects' responses to examples with changes in rhythmic regularity as well as slight changes in average onset frequency. 65

4-20 Example used to compare changes in loudness and onset frequency. 65

4-21 Subjects' responses to examples where loudness and onset frequency change in opposite directions. 66

4-22 Example used to compare changes in onset frequency with changes in harmony. 66

4-23 Subjects' responses to examples where onset frequency and harmony are in opposing directions. 67

4-24 Example used for testing combinations of changing loudness, harmony, and onset frequency. 67

4-25 Graphs of subjects' responses to combinations of changing loudness, harmony, and onset frequency. 68

4-26 Example used for testing combinations of changes in harmony, pitch height, loudness. 68

4-27 Graph of subjects' responses to examples where harmony, pitch height, and loudness are in parallel and examples where harmony changes in the opposite direction of pitch height and loudness. 69

4-28 Example showing change in harmony against changes in onset frequency, pitch height, and loudness. 69

4-29 Example showing change in loudness against changes in onset frequency, harmony, and pitch height. 70

4-30 Example showing change in pitch height against changes in onset frequency, harmony, and loudness. 70

4-31 Example showing change in onset frequency against changes in harmony, pitch height, and loudness. 70

4-32 Graph of subjects' responses to examples with harmony against three other features. 71

4-33 Graph of subjects' responses to examples where change in loudness goes against changes in three other features. 71

4-34 Graph of subjects' responses to examples where change in pitch height goes against changes in three other features. 72

4-35 Graph of subject responses to examples where change in onset frequency goes against three other features. 72

4-36 Graph comparing responses of subjects with considerable musical training with subjects with less musical training. 73

4-37 Graph showing responses of subjects with considerable musical training to examples with simple harmonic progressions and small changes in pitch. . . 74

4-38 Graph showing responses of subjects with less musical training to examples with simple harmonic progressions and small changes in the melody line. . . 74

4-39 Graph showing pitch height values over time for Q05 (Bach-Vivaldi). 77

4-40 First page of analysis showing how melodic expectation values are calculated for Q05 (Bach-Vivaldi). i values consist of direct note-to-note level expectations, and j values consist of high-level patterns based on salience and metrical placement. s represents stability ratings, m represents mobility ratings, p represents proximity ratings, and d represents direction ratings. The values between the two staves are the final melodic expectations ratings for each note. 77

4-41 Melodic expectation graph for Q05 (Bach-Vivaldi). 77

4-42 Loudness curve of Q03 (Beethoven). The blue line is the perceived loudness in dB produced by Jehan's psychoacoustic model and the red line is the smoothed version used for the feature. 78

4-43 Graph showing onset frequency in Q05 (Bach-Vivaldi). 79

4-44 Graph showing change in tempo for Q09 (cadence with *crescendo* and *ritardando*). 79

4-45 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. 79

4-46 Harmonic tension feature graph Q05 (Bach-Vivaldi). The x-axis represents time, and the y-axis, tension values. 80

4-47 Chart of harmonic tension calculations for Q05 (Bach-Vivaldi). δ is the sum of i, j, and k. 81

4-48 Prolongational reduction of Q05 (Bach-Vivaldi). 82

4-49 Subject and feature data graphs for Q09 (cadence with *crescendo* and *ritardando*). The first window shows 10 out of the 132 subject responses for Q09 (number of subjects multiplied by number iterations = total number of responses: 33 * 4 = 132). The second window shows the mean of the subject responses. The third window shows the corresponding feature graphs. The plus-signs (+) indicate the beat positions. H = harmony, L = loudness, M = melodic expectation, O = onset frequency, PB = pitch height of bass line, PS = pitch height of soprano line, and T = tempo (note that there is no graph line for PB because Q09 has no bass line). 83

4-50 Table showing correlation results for Q01–Q05. H = harmony, O = onset frequency, L = loudness, M = melodic expectation, PS = pitch height of soprano line, PB = pitch height of bass line, and T = tempo. The first two columns of r and p-values are for all data points (50 points sampled per second). The second two columns are for data sampled at every beat or note onset. Negative r-values indicate inverse correlation. 84

4-51 Table showing correlation results for Q06–Q10. H = harmony, O = onset frequency, L = loudness, M = melodic expectation, PS = pitch height of soprano line, PB = pitch height of bass line, and T = tempo. The first two columns of r and p-values are for all data points (50 points sampled per second). The second two columns are for data sampled at every beat or note onset. Negative r-values indicate inverse correlation. 85

4-52 Subject data mean for Q01 (simple chord progression with rising melodic line). The red line represents the average response of musically experienced subjects and the blue line represents responses of musically inexperienced subjects. 87

4-53 The first window shows the results of linear regression for Q09 (cadence with *crescendo* and *ritardando*). The blue line represents subject data and the red line represents the model. The vertical line indicates the division between training data for the model and out of sample data used for prediction. The second window shows the smoothed feature graphs. 89

4-54 The first window shows the results of polynomial (quadratic in this case) regression for Q09 (cadence with *crescendo* and *ritardando*). The blue line represents subject data and the red line represents the model. The vertical line indicates the division between training data for the model and out of sample data used for prediction. The second window shows the smoothed feature graphs. 89

4-55 The first window shows the results of linear regression for Q10 (Brahms). The blue line represents subject data and the red line represents the model. The vertical line indicates the division between training data for the model and out of sample data used for prediction. The second window shows the smoothed feature graphs. 90

4-56 The first window shows the results of polynomial (quadratic in this case) regression for Q10 (Brahms). The blue line represents subject data and the red line represents the model. The vertical line indicates the division between training data for the model and out of sample data used for prediction. The second window shows the smoothed feature graphs. 90

4-57 Screenshot of application written in OpenGL for visualizing data. 92

5-1 Screenshot of the first version of Hyperscore. The green line is drawn by the user, the blue, red, and white lines are added by the computer. The blue lines show the computed section divisions, the red lines show the min and max boundaries of the section, the white lines represent the estimated curves. 95

5-2 Screenshot of Hyperscore Version 2. The red pen, which has no motive attached to it, is used to draw the main curve. The other pens are for making annotations to the red line. The graphs below the drawing window indicate the prominence of each motive (colored lines) as well as the curviness of the tension line (white graph line) at any given point in time. The halo is a snapshot of a blinking cursor indicating playback position. 96

5-3 Screenshot of Hyperscore Version 4. This version allowed users to enter their own motives instead of using pre-composed ones. 97

5-4 A Version TS Hyperscore rendering of the exposition of the first movement of Beethoven's Fifth Symphony. 98

5-5 A motive window and sketch window with lines drawn in the color associated with the motive. 99

5-6 Musical realization of the Hyperscore piece shown in Figure 5-5. 99

5-7 Colored droplets representing individual chords. The colors were mapped to different chord qualities. 100

5-8 An empty Hyperscore sketch window showing the harmony line. The height or depth of the point indicates the key to which the section modulates (indicated by the text overlay). 102

5-9 Example of a Hyperscore piece that has been harmonized. 103

5-10 Musical realizations of motives in Figure 5-9. 104

5-11 The unharmonized musical realization of the Hyperscore piece shown in Figure 5-9. 106

5-12 Musical realization of the harmonized Hyperscore piece shown in Figure 5-9. Roman numerals indicate the chords generated by the harmony line. 108

5-13 Three versions of the latter part of the Hypscore piece shown in Figure 5-9. (A) does not have any harmonization applied. (B) is harmonized in the standard way. (C) Shows the effects of another harmonization option that allows users to selectively unharmonize lines within the current tonal context. 109

6-1 Example from Experiment 1 showing meter changes. 116

A-1 Q01 (Composed for the experiment by the author.) 117

A-2 Q02 (Composed for the experiment by the author.) 117

A-3 Q03 (Excerpt from Beethoven's First Symphony.) 117

A-4 Q04 (Composed for the experiment by the author.) 117

A-5 Q05 (Excerpt from Bach's organ transcription of Vivaldi's C major concerto.) 118

A-6 Q06 (Composed for the experiment by the author.) 119

A-7 Q07 (Composed for the experiment by the author.) 119

A-8 Q08 (Excerpt from Schönberg's Klavierstück, Op. 11, No. 1.) 119

A-9 Q09 (Composed for the experiment by the author.) 119

A-10 Q10 (Excerpt from Brahm's Piano Concerto No. 2.) 120

B-1 Graph showing onset frequency values over time of Q10 (Brahms excerpt). . 121

B-2 Graph showing pitch height graph of Q10 (Brahms excerpt). 121

B-3 Harmonic tension graph of Q10 (Brahms excerpt). The x-axis represents time, and the y-axis, tension values. 121

B-4 Prolongational reduction of excerpt of Q10 (Brahms excerpt). 122

B-5 Chart showing harmonic tension calculations of Q10 (Brahms excerpt). δ is the sum of i, j, and k. 123

B-6 First page of analysis showing how melodic expectation values are calculated for Q10 (Brahms excerpt). i values consist of direct note-to-note level expectations, and j values consist of high-level patterns based on salience and metrical placement. s represents stability ratings, m represents mobility ratings, p represents proximity ratings, and d represents direction ratings. The values between the two staves are the final melodic expectations ratings for each note. 124

B-7 Melodic expectation graph for Q10 (Brahms excerpt). 124

B-8 Loudness curve of Q10 (Brahms excerpt). The blue line is the perceived loudness in dB produced by Jehan's psychoacoustic model, and the red line is the smoothed version used for the feature graph in the analysis of Experiment 2. 124

CHAPTER ONE
Introduction

"*Dimidium facti qui coepit habet.*"

– Horace

Music is structured sound. Through parsing and interpreting these structures, listeners arrive at a musical experience that is highly personal. This experience can range from active interest to stronger emotional responses like happiness and sadness. How these constituent parts translate into something as high-level as affect 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 subjective, emotional response.[1]

To a greater or lesser extent, this phenomenon is evident to listeners and is relatively easy to define in qualitative terms. For example, increasing tension can be described as a feeling of building excitement or impending climax, or an increase in uncertainty, while decreasing tension can be described as a feeling of relaxation, resolution, or fulfillment. It can also be defined in terms of changing expectations, realized or denied. The difficulty lies in formalizing and quantifying this definition. While tension is a fundamental concept in theories of Western music, there exists no universal theory that describes how disparate musical features combine to produce a general feeling of tension.

[1]This work is concerned with what Meyer calls the "absolutist" as opposed to the "referential" meaning of music [Meyer, 1956]. It only considers what is within in the context of the musical work itself, not extramusical concepts.

1.1 Musical parameters

In most types of music throughout the world, sound dimensions such as pitch, duration, loudness, and timbre are categorized and organized into ordered relationships. Well studied, particularly in the case of tonal hierarchies in Western music, is the fact that listeners implicitly recognize these relationships common to music from their culture, even if they're incapable of naming these rules or structures. Trained musicians learn to identify these musical rules through study, much like a child who understands the rules of grammar in his or her native tongue but is unable to name the rules explicitly until learning them in school [Bigand, 1993].

Many of the features of music perception that appear early in development are also found in the universal features of music across cultures. Dowling and Harwood (1986) suggest that several features are common to virtually all of the world's musical systems. These include (1) the octave as a basic principle in pitch organization, (2) a logarithmic pitch scale, (3) discrete pitch levels, (4) five to seven unequally spaced pitches in a scale, (5) hierarchies of stability for pitch, and (6) melodic contour as an important organizational device [Justus and Bharucha, 2002].

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.

Research has shown that musical expectancy is governed by several factors, including melodic interval size [Carlsen, 1981] [Unyk and Carlsen, 1987], melodic contour [Boltz and Jones, 1986], rhythmic features [Boltz, 1993] [Jones, 1987] [Jones and Boltz, 1982] [Jones et al., 1993] [Schmuckler, 1990], and tonal and harmonic structures [Abe and Hoshino, 1990] [Bharucha and Stoeckig, 1986] [Bharucha and Stoeckig, 1987] [Schmuckler, 1989] [Schmuckler and Boltz, 1994] [Lerdahl and Krumhansl, 2003] ([Bigand et al., 1996]). In another study [Krumhansl, 1996], subjects were asked to move a slider in response to the degree of tension they heard in a movement from a Mozart piano sonata. 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.

1.2 Objective and approach

This thesis presents 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. Previous studies focused on a small subset of features (particularly harmony and melodic contour or some combination of both) without attempting to understand how other parameters or expressive features such as loudness and tempo directly affected listeners' sensitivity to them.

The model described here is based on a significant amount of empirical data gathered in two experimental settings. 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.

1.3 Motivation

Aside from the value of having a model that attempts to define musical tension from a new perspective, this work also provides a foundation for implementation and incorporation into applications for computer-assisted analysis of music as well as computer-assisted and purely automated composition. Composers commonly manipulate tension and relaxation in their music to influence a listener's response to the music. How this is done is often critical to the effectiveness of a piece. Applications that help both professionals and novices compose music could be improved with the incorporation of a model that can analyze tension or generate music given a tension description and some musical material. In the case of professional composers, having a way of generating exploratory material based on tension descriptions could be useful, particularly for sketching out ideas. For novices looking for help and direction, such a tool could be highly instructive or even critical to producing original, creative work.

The author's past research has been in the domain of computer-assisted composition systems. Hyperscore, one such example, is discussed as a possible application for the tension model. Hyperscore is an application that facilitates composition through the intelligent mapping of musical features to graphical abstractions, providing a visual analogue for what is happening structurally in the music. Users without musical training are able to compose with Hyperscore because it abstracts away complex musical features such as harmony and counterpoint. Hyperscore is a forward (or synthesis) model: people are given the tools to create music. The musical feature space is reduced by taking away the focus from certain aspects (for example, removing the requirement to read staff notation, a big barrier to entry for many potential composers) and providing certain knobs that allow for easier use (such as abstraction of harmony).

The original prototype of Hyperscore was based on the idea that a graphical curve is an intuitive way to represent musical tension. A person using Hyperscore could draw a “tension” curve to describe the arc of a piece, and would provide the system with some melodic material which would then be used by the system to generate music. In subsequent versions of Hyperscore, this idea eventually evolved into something less encompassing—the general tension curve became a harmonic tension curve and the compositional process became less automated.

While the final version of Hyperscore successfully assisted users in composing original and often highly sophisticated pieces of music, anecdotal observation made it clear that the ability of the system was limited. In order for it to evolve into an intelligent tool capable of helping users on multiple musical and creative levels, it needed to “understand” the concept of musical tension. For this to be possible, an analytical, first-principles approach to understanding and quantifying tension was necessary. The result is the work presented in this thesis.

1.4 Overview of Thesis

Chapter 2 provides an overview of work related to musical tension in the fields of music theory and music cognition. Lerdahl’s tonal tension model and Margulis’ melodic expectation model are discussed in particular detail because they are used in the data analysis.

Two experiments designed to collect the data are described in Chapter 3. Experiment 1 was a web-based study that collected data from nearly 3000 subjects. Most of the musical examples were short and the questions were in multiple choice format. Experiment 2 was a more traditional music cognition study that collected real-time responses to more complex musical stimuli including excerpts from pieces by Beethoven, Brahms, and Schönberg. Subjects were asked to move a slider up and down to indicate changes in tension.

Chapter 4 describes the analysis of data collected from the two experiments. For Experiment 1, statistics are presented comparing how subjects responded to different musical features as well as how responses between musically experienced subjects and musically inexperienced subjects differed. For Experiment 2, descriptive graphs were generated for each excerpt in order to quantify all relevant musical features. These graphs were then used to perform regression analysis on the empirical data. The first part of each excerpt was used as training data in an attempt to come up with either a linear or nonlinear model that accurately predicted the remaining data. Correlation between features and the empirical data is discussed as well as differences between responses of musicians and non-musicians.

Chapter 5 discusses possible applications of the model. Early versions of Hyperscore and how they attempted to address the issue of musical tension are described. The final version of Hyperscore is discussed in detail with a particular focus on how the harmonic tension line was implemented. Ideas for how the general tension model could be incorporated into Hyperscore are outlined.

Chapter 6 summarizes the findings and outlines the next steps for improving the model.

CHAPTER TWO

Prior Work

"*Et nunc, reges, intelligite; erudimini qui iudicatis terram.*"

– Psalm 2:10

The study of tension and resolution is central to Western music theory and has been a matter of ongoing debate for both music theorists and cognitive psychologists. Schenkerian analysis, probably the most original and influential analytical theory of the twentieth century, was developed over a period of 40 years [Schenker, 1906] [Schenker, 1935]. It suggested a new level of logic in tonal music, arguably becoming the dominating music-theoretical paradigm of the second half of the 20th century. Schenkerian analysis has influenced the most recent theoretical developments including Lerdahl and Jackendoff's *Generative Theory of Tonal Music* (1983). Lerdahl and Jackendoff developed an influential theory that formalized the listener's understanding (i.e. mental representation) of tonal music. Much of what they investigate is psychological in nature. The formulation of their theory coincided with and provided stimulus to the growth in studies of what has since become an independent, mostly experimentally based discipline: music cognition. [Palisca and Bent, 2003]

2.1 Schenkerian analysis

Schenkerian analysis is a technique for answering the question, "How are harmonic progressions directed towards a goal?" Schenker saw music as the temporal unfolding, or prolongation, of a major triad, and composition as the large-scale embellishment of a simple underlying harmonic progression—in essence a massively-expanded cadence. His method is particularly designed to show the special importance that large-scale linear formations have in the creation of directed motion toward harmonic goals. Schenker attempted to combine harmonic analysis with the principles of strict counterpoint in such a way as to overcome the limitations of each and show that even artistry and taste were not wholly inaccessible to rational explanation. [Bent, 1987]

There are three levels of analysis: the foreground, consisting of the surface details, the middleground consisting of mid-level deep structures, and the background or *ursatz*, the fundamental structure. Most of the real analysis takes place in the middleground (Figure 2-1). The *ursatz* also includes a descending upper voice called the *urlinie*.

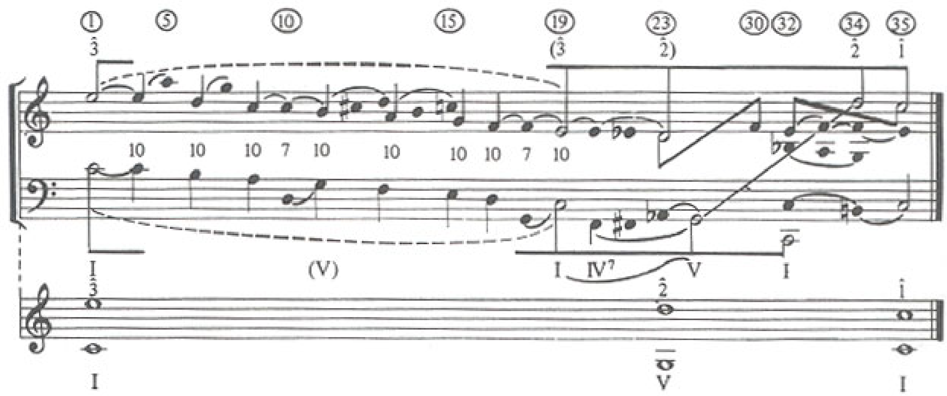


Figure 2-1: Middleground analysis of Bach Prelude in C Major, *Well-Tempered Clavier* Bk. I [Forte and Gilbert, 1982].

This structure fits many tonal pieces although it is an abstraction far removed from the listener's experience. The lack of direct correlation between score and analysis does create certain difficulties in judging or verifying Schenkerian interpretations. If Schenkerian analysis explains how people normally hear music, why would it be necessary to learn a new way of hearing music in order to carry out the analysis? Schenker believed that the most fundamental part of musical experience is directed motion towards an endpoint, and that almost all music exhibits more or less the same structure at this background level. However, it may be argued that listeners do not work backwards—it seems unlikely that any such unconscious understanding exists. Nevertheless, Schenker's ideas have been fertile ground for both music theorists and composers of computer-based music.

2.2 Generative theory of tonal music

Fred Lerdahl and Ray Jackendoff's generative theory of tonal music (GTTM) [Lerdahl and Jackendoff, 1983] attempts to characterize the way listeners perceive hierarchical structures in tonal music by developing a grammar based in part on the goals, though not the content, of generative linguistics. This grammar is intended to model musical intuition and takes the form of explicit rules that assign or "generate" structures that listeners unconsciously infer from the physical signal (or "musical surface") of a piece. These principles define components of musical intuition that are hierarchical in nature:

- **grouping structure** - segmentation of music into motives, phrases, and sections.
- **metrical structure** - hierarchy of alternating strong and weak beats.
- **time-span reduction** - hierarchy of structural importance of pitches with respect to their position in the grouping and metrical structures.
- **prolongational reduction** - hierarchy that expresses harmonic and melodic tension and relaxation (this component is the closest to Schenkerian reduction).

Each of these structures is described formally by a separate component of the musical grammar and within each component there are three rule types. Well-formedness rules specify the possible structural descriptions. Transformational rules apply certain distortions to the otherwise strictly hierarchical structures provided by the well-formedness rules. Preference rules[1] do the major work of analysis within the theory by picking structural descriptions that correspond more closely to experienced listeners' hearing of any particular piece.

2.3 Lerdahl's tonal tension model

Lerdahl has significantly extended GTTM by developing a precise model of how a piece is heard as it unfolds in terms of paths in pitch space 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. He has developed a quantitative model of these intuitions.

2.3.1 Perception of tonal hierarchies

Cognitive approaches like Lerdahl and Jackendoff's theory emphasize the importance of tonal function (see also [Bharucha, 1984], [Krumhansl, 1990], [Bigand, 1993], and [Lerdahl, 2001]), while more perceptual theories underline the psychoacoustic features of chords ([von Helmholtz, 1877], [Mathews et al., 1987], [Parncutt, 1989], [Roberts and Shaw, 1984]).

[1] More recently David Temperley has extended and elaborated the concept of preference rules [Temperley, 2001].

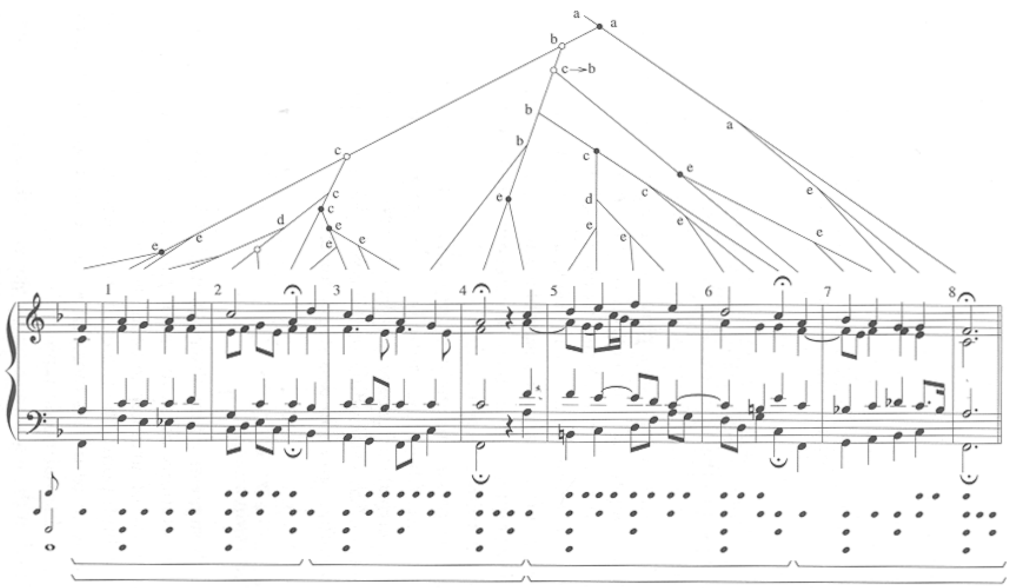


Figure 2-2: GTTM analysis of a Bach chorale showing the prolongational tree, metrical grid, and grouping analysis [Lerdahl, 2001].

It has been established that Western listeners have an implicit knowledge of tonal hierarchies [Krumhansl and Shepard, 1979] [Krumhansl and Kessler, 1982]. Experiments have shown that even inexperienced listeners have internalized within-key hierarchies and between–key distances ([Dowling and Harwood, 1986], [Francès, 1958], [Krumhansl, 1990]). In one study [Krumhansl and Shepard, 1979], all seven notes of the C major scale were played followed by one of the twelve notes of the chromatic scale. Subjects had to rate how well the last note completed the previous notes. The results indicted that the notes judged to provide the least adequate completion were the ones outside the key of C major, confirming the implicit knowledge of various keys. Among the pitches within the key of C there was also a clear hierarchy; all listeners felt the tonic pitch C provided the best completion followed by G and E. Other studies have explored the implicit learning of tonality and have obtained similar results [Bharucha, 1987].

On the psychological level, differences in the hierarchical level of pitches generate patterns of musical tension and relaxation. Notes low on the tonal hierarchy produce strong tensions that then can be resolved by the arrival of more stable notes [Bigand, 1993]. Likewise chords in a tonal context convey tension and relaxation through resolution (or lack of resolution). For example, a dominant seventh chord generates the expectancy that it will resolve to the tonic. A chord that realizes or confirms this expectation creates a feeling of relaxation. Musical tension and musical expectancy may be viewed as two co-occurring phenomena.

Therefore, we may assume that studying either should provide a complimentary insight about similar or related aspects of music cognition [Bigand et al., 1996].

2.3.2 Lerdahl's criteria for tonal tension

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 [Lerdahl, 2001] [Lerdahl and Krumhansl, 2003]:

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

The first component is equivalent to GTTM's prolongational reduction and can be represented in tree notation (see Chapter 4 and Appendix B for examples of analyses).

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 is represented by three embedded spaces, the first two representing within-key hierarchies, and the third one between-key distances. The first space is pitch-class proximity. It consists of five levels: chromatic, diatonic, triadic, fifth, and root. The second space is chord proximity within a key (or region), and the third space distances between keys or regions. The diatonic chord distance rule is thus defined as follows:

$$\delta(x, y) = i + j + k \tag{2.1}$$

where $\delta(x, y)$ = the distance between chord x and chord y, i = the number of steps between two regions on the chromatic fifths circle (i.e. distance between two chords with regard to key), j = the number of steps between two chords on the diatonic fifths circle (distance with regard to chord function), and k = the number of distinctive pitch classes in the basic space of y compared to those in the basic space of x.

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 rule is defined as follows:

$$T_{diss} = f + g + h \tag{2.2}$$

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

The fourth component describes how listeners experience the relative pull of pitches toward other pitches in a tonal context; for example, a leading tone has a strong attraction to the tonic. Bharucha's notion of anchoring provides a psychological account of this phenomenon of less stable pitches tending to resolve on subsequent, proximate, and more stable pitches [Bharucha, 1984]. The melodic attraction is defined as follows:

$$\alpha(p_1, p_2) = \left(\frac{s_2}{s_1}\right)\left(\frac{1}{n^2}\right) \tag{2.3}$$

where $\alpha(p_1, p_2)$ = the melodic attraction of pitch p_1 to p_2, s_1 = the anchoring strength of p_1, s_2 = anchoring strength of p_2 in the current configuration of the basic space, and n = the number of semitone intervals between p_1 and p_2.

The final tonal tension equation is defined as the sum of all four components:

$$T_{global} = \delta(x, y) + T_{diss} + \alpha(p_1, p_2) + t \tag{2.4}$$

where t is the inherited tension value derived from a GTTM prolongational analysis.

An ongoing, large-scale study has already resulted in empirical support for Lerdahl's tonal tension model. See [Lerdahl and Krumhansl, 2003] for more details.

2.4 Melodic structures

Veridical expectations represent specific knowledge about the way a particular piece progresses, for example, a listener's knowledge of Beethoven's Seventh Symphony after listening to it 100 times. On the other end of the spectrum are expectations automatically predicted from patterns that can be implicitly extrapolated from an extensive corpus of music, as in the case of tonal hierarchies; these are classified as schematic expectations.

Within this category, expectations can range from "shallow" to "deep." For example, the ability to identify cadences in the style of Mozart is learned through listening to various examples of Mozart's music. After hearing enough examples, certain harmonic and melodic patterns become expected when these cadences occur. These expectations can be violated and give rise to a feeling of tension. Unlike the implicit understanding of tonal hierarchies, however, these expectancies are considerably more shallow because new examples of such cadences can more easily modify previous expectations and the corpus from which the patterns are learned is far smaller (e.g. works only by Mozart versus all Western music). [Margulis, 2005]

At the very lowest level are "deeply schematic" structures. These structures are universals that not only apply to Western music but all types of music. Expectations arising from these structures still exert an influence on listeners' expectations, even if they go against veridical knowledge of a piece [Tillmann and Bigand, 2004a] [Justus and Bharucha, 2001].

2.4.1 Implication-Realization theory

Narmour's implication-realization (I-R) theory [Narmour, 1977] [Narmour, 1992] concentrates almost exclusively on deeply schematic, note-to-note melodic relations. In effect, the analytic method he purposes is intended to track and reflect fluctuations in affect that arise from the validation or invalidation of the moment-to-moment subconscious inferences made by a listener concerning the way a piece of music unfolds [Cross, 1995]. The Implication-Realization theory was conceived as an alternative to Schenker's idea of prolongation. It is an extension of the work of Leonard Meyer, who used Gestalt principles of perceptual organization to account for the way listeners' expectations are generated and manipulated [Meyer, 1956].

Narmour's main hypothesis is that any two successive pitches imply a third pitch. Any pair of melodic pitches transmits separate intervallic and registral messages to the listener that imply *continuation* when the interval is small (e.g. a major second going up followed by another major second going up) or *reversal* of direction and *differentiation* of interval size when it's large (e.g. a major sixth going up followed by a minor second going down). These universal bottom-up principles are also combined with a top-down approach regarding learned or style-specific principles. These concepts are particularly useful in codifying the analysis of melodic contour and how it affects the listener's perception of tension.

While Narmour's I-R model clearly categories melodic segments, it only secondarily denotes the theorized expectedness of each segment—expectations are given only every third note sometimes. The Implication-Realization theory's foundational reliance on Gestalt principles using notions of "good" and "best" to describe groupings is also vague. Furthermore, the definition of similarity implying another similarity versus differentiation implying differentiation seems inconsistent; e.g. although pitches nearer to one another are closer in frequency, that doesn't necessarily imply that they sound more "similar." The notational system in I-R is also opaque—it shows structures but does not categorize them according to their degree of realization or denial. As a result, raw taxonomy obscures the main intent. [Margulis, 2005]

2.4.2 Margulis' model of melodic expectation

Margulis' model of melodic expectation [Margulis, 2005] extends the I-R model while addressing some of its problems. The key connections between the I-R model and Margulis' model are a commitment to an account of purely schematic expectations and the centrality of proximity and distance. An additional factor, stability, is derived from Lerdahl's tonal pitch space and melodic attraction models, thus addressing tonal relationships as well as melodic interval relationships. Aside from incorporation of tonal relatedness (stability) and treatment of hierarchy and factors such as meter and harmony, Margulis' model provides quantitative predictions of expectedness.

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

$$z = (smp) + d \tag{2.5}$$

where z is the amount by which pitch x is expected to follow pitch y, $s =$ the stability rating of x (see Table 2.1), $m =$ the mobility rating of x (2/3 if x repeats y and 1 in all other cases), $p =$ the proximity rating of x (see Table 2.2, and $d =$ the direction rating of x (see Table 2.3).

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	Chord root (Chromatic pitches

Table 2.1: Table showing stability ratings and the requirements which govern them.

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

- A secondary chord has occurred, in which case the context shifts to I in the tonicized key.

- The melody note constitutes a nonchord tone with respect to the current harmony, in which case the context shifts to the current chord in the current key.
- The previous melody note was the seventh of the chord, in which case its lower diatonic neighbor is promoted to the highest stability rating.
- A strong predominant chord such as an augmented sixth or Neapolitan has occurred, in which case the context shifts to V in the current key.

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
$\geq$14 (M9)	0.01

Table 2.2: Table showing proximity ratings. Proximity ratings increase as semitone distance increases, reflecting the expectation that pitches will proceed to relatively nearby continuations.

Interval size 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
$\geq$10 (m7)	75 for reversal

Table 2.3: Ratings given to events that lie in the direction implied by the interval between the preceding two notes.

Deeply schematic structures are not just dependent on the progression of a few notes, but a larger context as well. Thus the final values are calculated by averaging the weighted values of expectations at each hierarchical level defined by a GTTM time-span reduction. 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 levels with spans of two-second duration, receive a weight of 5. Ratings at levels with time-span durations from two up to six seconds receive a weight of 2. No levels with durations longer than six seconds are considered. Thus the formula for overall expectedness of a note is as follows:

$$\frac{\sum_{i=1}^{n} w_i z_i}{\sum w_i} \tag{2.6}$$

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

Margulis' model also addresses issues concerning affect by describing three kinds of responses to changes in tension:

- **surprise-tension** correlates inversely with expectancy ratings, e.g. highly predictable events generate little surprise-tension. Listener's associated experience is a subtle feeling of intensity or dynamism.
- **denial-tension** correlates directly with implicative denial—it's the difference between the maximally expected pitch and the expectancy rating of the pitch that actually occurred. Listener's associated experience is a sense of will, intention, or determinedness.
- **expectancy-tension** pertains to the strength of expectation generated by an event about future events. It is directly proportional to expectancy rating of the most-expected possible continuation. Listener's associated experience is an impression of strain or yearning.

A slightly modified version of Margulis' model is utilized in the analysis of empirical data from Experiment 2. See Chapter 4 for examples of the analysis process and how the model is used.

2.4.3 Melodic contour and tension

Melodic expectancy is one important factor that contributes to the experience of musical tension. The results of the comparison between Margulis' tension predictions and Krumhansl's tension data [Krumhansl, 1996] suggest that expectancy-based tension forms an important part of the generation of overall experiences of tension. In other studies, the responses of listeners in continuity-rating and melody-completion tasks have provided empirical support for some of the principles described in Namour's and Margulis' models [Krumhansl, 1995] [Cuddy and Lunney, 1995] [Thompson et al., 1997] [Schellenberg, 1996], [Schellenberg, 1997].

There also exists a body of research that has examined the memory and mental representation of specific melodies. These include studies of melody recognition when transposed to a new key, suggesting that melodic fragments are encoded with respect to scales, tonal hierarchies, and keys ([Cuddy and Cohen, 1976] [Dewar et al., 1977] [Cuddy et al., 1979] [Cuddy et al., 1981] [Cuddy and Lyons, 1981]. Melodies are processed and encoded not only in terms of the musical scale in which they are written, but also in terms of their melodic contour. When discriminating between atonal melodies, where there is no tonal hierarchy, listeners rely mainly on the melodic contour [Dowling and Fujitani, 1971]. Furthermore, within tonal contexts, melodies and their transpositions that alter particular intervals by semitones to preserve the key are just as easily confused as exact transpositions to new keys [Dowling, 1978]. One explanation of this result is that the contour, which is represented separately from the specific interval information, is processed relative to the framework provided by the scale [Justus and Bharucha, 2002].

Due to these factors, the analysis of musical excerpts used in Experiment 2, described later in Chapter 4, employs separate descriptions of melodic expectancies and pitch height.

2.5 Some comments on tension outside the tonal context

Tension in atonal music—by its very definition—cannot be measured with regard to tonal structures. While deeply schematic melodic rules described by Narmour are still applicable, melodic attraction rules dependent on harmonic functions are not.

Lerdahl proposes a flat pitch-space distance rule which measures distance between two chords by comparing the number of interval classes and pitch classes in each chord [Lerdahl, 2001]. He also suggests that listeners infer an event hierarchy based on the relative salience of events. An event is deemed to have greater structural importance ([Lerdahl, 1989] if it meets the following criteria:

- attacked within the region (i.e., within the time span) [3]
- in a relatively strong metrical position [1]
- relatively loud [2]
- relatively prominent timbrally [2]
- in an extreme registral position [3]
- relatively dense [2]
- relatively long in duration [2]

- relatively important motivically [2]
- next to a relatively large grouping boundary [2]
- parallel to a choice made elsewhere in the analysis [3]

(The numbers in brackets indicate the relative strength of the condition.)

The findings of Dibben (1999) indicate that there is empirical support for theories of atonal prolongational structure. She investigates the perception of structural stability in atonal music through two experiments. The first experiment suggests that listeners are greatly influenced by metrical and durational structures and hear atonal music in terms of the relative structural importance of events. The second experiment suggests that listeners infer relationships of relative structural stability even in the absence of clear rhythmic, timbral, dynamic, and motivic information. They are influenced by a number of other factors including horizontal motion and dissonance in particular.

Beyond discrete-pitch-based analyses, the feature of timbre is of particular importance. In contrast to cognitive approaches to tension, psychoacoustical models predict the strength of pitch relationships between successive chords without considering the listener's implicit knowledge of tonality [Bigand et al., 1996]. Pressnitzer et al. (2000) suggest that psychoacoustic roughness plays an important role in the perception of musical tension in the absence of tonality, dynamics, or rhythmic cues. Roughness is an auditory attribute that has been proposed as a sensory basis for musical consonance within the tonal system [von Helmholtz, 1877].

2.6 Modeling expressive characteristics of music

The model described in this thesis takes into account the expressive features of dynamics and tempo variation as well as inherently musical, higher-level features such as harmony and melody. While empirical studies [Krumhansl, 1996] suggest that performed tempo and dynamics have remarkably little effect on tension judgments, perhaps more extensive and less subtle variations in loudness and tempo would produce a different result (see Chapter 4). In any case, the reverse doesn't seem to apply: it appears that tension judgments stemming from musical features such as melodic and harmonic structures influence the way people expressively interpret music.

This relationship between dynamics and tempo and other musical features is particularly well presented in the work of Widmer, who combines machine learning with music theory to model musical expression. From a machine-learning perspective, his objective is to study various types of imprecise and incomplete domain knowledge and ways of using them to form better hypotheses. From a musical perspective, he tries to answer the following questions: What kind of general musical knowledge do listeners possess? How can it be formalized? What is the relationship between this knowledge and expressive performance? What structural aspects of music determine or influence the acceptability of performances? [Widmer, 1998]

The parameters used by Widmer to analyze expressive musical performances were dynamics (*crescendo* vs. *diminuendo*) and timing (*accelerando* vs. *ritardando*). Training data for the learning system were melodies (i.e. sequences of notes), along with actual (human) performances of them. The system's task was to learn when and how to apply dynamics and tempo variations to given melodies. The learned rules then allowed the system to play new pieces expressively. The approach was guided by three important assumptions:

- Expressive interpretation is in part the communication of the understanding of musical structures.

- Some knowledge of musical structures can be described explicitly. This background knowledge is based on Lerdahl and Jackendoff's generative theory of tonal music and (more loosely) on Narmour's implication-realization theory.

- The level of symbolic notes (as opposed to sound spectra) is a reasonable abstraction level.

Widmer's approach tested whether expressive principles could be learned at all by a machine and whether the addition of explicit musical knowledge could help the learning process. There were three approaches taken: the first was note-based and didn't use the background knowledge (for comparison purposes); the second approach was also note-based but did use the background knowledge; the third approach abandoned the note level and tried to learn expression rules directly at the level of the musical structures expressed by the knowledge base. The use of a knowledge base in the latter two cases led to a clear improvement of learning results. The third approach worked the best, suggesting that learning at the abstraction level of musical structures produces performances of much higher quality than rules learned and formulated at the note level [Widmer, 2000].

One caveat Widmer mentions is the problem that "correct" musical interpretation is not definable, since it's an aesthetic judgment. However, both the qualitative and quantitative results of his experiments indicate how higher level musical structures are essential in understanding how humans perceive music and provide yet more evidence in support of theories such as GTTM.

2.7 The role of prior work in defining a new model

Since the model described in this thesis 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 and Margulis' models, 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 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.

CHAPTER THREE

Method and Experimentation

"*Quot homines, tot sententiae.*"

– Terence

Two studies were conducted in order to collect empirical data required to inform the tension model. Experiment 1 was a web-based study designed to attract as many subjects as possible (as of December 2005, it had collected data from nearly 3000 subjects). The musical examples and corresponding response choices for Experiment 1 were mostly short and simple. Experiment 2, in contrast, had far fewer subjects but more varied, complex examples. Unlike the first experiment, real-time responses to stimuli were recorded.

3.1 Experiment 1: Web-based study

3.1.1 Procedure

The format for this experiment was modeled after the Moral Sense Test, a web-based study designed by Marc Hauser and his research group at Harvard [Hauser et al., 2005]. Over 40,000 people have taken this test due to its prominence online and its welcoming format. Because so many subjects have taken part in the study, the investigators can give a small subset of all the possible questions to each participant.

The web-based interface[1] for Experiment 1 was implemented in collaboration with Josh McDermott, a researcher in the Brain and Cognitive Sciences department at MIT. The collaborative test was officially entitled, "Music Universals Study." The first two parts of the exam were designed by McDermott and are not related to the research described here. The third and final part was the tension experiment. This part consisted of 11 questions selected from a set of 207 musical examples.

The test as a whole contained two surveys, one taken before the test and one taken afterwards. The first survey asked some technical questions (type of computer, operating system, and speakers used by the subject), and some questions about cultural background (country of origin, country where participant grew up, educational level). The survey at the end asked questions pertaining to prior musical experience (formal training in instrumental and/or vocal performance, composition, music theory). Following the initial survey, a series of tones ranging from very low to very high frequencies were played to help subjects calibrate their speakers properly. This was necessary since observers were not there in person to make sure subjects heard the stimuli with minimal distortion.

The tension section was introduced by the following text: *Many different aspects of music can contribute to the feeling that a piece of music is "going somewhere." This last experiment explores how these aspects combine to create feelings of tension and resolution. You will be played short a series of sound files and asked to choose a graphical representation that best fits how you feel the tension is changing in the music.*

Each musical example in the test was played twice. Subjects chose from a selection of curves graphically depicting changes in tension (Figure 3-1). The graphics were partially designed with Hyperscore in mind as a future application of the model (see Chapter 5 for details).

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.

3.1.2 Selection of musical examples

At the beginning of the test, subjects were presented with a sample question that was answered for them (Figure 3-2). This question was assumed to have an obvious answer (or at least as obvious an answer as possible given the subjective nature of the topic). 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.

[1] Implemented by Joe Presbrey with additional scripts by Rich Chan.

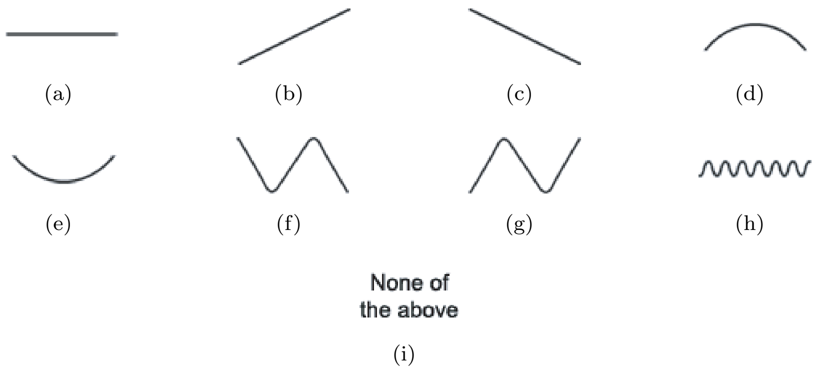
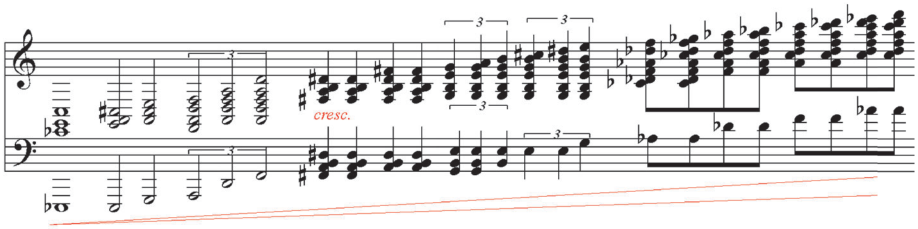


Figure 3-1: Response choices for Experiment 1.



(a)

(b)

Figure 3-2: (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.

The examples were 2 to 60 seconds long and recorded using piano, strings or 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). The volume scaling for examples with loudness changes was done in an audio editing program. Due to perceptual differences in loudness sensitivity depending on the frequency range (see Section 4.2.1 for further discussion on loudness perception) some examples were adjusted by hand to sound more consistent.

Most examples were recorded with both piano and string sounds, while only a small subset were recorded with unpitched percussion. This is due to the fact that most excerpts required pitched sounds (e.g. those with harmony or pitch height changes). Out of the 207 audio files that were used in the experiment, 102 were musically unique excerpts. All were recorded with piano, 85 with strings, and 20 with percussion. Some of the harmony examples only varied with regard to major or minor mode. If those pairs are counted once, the total number of unique examples is 85.

Each example file was designated by a group number (Table 3.1). Examples within the same group were very similar and highlighted the same features in slightly different ways. Figure 3-3 shows several examples from group 001.

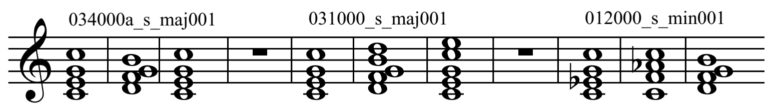


Figure 3-3: Three musical examples from group 001.

The sample question (Figure 3-2) was taken from group 011. Out of the 11 following questions, only the first (taken from group 013) was not chosen randomly. This question was used as a control (Figure 3-5); like the sample question, it was considered as close to an "easy" question as possible (the presumed answer was that tension decreased). It could also be used as a measure for whether or not the subject understood what was being asked. A response specifying the opposite of what was expected might indicate that the subject's other responses were unreliable.

Questions 2 through 9 were selected by randomly choosing a group less than 111 and then randomly selecting an example in that group. If the example chosen had been recorded with different instrumental sounds, the timbre was selected randomly as well. Only one example per group could be chosen. As a result, previously selected group numbers could not be considered in subsequent questions. An example from group 111 was always chosen for question 10 and an example from group 112 was always chosen for question 11. These last two questions were different from the others. Instead of being composed specifically for the test, they consisted of all or part of an excerpt from Bach's organ transcription of Vivaldi's C Major concerto (see Figure A-5 in Appendix A for score). Excerpts selected included a large one-minute long section of the piece as well as various smaller sections that encompassed whole and partial phrases, transitions, and portions with altered tempo and dynamics.

One result of the unequal distribution of examples among groups as well as the method of selecting the group themselves, was an uneven distribution of responses per question. With so many subjects, however, this was not significant problem. Another reason for the unbalanced response distribution was due to some examples being added midway through the data collection process. These included examples in groups 020 and higher and mostly covered examples that featured rhythmic irregularity and parts of the Bach-Vivaldi excerpt.

Group	Num. examples	Description
001	7	Simple harmonic progressions with subtle changes in voicing
002	1	Changes in onset frequency
003	1	Changes in pitch height
004	1	Changes in harmony and pitch height corresponding
005	4	Changes in harmony against changes in onset frequency
006	1	Changes in harmony and onset frequency against loudness
007	4	Changes in harmony with differing degrees of changes pitch height
008	4	Different combination of changes in onset frequency, harmony, and pitch height
009	1	Changes in onset frequency, harmony, and pitch height against loudness
010	1	Changes in onset frequency, pitch height, and loudness against harmony
011	2	Changes in onset frequency, harmony, pitch height, and loudness all increasing in intensity
012	1	Changes in harmony, pitch height, and loudness against onset frequency
013	1	Onset frequency, harmonic tension, pitch height, and loudness all decreasing in intensity
014	1	Changes in onset frequency, harmony, and loudness against pitch height
015	1	Increasing dissonance
016	2	Changes in rhythmic regularity
017	1	Changes in loudness
020	2	Changes in tempo
021	2	Changes in tempo, some examples having notes with shorter articulation
022	9	Changes in rhythmic patterns (or rhythmic regularity) coupled with slight changes in onset frequency
023	4	Simpler changes in rhythm (than in group 022) coupled with slight changes in onset frequency
024	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

Table 3.1: Different groups of musical examples in Experiment 1.

3.1.3 Design of musical examples

The examples composed for Experiment 1 attempted to systematically isolate and combine the following parameters:

- onset frequency and tempo
- dynamics (loudness)
- pitch height
- harmony
- rhythmic irregularity

While these are not the only possible features that could be used to evaluate musical tension, they form an adequate foundation for analyzing score-based music.

For this experiment, onset frequency and tempo were placed in a single category. Tempo change was classified as a type of onset frequency change, though one that might be difficult to notate precisely. In Experiment 2, however, they were analyzed separately (see Section 4.2.1).

The main hypothesis being tested was that changes in each parameter would contribute 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 was used to provide a quantitative assessment of tension for each chord (for more details on how this was done, see Chapter 4). 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.

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. Figure 3-4 shows an example where 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 in what percentage of the participants are uncertain and to what degree they are uncertain, not necessarily in finding the one right answer.

3.1.4 Demographics

People from 108 different countries took part in the study. Figure 3-6 shows the percentage of subjects from the top 12 countries and Figure 3-7 presents a graphical view of subjects' countries of origin.

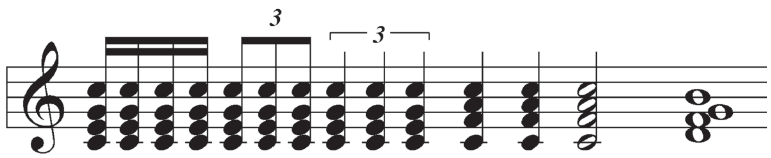


Figure 3-4: Example where two parameters, harmony and tempo, appear to be conflicting in their contribution to tension.

Figure 3-5: The first question in Experiment 1. All parameters are moving in the same direction: the harmony is relaxing, the tempo is slowing down, the volume is decreasing, and the pitch contour is falling consistently.

Figure 3-6: Map showing the origin of subjects in Experiment 1. Numbers shown include only the people who completed both the first and final surveys (image courtesy of Rich Chan and Google).

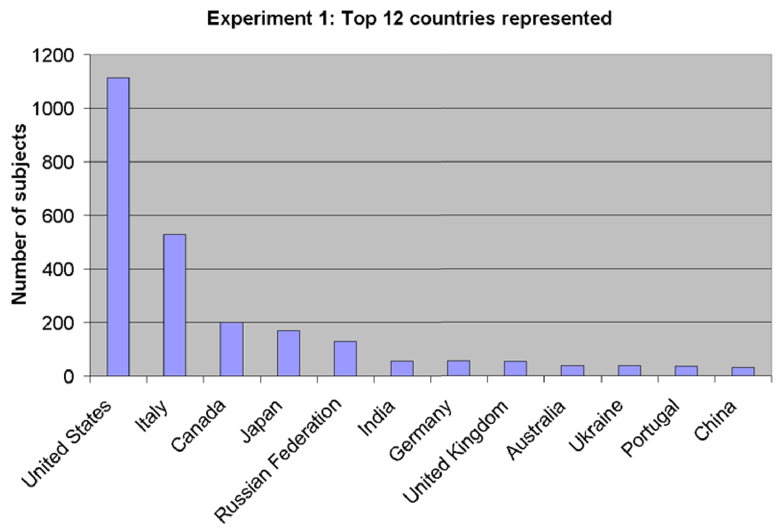


Figure 3-7: Top twelve countries of origin for subjects in Experiment 1.

3.2 Experiment 2

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. While it might be possible to overcome the latter limitation by presenting excerpts of varying length, taken from one long example—as in the case of the Bach-Vivaldi examples from Experiment 1—they still may not adequately convey the overall flow of a larger excerpt. Furthermore, the collection of data may be time-consuming, especially if long musical excerpts are used. It is because of these limitations that Experiment 2 was designed to measure responses to music in real time. This method provides a relatively efficient way of capturing a richer response to the data. [Toiviainen and Krumhansl, 2003]

3.2.1 Test format

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 composed 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: Schönberg Klavierstück, Op. 11 No. 1[2], the Bach-Vivaldi concerto from Experiment 1, 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. Seven participants had no more than five years of instrumental lessons or any other type of musical training. Three subjects recognized the Brahms example, one subject recognized the Bach-Vivaldi example, three subjects recognized the Beethoven example, and two subjects recognized the Schönberg.

Test subjects were presented with a computer interface written in C++ for Windows (Figures 3-8 and 3-9). 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. An increase in tension was defined as the feeling of building excitement, perhaps moving toward a climactic moment, or the heightening of feeling that something was unresolved. A decrease in tension was defined as a feeling of relaxation, rest, or resolution.

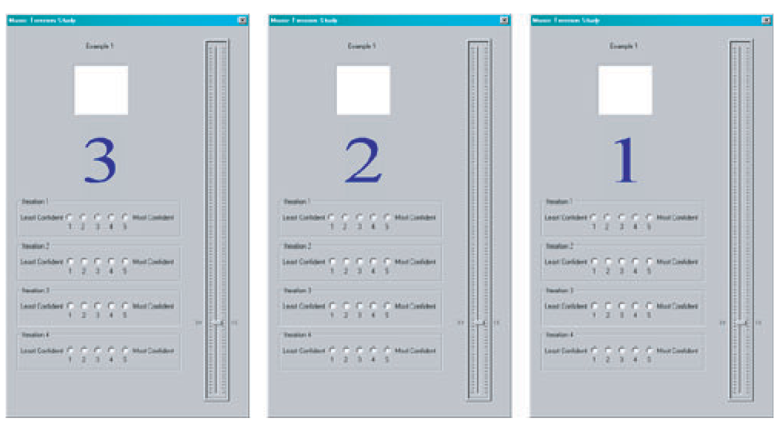


Figure 3-8: Experiment 2: countdown to prepare subjects for excerpt playback.

[2]This excerpt was slightly modified. The most significant change was the simplification of the dynamics.

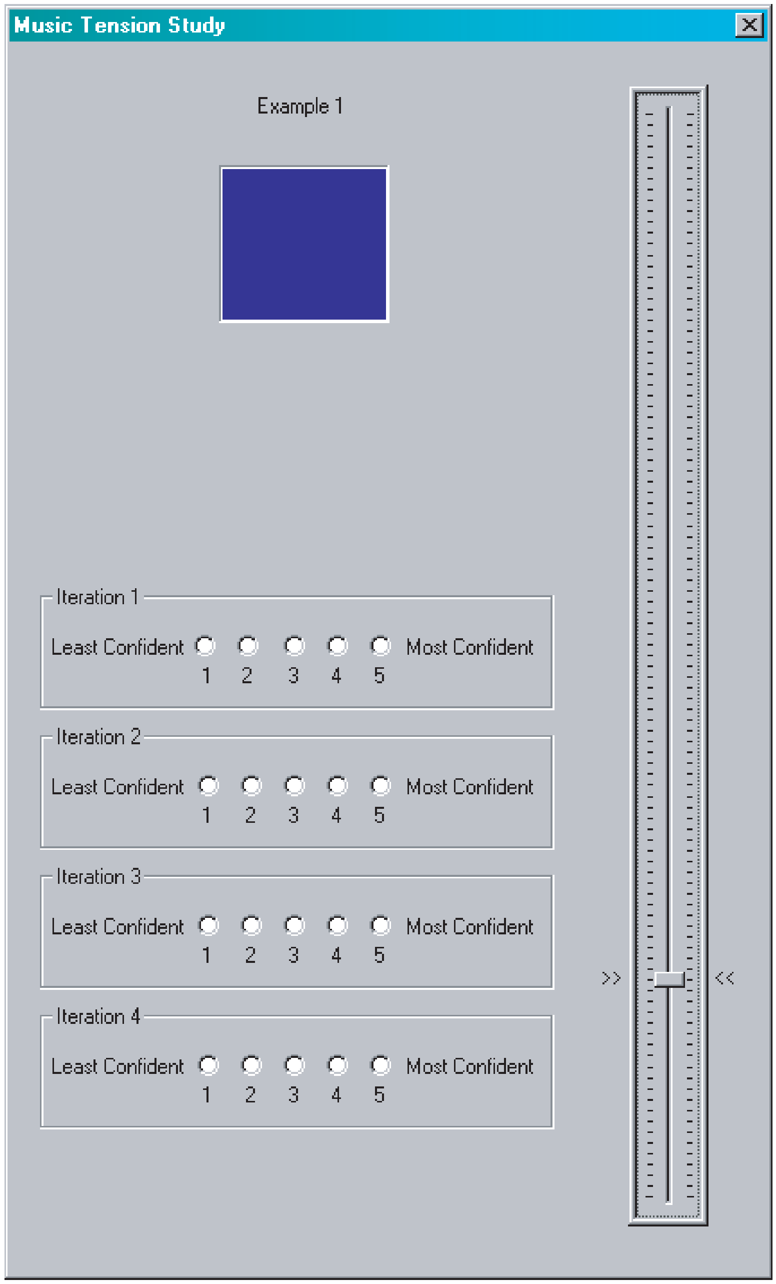


Figure 3-9: Screen shot of Experiment 2 interface in data recording/subject listening mode.

Two pairs of arrows marked the baseline on the slider. It was designated as the point of the most relaxation, and subjects were encouraged to use it as a starting point.[3] There was some space below it in case the subject felt that the lowest point of relaxation had been reached but then afterwards exceeded. The position of the baseline was decided after observing subjects participate in a pilot study that experimented with different interface configurations. It became clear that subjects starting in the middle tended to use only the top half of the slider, and subjects starting at the very bottom often felt they needed more space below the starting point.

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. The numbers, "3 2 1" would flash on the screen and then a white square on the interface would turn blue as the music started playing (Figures 3-8 and 3-9). During playback, as slider motions were being recorded, the square would remain blue. When the music stopped, the square would turn white again. The position of the slider was recorded at a sampling rate of 64Hz; this was the maximum rate possible on the computer used for the experiment. Slider values ranged from 0 to 100, the baseline being at 20.

Most of the subjects understood the instructions clearly and were able to respond as instructed. Two of the 35 data sets were thrown out due to confusion or admitted error on the part of the subject. In one case, a test subject felt the slider should have been horizontal instead of vertical and remarked that he found himself responding by moving the mouse left to right (which did nothing) instead of up and down. After careful observation of all 35 test subjects, it appears that the best sort of interface for this kind of data collection might be a device that allows only one degree of freedom without the added distraction of a visual component on a computer screen. For example, a slider on a mixer would have served as a better input device. In any case, the data collected did provide sufficient results regardless of possible future improvements in experimental design.

[3] In retrospect, they should have been required to start there.

CHAPTER FOUR
Model Results

"*Quod erat demonstrandum.*"

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.

4.1 Experiment 1 analysis

As discussed in Chapter 3, there were nine possible graphical responses subjects could choose from in Experiment 1 (Figure 3-1). 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 3-1(f) and 3-1(g) depict more changes in tension than subjects could track and recall with certainty.

4.1.1 Pitch height

In the case of pitch height and all subsequent parameters, subjects were considered to respond to the feature *if they selected a tension curve that matched the hypothesized tension curve for that feature*. Figure 4-1(a) shows an example that was used to determine how subjects responded to pitch height in isolation. Figure 4-1(b) depicts the hypothesized answer for the example.

(a) (b)

Figure 4-1: (a) Example where pitch height is the only feature changing. (b) The "correct" answer corresponding to pitch height for (a).

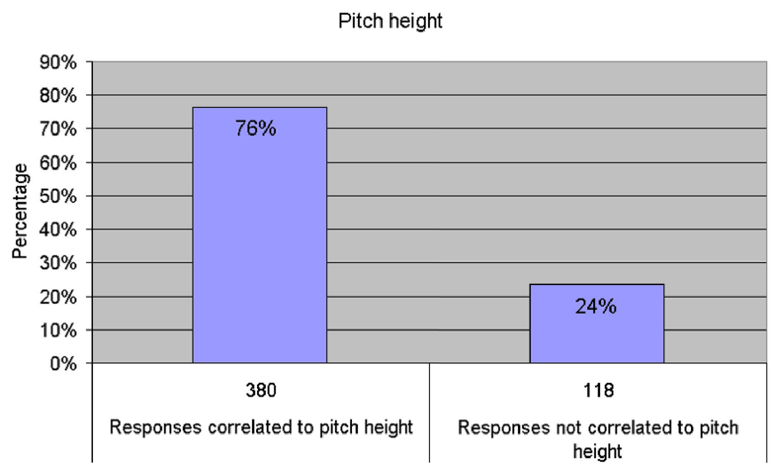


Figure 4-2: Graph showing subjects' responses to changes in pitch height.

Subjects responded very strongly to change in pitch height. There were a total of 498 responses to questions where pitch height was the only feature present. 76% chose an answer which corresponded with the feature. Only 24% of subjects chose any of the other eight responses (see Figure 4-2).[1] In other words, if a subject were given the example in Figure 4-1(a), the response was judged to correspond to pitch height if the curve shown in Figure 4-1(b) was chosen.

It might be argued that one reason for this strong response could be the clear mapping of pitch contour to graphical curve (at least more directly than for other musical features).[2]

[1] N.B. the number of total responses for each category of musical examples analyzed is present under each bar for all graphs.

[2] In the future, text-only response choices should be added as well. One reason for having graphical response choices was to make it easier for people with limited knowledge of English to take the study. Shortly before the data set for the analysis was downloaded, instruction comprehension questions were added to assess if subjects understood English well enough to understand the instructions. This makes it possible to evaluate the validity of data gathered from subjects selecting from all-text response choices in the future.

(a) (b) (c) (d)

Figure 4-3: Four examples with varying degrees of change in pitch height.

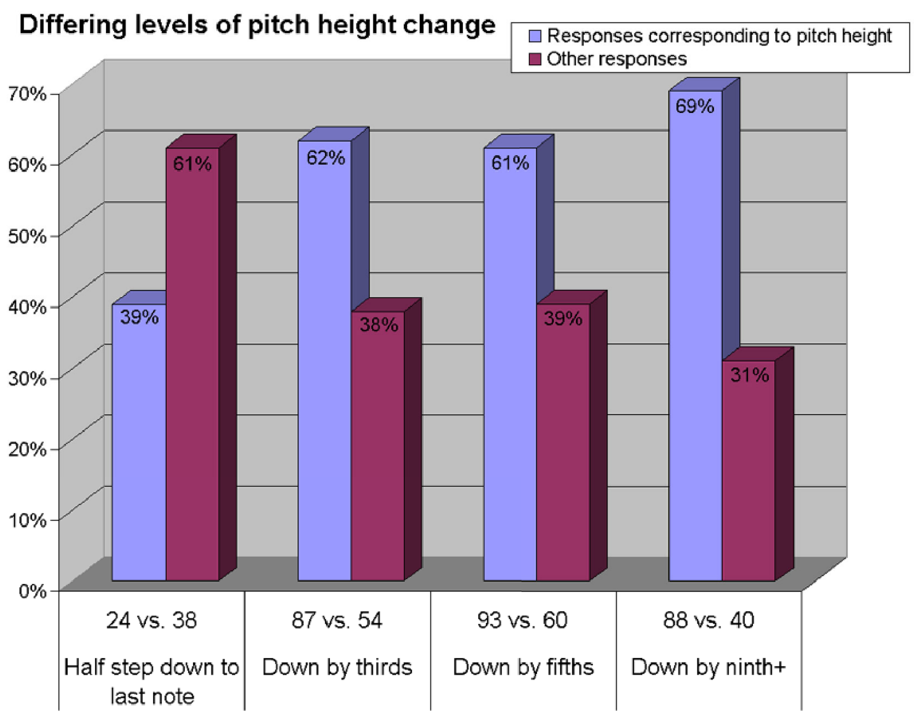
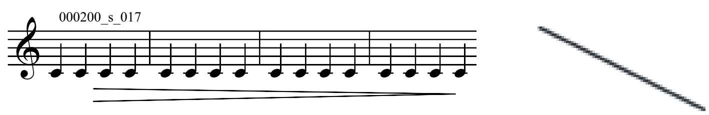


Figure 4-4: Subjects' responses to musical examples shown in Figure 4-3.

Results from other examples with changes in pitch height seem to indicate that increases in magnitude of change are not directly proportional to the perception that tension is changing. Figure 4-3 shows four examples with the same chord progression, rhythmic values, and loudness. The only difference is the amount of change in pitch height. The percentage of subjects responding to the pitch height does not increase linearly with respect to the degree of change (Figure 4-4).

4.1.2 Loudness

Pitch height is clearly an important parameter that contributes to tension. Even the very presence of pitch, without any change, appears to have an effect. This is shown in the next set of examples on loudness, which consist of notes that were recorded with pitched instruments (piano and strings) even though the pitch remained stationary (Figure 4-5). This most likely lessened the perceived effect of loudness; while it was clear that loudness was an important factor—out of 645 total responses, 42% chose the answer which corresponded to it while 58% chose one of the other eight responses—a large percentage (65%) of those choosing an answer that did not correspond to change in loudness selected the response indicating no change in tension (Figure 4-6). It might be safe to assume that the monotonous presence of an unchanging, repeated pitch blunted the results produced by change in loudness.



(a) (b)

Figure 4-5: (a) An example designed to measure subjects' responses to loudness. (b) Answer corresponding to loudness for example shown in (a).

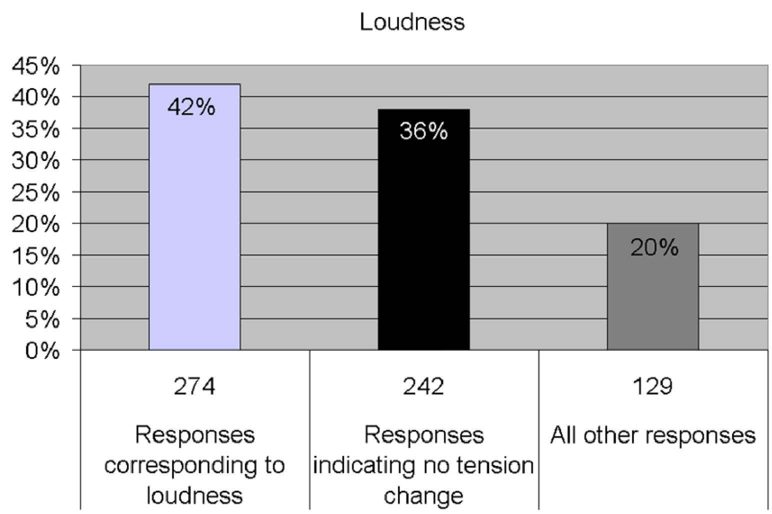
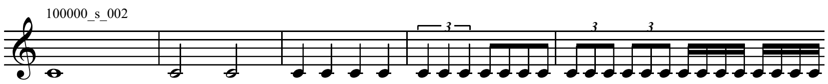
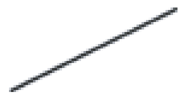


Figure 4-6: Graph showing subjects' responses to all examples that have loudness as the only changing feature.



(a)



(b)

Figure 4-7: (a) An example designed to test subjects' responses to onset frequency. (b) Answer corresponding to onset frequency for example shown in (a).

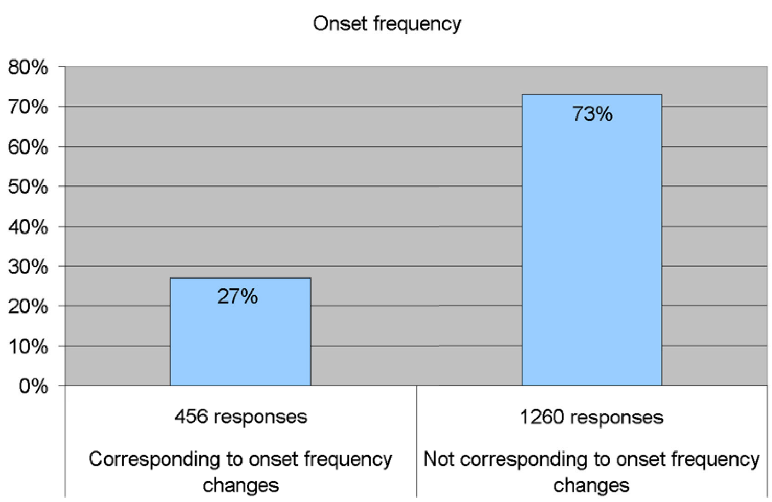


Figure 4-8: Graph showing subjects' responses to all examples where the only changing feature is onset frequency.

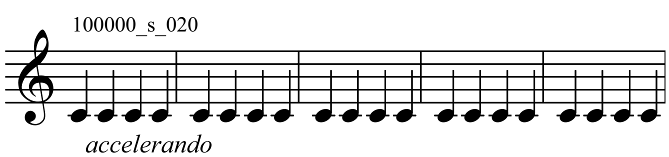


Figure 4-9: An example designed to isolate and test tempo changes (a subcategory of onset frequency). While tempo is the only feature changing, pitched notes are present, which might be affecting the results. Subjects might not only be responding to changes tempo, but lack of pitch change as well, thus dampening the perceived tension change.

4.1.3 Onset frequency

There is further evidence for this in examples that isolate onset frequency. In general, subjects responded to onset frequency and tempo changes (Figure 4-

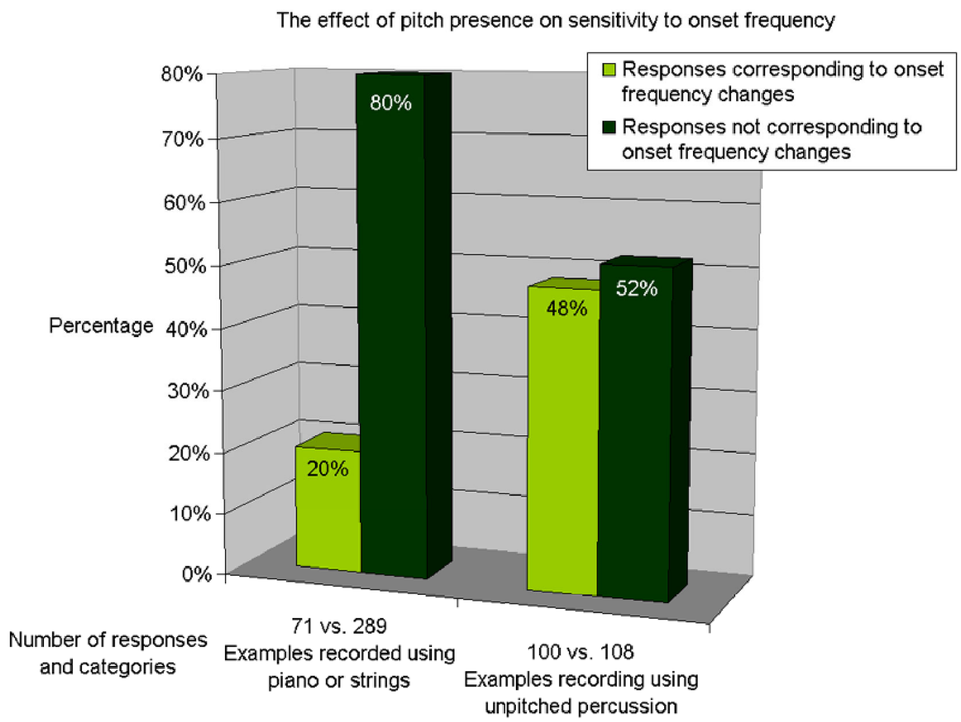


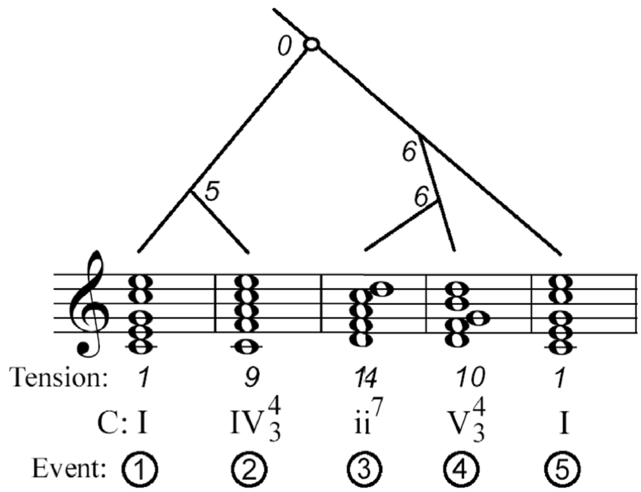
Figure 4-10: Subjects' responses to two types of examples featuring change in onset frequency. The first set consists of examples recorded with pitched instruments (strings or piano) and the second set consists of examples recorded with unpitched percussion. The presence of pitch in the first set appears to have affected subjects' responses to onset frequency.

8), though not quite as strongly as for loudness; out of 1716 total responses, 27% chose an answer which corresponded to onset frequency.[3] However, this percentage appears to be artificially low. In the case where examples (Figure 4-9) were recorded with three different timbres (piano, strings, and unpitched percussion), sensitivity to onset frequency varied greatly. Out of a total of 360 responses to examples recorded with piano or strings, only 20% of the answers corresponded with onset frequency changes while out of a total of 208 responses to the same examples recorded with unpitched percussion sounds, 48% corresponded with onset frequency (Figure 4-10).

4.1.4 Harmony

Harmony is considerably more complex than any of the other parameters because it is multidimensional and cannot be completely separated from pitch contour. The harmonic tension values are calculated with Lerdahl's tonal tension model without the melodic attraction component. Figure 4-11(a) shows

[3]While a response of 27% might not seem very significant, it is still far above the 11% that a single response would receive on average if chosen randomly (1 out of 9).



(a)

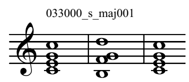
(b)

Figure 4-11: (a) An example used to test subjects' responses to harmony. The prolongational reduction is shown. (b) Answer corresponding to harmony for example shown in (a)

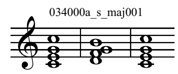
Event	Distance	δ	Inherited	Scale degree	Inversion	Non-harmonic tones	TOTAL TENSION
1	δ(1,5)	0	0	1	0	0	1
2	δ(2,1)	5	0	1	2	1	9
3	δ(3,4)	6	6	1	0	1	14
4	δ(4,5)	6	0	1	2	1	10
5	-	0	0	1	0	0	1

Figure 4-12: Chart showing values required to calculate harmonic tension values using Lerdahl's model. Values correspond to the example in Figure 4-11.

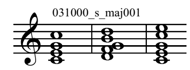
one example designed to isolate harmony. It is annotated with the prolongational reduction necessary to calculate the tension values that determine the shape of the corresponding tension curve (shown in Figure 4-11(b)). Figure 4-12 displays all of the variables that are required to evaluate the final tension value for each chord in the example. For some of the harmony examples, different voicings of the same chords were used in order to get an idea of how subtle changes in pitch contour affected the results (Figure 4-13).



(a)



(b)



(c)

Figure 4-13: Three examples designed to test subjects' responses to changes in harmony. All three contain the same chord progression but voiced slightly differently.

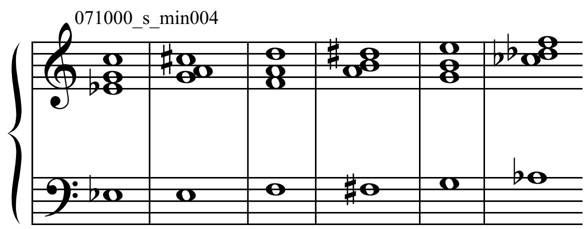


Figure 4-14: Harmonic changes paralleled by clear changes in pitch height (i.e. harmonic tension increasing while pitch height increasing).

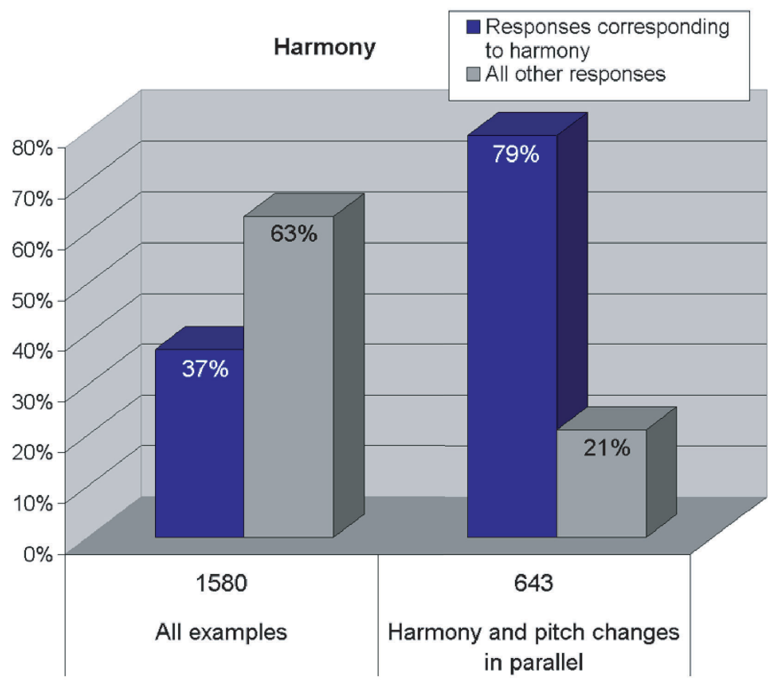


Figure 4-15: Graph showing subjects' responses to all harmony examples and a subset of those examples where pitch changes parallel harmonic movement.

In general, subjects responded to harmony quite strongly. Out of 1580 responses, 37% corresponded to changes in harmony (Figure 4-15). The strongest responses were for examples where the pitch contour very clearly coincided with the change in harmonic tension (e.g. Figure 4-14); the response rate corresponding to harmony for these example was much higher (79%). In cases where the pitch changes were more subtle, as in Figures 4-13(a), 4-13(b), and 4-13(c), the results were less clear but still seemed to indicate that movement in the salient voices, either in correspondence or opposition to the harmony, were influential. Another factor that mostly likely affected the results was melodic expectation influenced by the tonal context. See Section 4.2.1 for further discussion on melodic expectation.

4.1.5 Rhythmic irregularity

Rhythmic irregularity was the only feature tested in Experiment 1 that did not yield positive results. Figure 4-16 shows one example in this category and its hypothesized tension curve. Out of 602 total responses, only 2% felt changes in rhythmic regularity affected their perception of musical tension. The single most popular reply by far was no change in tension (Figure 4-17).

(a)

(b)

Figure 4-16: (a) An example used to gauge subjects' responses to rhythmic irregularity. (b) Answer corresponding to rhythmic irregularity for example shown in (a)

In other examples where there were slight changes in the regularity of rhythmic patterns (e.g. Figure 4-18), there seemed to be at least some effect from small changes in onset frequency. For example, when the average onset frequency increased slightly while the rhythm became less regular there appeared to be a higher number of subjects feeling tension increase as irregularity increased. Nevertheless, the numbers in both of these cases appears to be insignificant in general (Figure 4-19). There is no evidence to support that the results were affected by the presence of constant pitch because responses to both pitched and unpitched versions of rhythmic irregularity examples did not differ significantly, unlike the case of onset frequency.

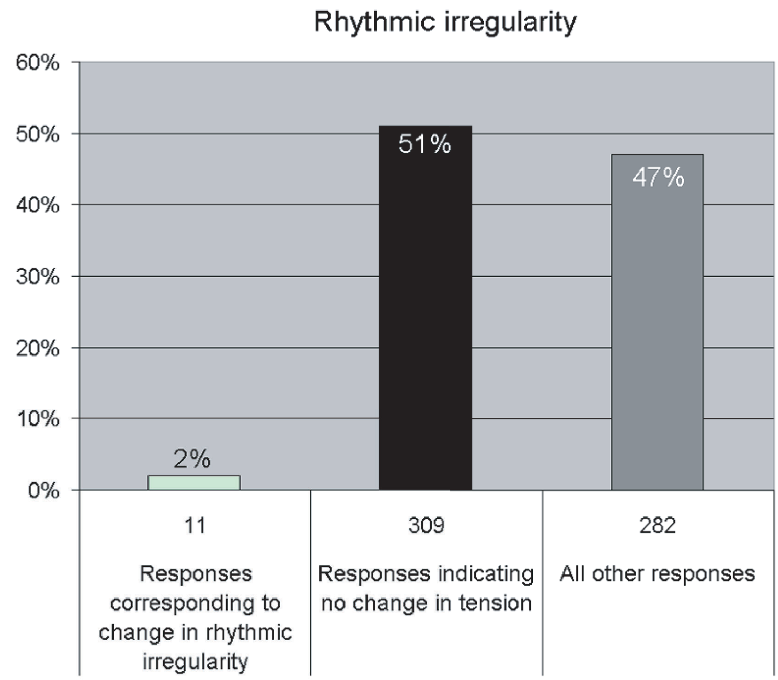


Figure 4-17: Subjects' responses to examples where the only feature changing is rhythmic regularity.

Figure 4-18: An example showing changes in rhythmic regularity as well as slight changes in average onset frequency.

4.1.6 Interaction of features

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.

Figure 4-20 shows a musical example pairing loudness with onset frequency. In the process of analyzing examples in this category, a problem was noted concerning files with a *crescendo* on the final whole note. When recorded with string sounds, there was a lack of decay on the final note; loudness continued to increase at the end of the example while there were no more note onsets,

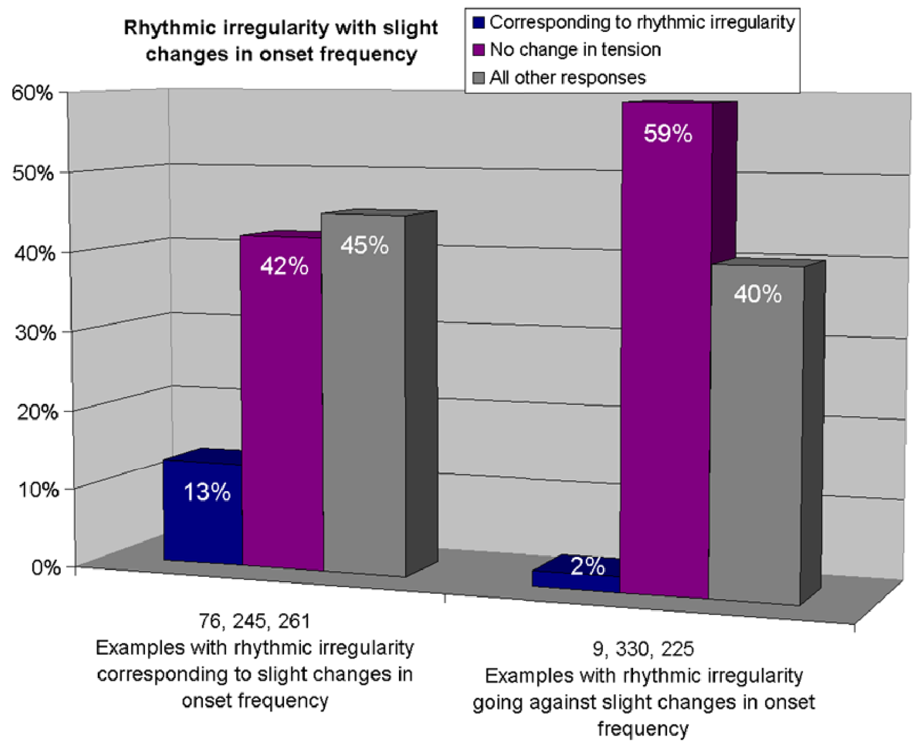


Figure 4-19: Subjects' responses to examples with changes in rhythmic regularity as well as slight changes in average onset frequency.

resulting in the perception that loudness was the only feature changing. When those examples were taken out (Figure 4-21), subject responses were closer to even between the two features, with a slight edge given to loudness.

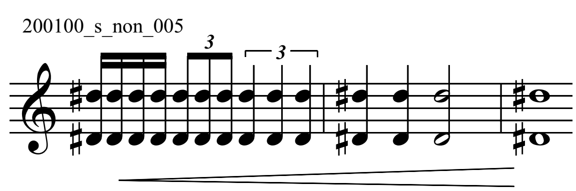


Figure 4-20: Example used to compare changes in loudness and onset frequency.

In the case of harmony versus onset frequency (see Figure 4-22 for an example), subjects appeared to respond more strongly to harmony. Out of 273 responses, 8% corresponded to onset frequency, and 17% corresponded to harmony. It is possible that the numbers for harmony were a bit low because all of the

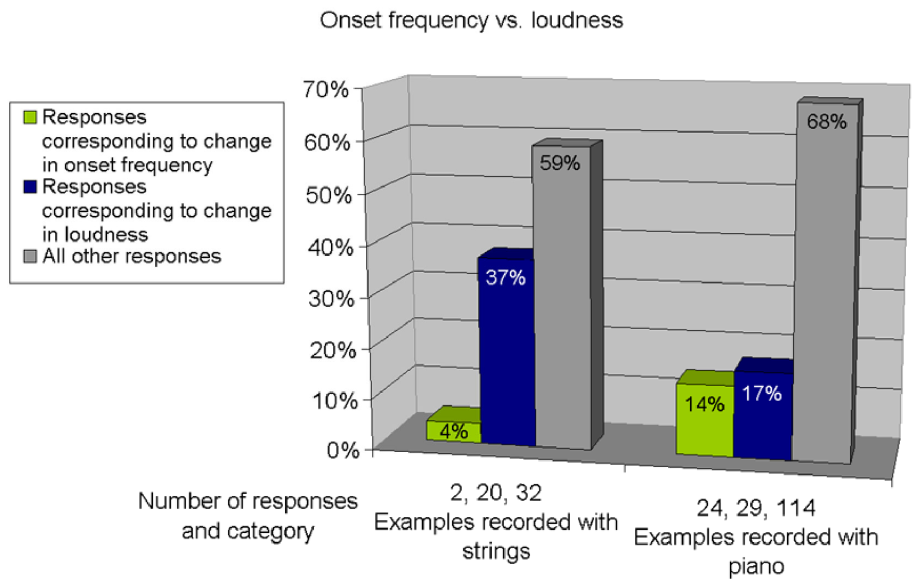


Figure 4-21: Subjects' responses to examples where loudness and onset frequency change in opposite directions.

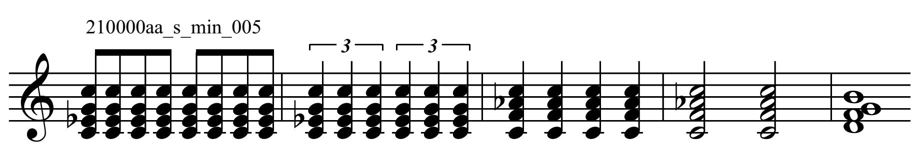


Figure 4-22: Example used to compare changes in onset frequency with changes in harmony.

examples analyzed had a stepwise descent to the last note in the soprano line. This might have strengthened the effect of decreasing onset frequency, adding to the perception of decreasing tension at the end.

The comparison of features in various combinations of triplets yielded interesting results. Figure 4-24 shows an example that combines loudness and onset frequency against harmony. Graph 4-25 describes the different combination (pairs against single features) and the subjects' responses to them. In all cases, the pairs together were stronger than the single feature. In all cases, loudness seemed to be the most salient feature that listeners responded too, while harmony and onset frequency had about the same amount of influence. These results might also have been slightly affected by the fact that the soprano line stepped down on the last change in harmony.

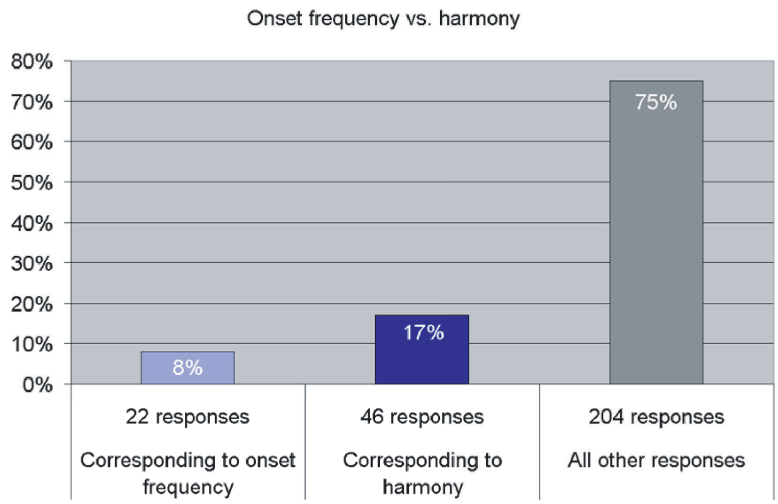


Figure 4-23: Subjects' responses to examples where onset frequency and harmony are in opposing directions.

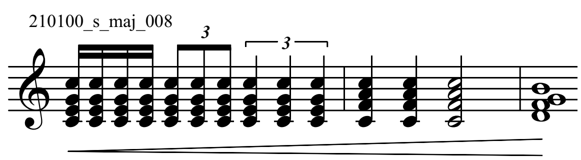


Figure 4-24: Example used for testing combinations of changing loudness, harmony, and onset frequency.

When harmony, pitch height, and loudness were combined so that they changed in the same direction (Figure 4-26), the results were very strong: out of 93 responses, 80% chose the answer that corresponded with the three features. However, when pitch height and loudness were paired together against harmony (e.g. Figure 4-26), the responses showed more ambivalence: out of 485 responses, 60% chose the response which corresponded with pitch height and loudness (Figure 4-27).[4]

4.1.7 Three corresponding features opposed by a fourth feature

This category of examples explored the effect of one particular feature in opposition to three other features in concordance. In the case of harmony, the results seemed to indicate that it had little effect on the overall perception of

[4]It must be noted that since the examples did not have the same harmonic progression, a definite conclusion can't be reached.

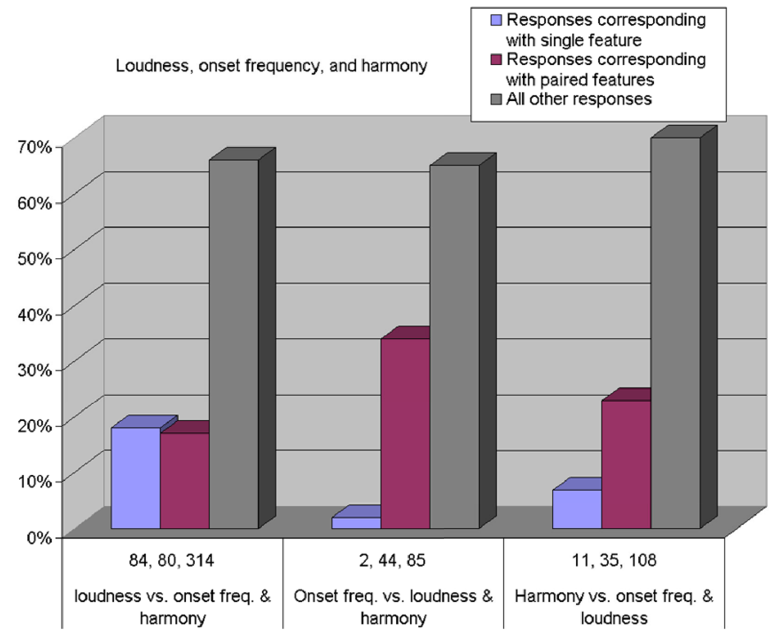


Figure 4-25: Graphs of subjects' responses to combinations of changing loudness, harmony, and onset frequency.

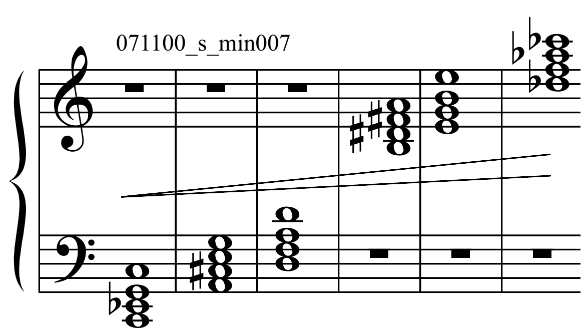


Figure 4-26: Example used for testing combinations of changes in harmony, pitch height, loudness.

increasing tension (Figures 4-28 and 4-32). Nonetheless, although the harmonic progression clearly relaxes at the end, it fits a rather complex shape as a whole (this shape is shown in Figure 3-1(f)); thus, it does less to directly counteract the other features' contribution to increasing tension.

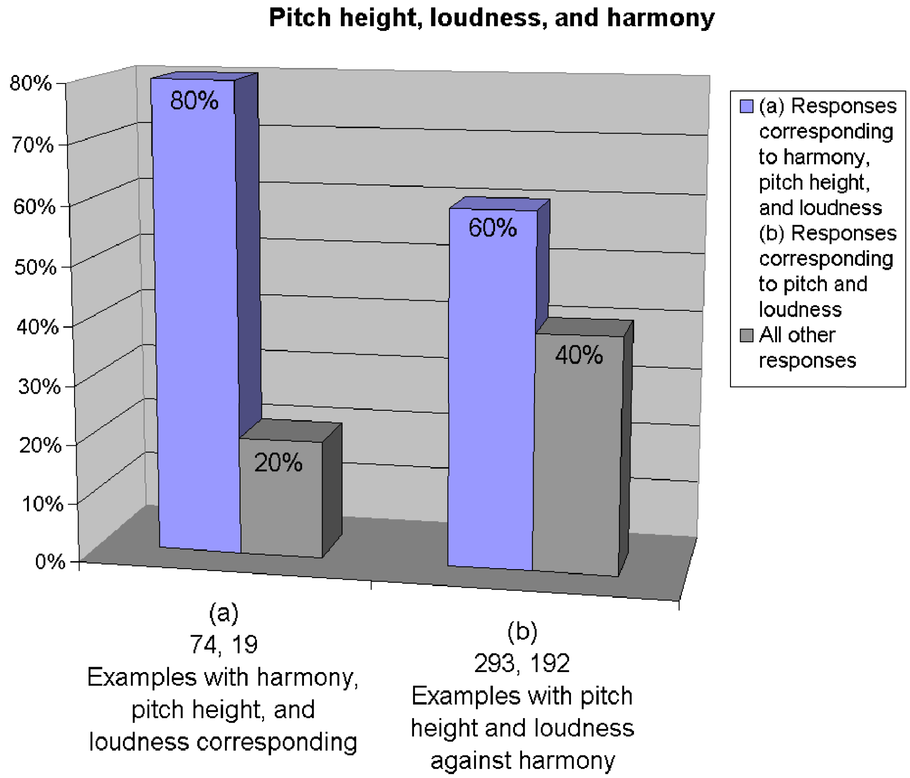


Figure 4-27: Graph of subjects' responses to examples where harmony, pitch height, and loudness are in parallel and examples where harmony changes in the opposite direction of pitch height and loudness.

Both loudness and onset frequency seem to have weak effect when up against the strength of three united features (Figures 4-29, 4-31, 4-33, and 4-35). Pitch height fares considerably better (Figures 4-30 and 4-34).

Figure 4-28: Example showing change in harmony against changes in onset frequency, pitch height, and loudness.

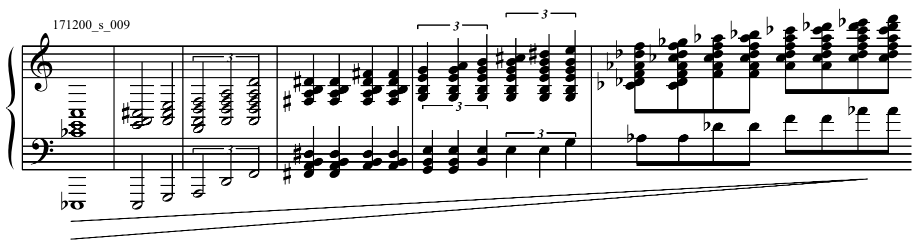


Figure 4-29: Example showing change in loudness against changes in onset frequency, harmony, and pitch height.

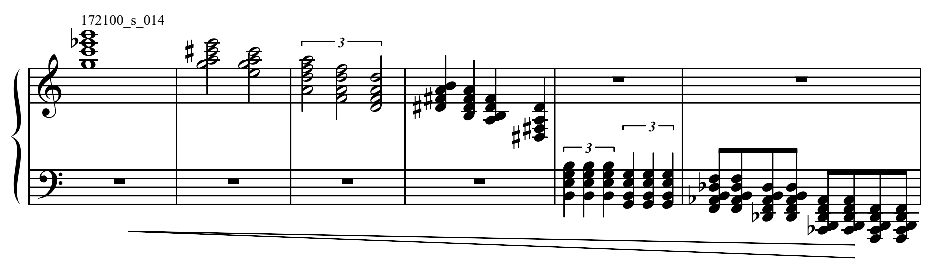


Figure 4-30: Example showing change in pitch height against changes in onset frequency, harmony, and loudness.

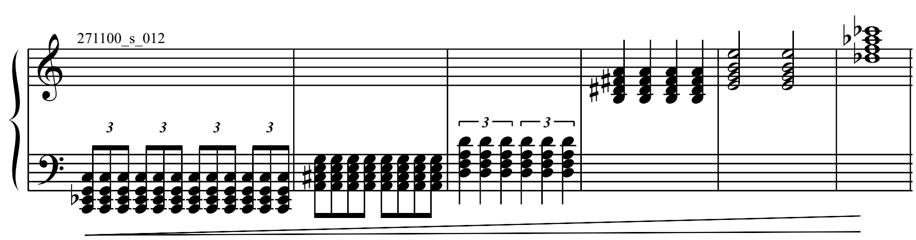


Figure 4-31: Example showing change in onset frequency against changes in harmony, pitch height, and loudness.

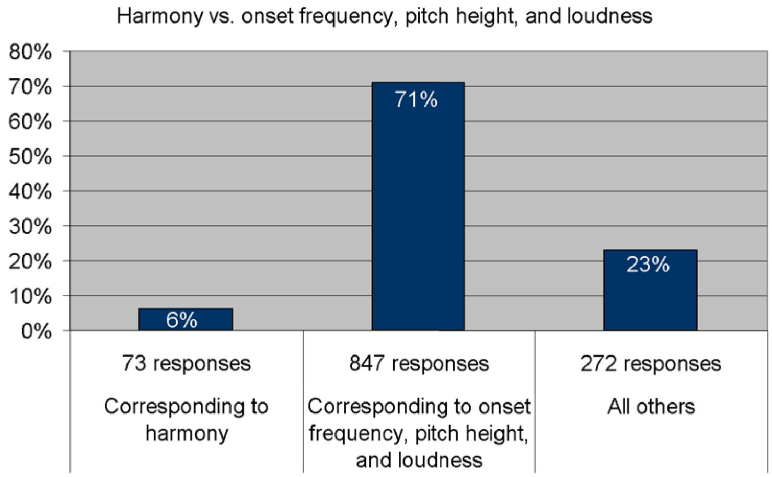


Figure 4-32: Graph of subjects' responses to examples with harmony against three other features.

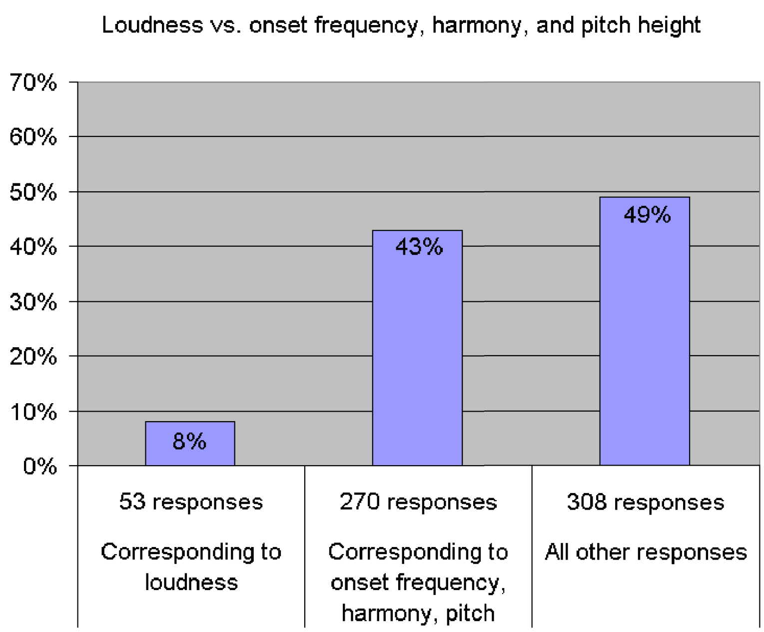


Figure 4-33: Graph of subjects' responses to examples where change in loudness goes against changes in three other features.

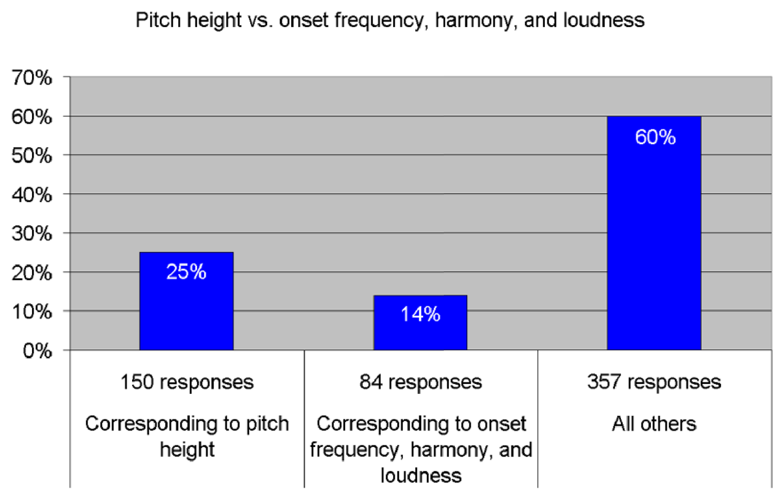


Figure 4-34: Graph of subjects' responses to examples where change in pitch height goes against changes in three other features.

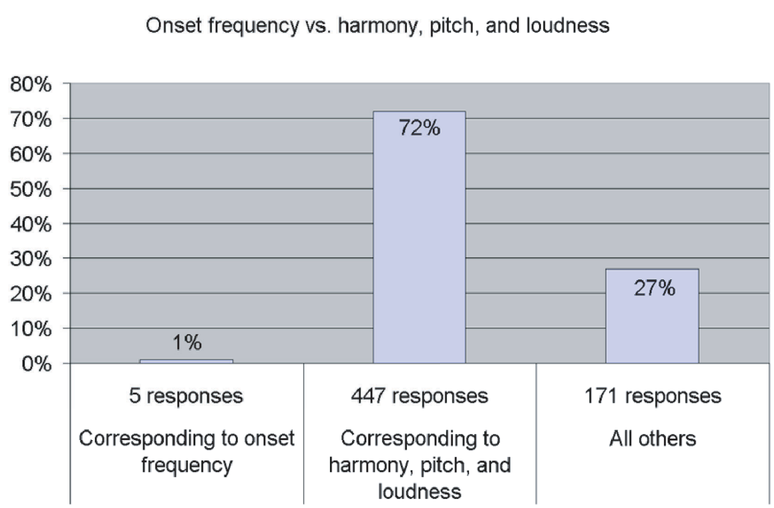


Figure 4-35: Graph of subject responses to examples where change in onset frequency goes against three other features.

4.1.8 Musicians versus non-musicians

In order to compare results between musically experienced and musically inexperienced subjects, the data was divided into two categories. In the first category (17% of total) were subjects who rated their level of musical training a 4 or higher on a scale of 1 to 5. The second category contained the rest of the subjects (83%).

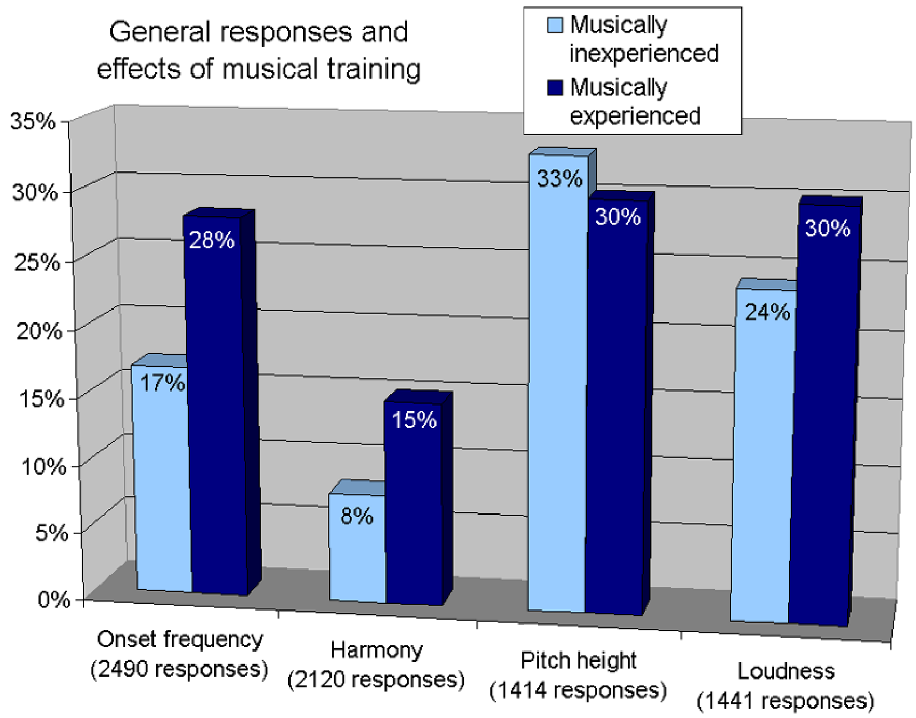


Figure 4-36: Graph comparing responses of subjects with considerable musical training with subjects with less musical training.

For onset frequency, harmony, pitch height, and loudness, all of the examples where each feature was present and not in correspondence with any other features were selected to determine overall sensitivity to that feature. Responses were categorized as either corresponding or not corresponding to the feature in question. A graph showing the results is shown in Figure 4-36. Each feature has two bars—one representing the response of musically experienced subjects and the other representing musically inexperienced subjects. The absolute values of the bars are not important; only the proportional differences between the two bars in each group matter.

As can be seen from the graph, the clearest difference between musically experienced and inexperienced subjects was sensitivity to harmony, followed by sensitivity to onset frequency. Other studies have also indicated that musicians are particularly sensitive to harmony [Bigand et al., 1996], [Parncutt, 1989].

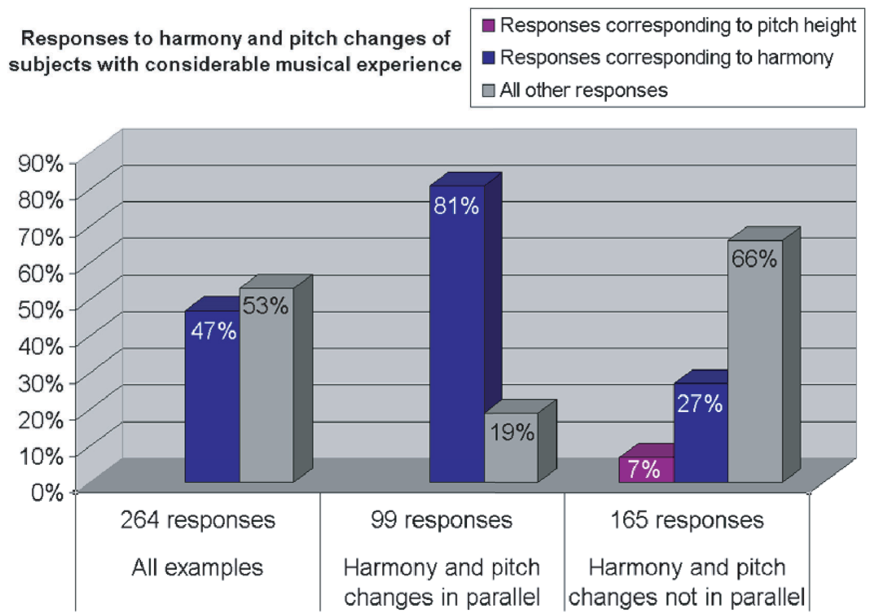


Figure 4-37: Graph showing responses of subjects with considerable musical training to examples with simple harmonic progressions and small changes in pitch.

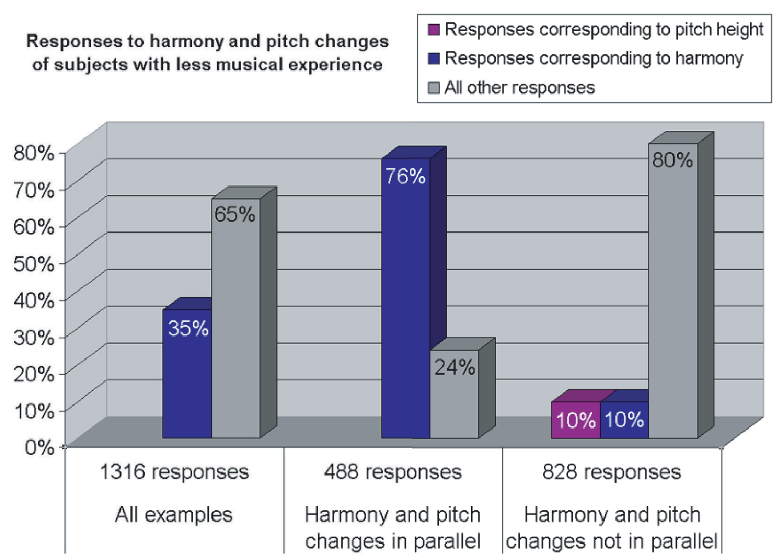


Figure 4-38: Graph showing responses of subjects with less musical training to examples with simple harmonic progressions and small changes in the melody line.

There did not appear to be significant differences in responses to pitch height and loudness. However, given that other studies have shown that musically experienced subjects are more sensitive to harmony while musically inexperienced subjects appear to be more responsive to pitch changes in the salient voice ("horizontal motion" in [Bigand et al., 1996]), additional analysis was done in order explore how musical background might have affected responses to pitch height in the context of harmony. The examples analyzed were the same subset as those discussed in Section 4.1.4. In general, the results (see Figures 4-37 and 4-38) lend some support to the results of previous studies.[5]

4.1.9 Summary

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.

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 (Figure 4-3).

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.

It is important to note that in order to make it absolutely clear that musical features were changing in one "direction" or another, the examples had to be short and often exaggerated—pitch height dropped or jumped dramatically, loudness levels went from one extreme to another, the rhythm sped up or slowed down without any subtleties in timing. The only feature where such exaggerations

[5] As noted in [Bigand et al., 1996], even small changes in the soprano voice seem to have an effect. The perceptual salience of the higher, outer voice has been reported in other experimental studies as well ([Francès, 1958], [Palmer and Holleran, 1994], [Thompson and Cuddy, 1989]).

were not utilized was harmony. In general, the obviousness of a changing feature most likely had an effect on how much it influenced the perceived tension.

4.2 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.

4.2.1 Feature graphs

All of the musical parameters confirmed in Experiment 1 as well as one additional parameter were quantitatively described for each excerpt. 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.

For discussion purposes, the 10 musical examples used in Experiment 2 will be referred to as Q01–Q10. See Appendix A for scores of all the excerpts.

Pitch height and melodic expectation

Most examples did not require more than one or two pitch height graphs (soprano and bass lines for the most part). The only example that had four separate graphs (for soprano, two inner voices, and bass) was Q08 (Schönberg).

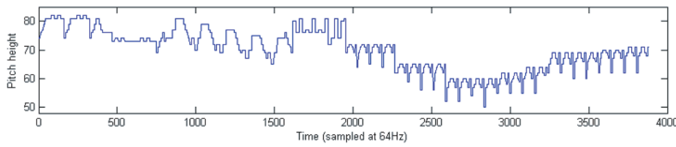


Figure 4-39: Graph showing pitch height values over time for Q05 (Bach-Vivaldi).

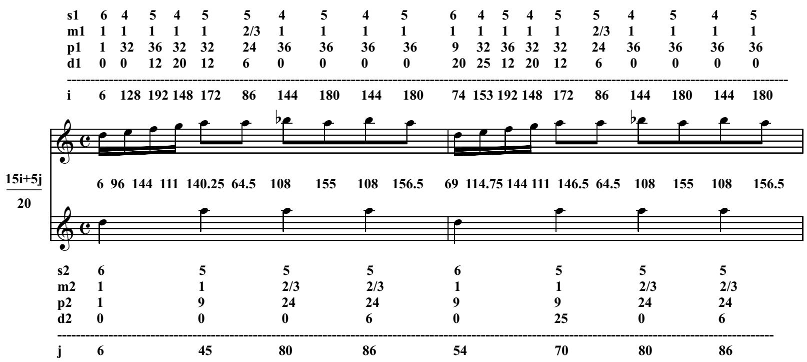


Figure 4-40: First page of analysis showing how melodic expectation values are calculated for Q05 (Bach-Vivaldi). i values consist of direct note-to-note level expectations, and j values consist of high-level patterns based on salience and metrical placement. s represents stability ratings, m represents mobility ratings, p represents proximity ratings, and d represents direction ratings. The values between the two staves are the final melodic expectations ratings for each note.

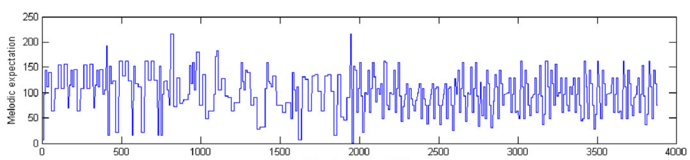


Figure 4-41: Melodic expectation graph for Q05 (Bach-Vivaldi).

The individual pitch height values were not connected by linear interpolation. Since the x-axis of the graph spanned the time of the excerpt, the values were extended for their respective durations, resulting in a step-like graph. This graph format was the same for other features with values that change at discrete time intervals.

While pitch height is an important factor in how listeners perceive tension, it's also somewhat crude. It does not take into account some of deeply schematic expectations of melodic contour described in Narmour's theory 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 (see

Chapter 2), with a few minor adjustments, was used to analyze the examples (see Figures 4-40 and 4-41).

Loudness model

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[6]), frequency warping into a cochlear-like frequency distribution, frequency masking, and temporal masking [Moore and Glasberg, 1995] [Zwicker and Fastl, 1999] [Glasberg and 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.

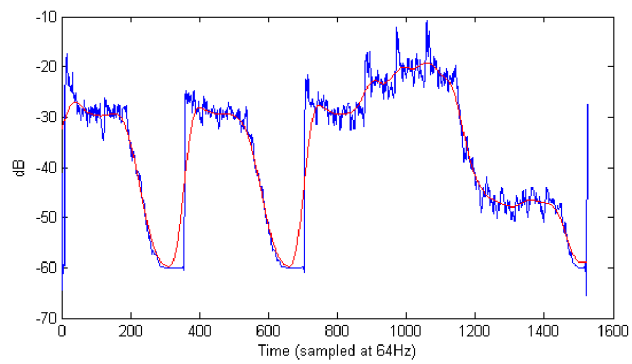


Figure 4-42: Loudness curve of Q03 (Beethoven). The blue line is the perceived loudness in dB produced by Jehan's psychoacoustic model and the red line is the smoothed version used for the feature.

The feature graph values are measured in dB with a reference silence at -60dB. A point was computed every 256 samples at 44100 Hz with a window length of 4096 samples. The results were then filtered to obtain a smooth curve (Figure 4-42).

[6]Fletcher-Munson curves show that human hearing is most sensitive in the frequency range of 2000Hz to 5000Hz [Fletcher and Munson, 1937].

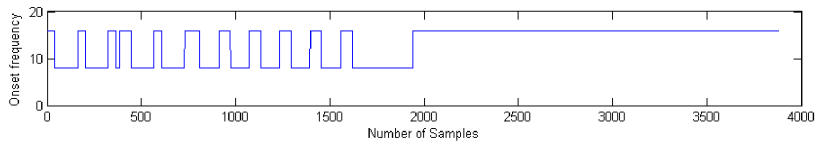


Figure 4-43: Graph showing onset frequency in Q05 (Bach-Vivaldi).

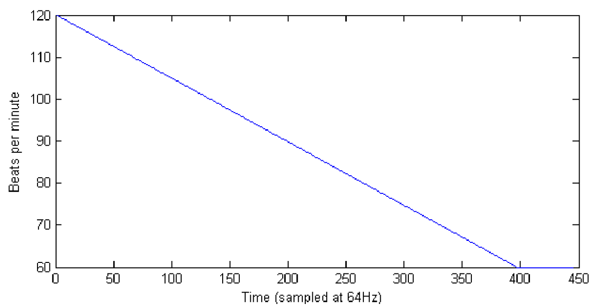


Figure 4-44: Graph showing change in tempo for Q09 (cadence with *crescendo* and *ritardando*).

Onset frequency and tempo

Even though there were some descriptive overlaps, tempo and onset frequency were treated separately and given individual feature graphs. Figure 4-43 shows the onset frequency graph for Q05 (Bach-Vivaldi).

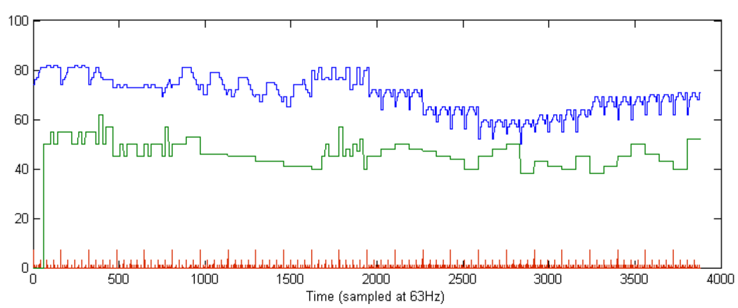


Figure 4-45: Graph showing beat and onset markings in red for Q05 (Bach-Vivaldi). The pitch height graphs for the melody and bass line have been added in blue and green for reference.

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 4-45). Binary values were assigned to the three labels: onset (1), beat (2), downbeat (4). These values were summed as needed.

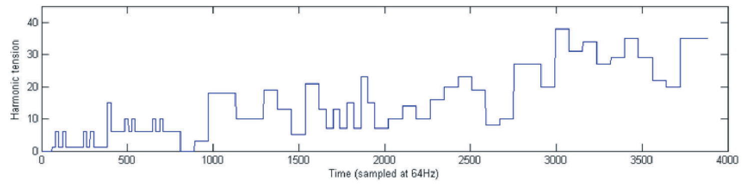


Figure 4-46: Harmonic tension feature graph Q05 (Bach-Vivaldi). The x-axis represents time, and the y-axis, tension values.

Harmonic tension

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. Lerdahl's theory is explained in Chapter 2; 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. Figure 4-48 shows the prolongational analysis for Q05 (Bach-Vivaldi); Figure 4-47 shows how the tonal tension values are calculated. The resulting harmonic tension graph of Q05 is shown in Figure 4-46.

The harmonic tension graph for Q08 (the atonal Schönberg excerpt) was calculated differently from the tonal examples.[7] Each chord was assigned a tension value based on the interval classes it contained. This was done by assigning each interval class (six total) a value corresponding to its relative harmonic dissonance (Figure 4.1). The final tension value given to a chord consisted of the sum of the values associated with the interval classes found in that chord.

Interval Class	Value
P5 or P4	1
M3 or m6	2
m3 or M6	3
M2 or m7	4
m2 or M7	5
A4 or d5	6

Table 4.1: Interval classes and their assigned values based on relative dissonance.

[7] An atonal prolongational reduction (described in [Lerdahl, 2001]) was not considered but not used.

	Chords	i	j	k	δ	Inherited	Scale degree	Inversion	Non-harmonic tones	TOTAL	
d(1)	i	0	0	0	0	0	0	0	0	0	d(1)
d(2,1)	i - i	0	0	0	0	0	1	0	0	1	d(2,1)
d(3,4)	iv - i	0	1	4	5	0	1	0	0	6	d(3,4)
d(4,2)	i - i	0	0	0	0	0	1	0	0	1	d(4,2)
d(5,6)	iv - i	0	1	4	5	0	1	0	0	6	d(5,6)
d(6,4)	i - i	0	0	0	0	0	1	0	0	1	d(6,4)
d(7,1)	i - i	0	0	0	0	0	1	0	0	1	d(7,1)
d(8,9)	iv - i	0	1	4	5	0	1	0	0	6	d(8,9)
d(9.7)	i - i	0	0	0	0	0	1	0	0	1	d(9.7)
d(10,11)	iv - i	0	1	4	5	0	1	0	0	6	d(10,11)
d(11,9)	i - i	0	0	0	0	0	1	0	0	1	d(11,9)
d(12,1)	i - i	0	0	0	0	0	1	0	0	1	d(12,1)
d(13,14)	ii2 - V	0	1	5	6	5	1	2	1	15	d(13,14)
d(14,1)	V - i	0	1	4	5	5	1	0	0	11	d(14,1)
d(15,14)	V-V	0	0	0	0	5	1	0	0	6	d(15,14)
d(16,17)	i-V	0	1	4	5	5	0	0	0	10	d(16,17)
d(17,15)	V-V	0	0	0	0	5	1	0	0	6	d(17,15)
d(18,19)	i-V	0	1	4	5	5	0	0	0	10	d(18,19)
d(19,17)	V-V	0	0	0	0	5	1	0	0	6	d(19,17)
d(20,14)	V-V	0	0	0	0	5	1	0	0	6	d(20,14)
d(21,22)	i-V	0	1	4	5	5	0	0	0	10	d(21,22)
d(22,20)	V-V	0	0	0	0	5	1	0	0	6	d(22,20)
d(23,24)	i-V	0	1	4	5	5	0	0	0	10	d(23,24)
d(24,22)	V - V	0	0	0	0	5	1	0	0	6	d(24,22)
d(25,26)	V - i	0	1	4	5	0	1	0	0	6	d(25,26)
d(26,1)	i - i	0	0	0	0	0	0	0	0	0	d(26,1)
d(27,26)	i6 - i	0	0	0	0	0	1	2	0	3	d(27,26)
d(28,29)	iv6 - III6	0	2	6	8	7	0	2	1	18	d(28,29)
d(29,26)	III6 - i	0	3	4	7	0	0	2	1	10	d(29,26)
d(30,29)	ii6 - III6	0	2	6	8	7	1	2	1	19	d(30,29)
d(31,32)	IV6 - iii6	0	2	6	8	2	0	2	1	13	d(31,32)
d(32,26)	iii6/Bb - i/d	2	0	0	2	0	0	2	1	5	d(32,26)
d(33,34)	iv6 - V	0	2	6	8	11	0	2	0	21	d(33,34)
d(34,35)	V - i	0	1	4	5	7	1	0	0	13	d(34,35)
d(35,42)	i - i	0	0	0	0	7	0	0	0	7	d(35,42)
d(36,37)	iv - i	0	1	4	5	7	1	0	0	13	d(36,37)
d(37,35)	i - i	0	0	0	0	7	0	0	0	7	d(37,35)
d(38,39)	iv64 - i	0	1	4	5	7	1	2	0	15	d(38,39)
d(39,37)	i - i	0	0	0	0	7	0	0	0	7	d(39,37)
d(40,41)	ii65 - V7	0	1	5	6	13	1	2	1	23	d(40,41)
d(41,42)	V7 - i	0	1	5	6	7	1	0	1	15	d(41,42)
d(42,26)	i/a - i/d	1	1	5	7	0	0	0	0	7	d(42,26)
d(43,42)	i6 - i	0	0	0	0	7	1	2	0	10	d(43,42)
d(44,45)	iv7 - i6	0	1	4	5	7	1	0	1	14	d(44,45)
d(45,42)	i6 - i	0	0	0	0	7	1	2	0	10	d(45,42)
d(46,45)	ii - i6	0	2	6	8	7	1	0	0	16	d(46,45)
d(47,46)	ii2 - ii	0	0	1	1	15	1	2	1	20	d(47,46)
d(48,49)	vii7 - V9	0	3	4	7	14	1	0	1	23	d(48,49)
d(49,50)	V9 - i	0	1	6	7	7	1	0	4	19	d(49,50)
d(50,42)	i - i	0	0	0	0	7	1	0	0	8	d(50,42)
d(51,50)	i6 - i	0	0	0	0	7	1	2	0	10	d(51,50)
d(52,53)	V7 - i	0	1	5	6	19	1	0	1	27	d(52,53)
d(53,50)	i/g - i/a	2	2	8	12	7	1	0	0	20	d(53,50)
d(54,55)	V2 - I	0	1	5	6	28	1	2	1	38	d(54,55)
d(55,53)	I/C - i/g	2	1	6	9	19	1	2	0	31	d(55,53)
d(56,57)	V7-i	0	1	5	6	26	1	0	1	34	d(56,57)
d(57,53)	i/d-i/g	1	1	5	7	19	1	0	0	27	d(57,53)
d(58,57)	i 6 - i	0	0	0	0	26	1	2	0	29	d(58,57)
d(59,60)	vii6 - V7	0	3	4	7	25	1	2	0	35	d(59,60)
d(60,61)	V7 - i6	0	1	5	6	19	1	2	1	29	d(60,61)
d(61,53)	i/g-i/g	0	0	0	0	19	1	2	0	22	d(61,53)
d(62,61)	i/g-i/g	0	0	0	0	19	1	0	0	20	d(62,61)
d(63,62)	V7/ii - i	2	3	9	14	19	1	0	1	35	d(63,62)

Figure 4-47: Chart of harmonic tension calculations for Q05 (Bach-Vivaldi). δ is the sum of i, j, and k.

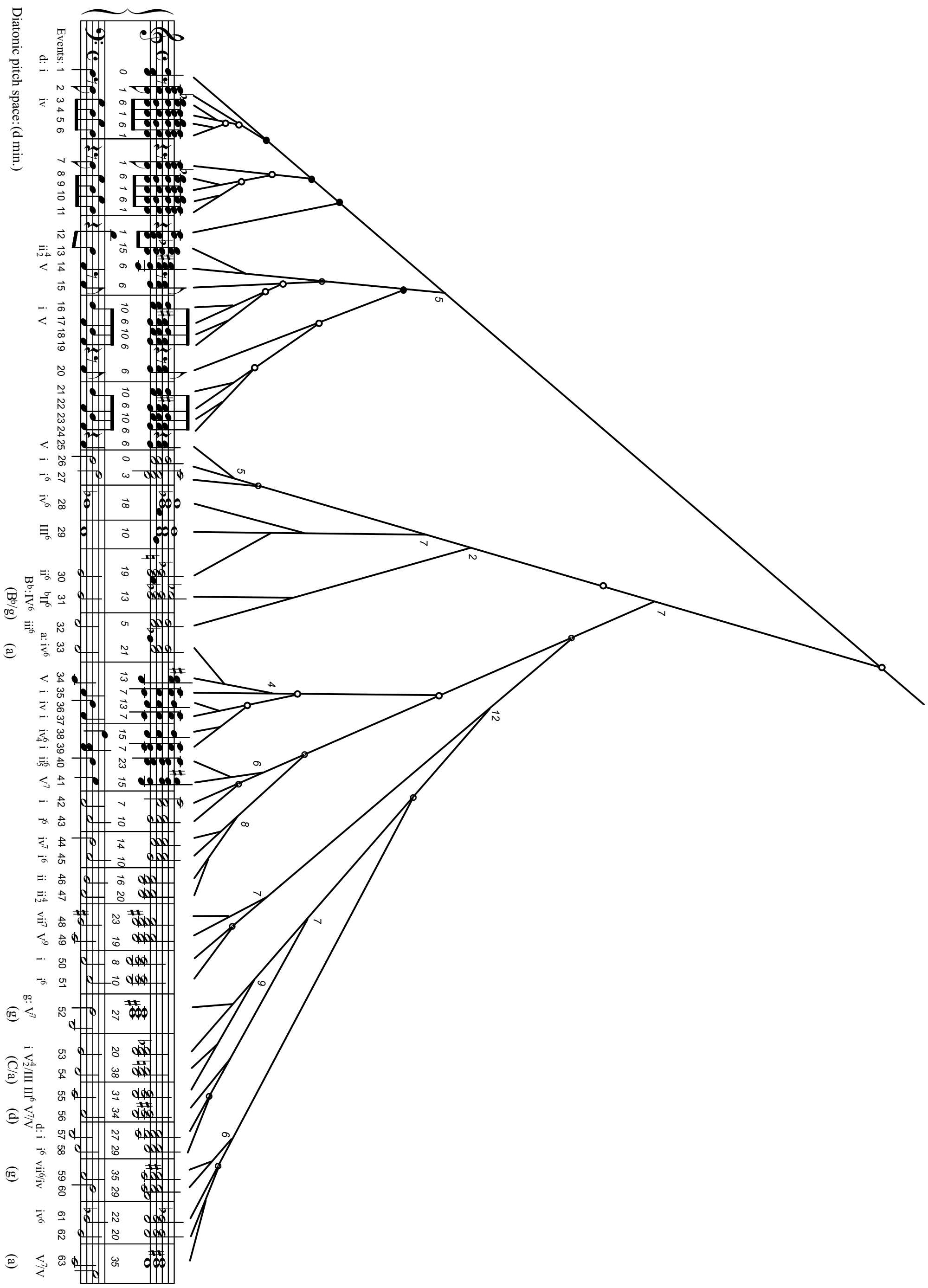


Figure 4-48: Prolongational reduction of Q05 (Bach-Vivaldi).

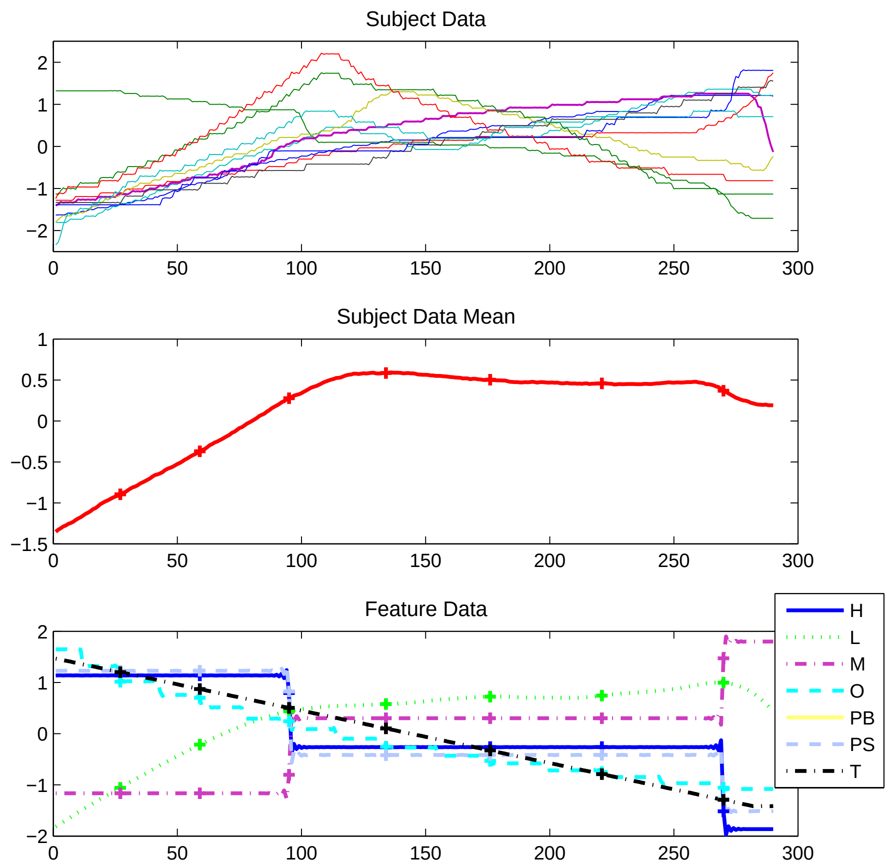


Figure 4-49: Subject and feature data graphs for Q09 (cadence with *crescendo* and *ritardando*). The first window shows 10 out of the 132 subject responses for Q09 (number of subjects multiplied by number iterations = total number of responses: 33 * 4 = 132). The second window shows the mean of the subject responses. The third window shows the corresponding feature graphs. The plus-signs (+) indicate the beat positions. H = harmony, L = loudness, M = melodic expectation, O = onset frequency, PB = pitch height of bass line, PS = pitch height of soprano line, and T = tempo (note that there is no graph line for PB because Q09 has no bass line).

4.2.2 Linear correlation of individual features with empirical data

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

	FEATURE	*r*	*p*	*r*	*p*
Q01					
	H	0.38326261	0.00000000	0.43355306	0.18280565
	L	0.95770874	0.00000000	0.96374109	0.00000183
	M	0.88679215	0.00000000	0.91281422	0.00008796
	PS	0.84125444	0.00000000	0.85739264	0.00074120
Q02					
	L	0.03171998	0.53975753	0.06167140	0.83410704
	M	0.29503266	0.00000001	-0.47081815	0.08928231
	O	-0.99303721	0.00000000	-0.99565072	0.00000000
	PS	0.99178999	0.00000000	0.99421407	0.00000000
	T	-0.98604293	0.00000000	-0.98737709	0.00000000
Q03					
	H	-0.56474740	0.00000000	-0.53191054	0.03394465
	L	0.68648549	0.00000000	0.75058018	0.00080790
	M	-0.15004340	0.00000060	-0.16156906	0.54997260
	O	0.20892201	0.00000000	0.22356635	0.40521871
	PB	0.32655890	0.00000000	0.29542534	0.26662375
	PS	0.17267138	0.00000001	0.10862305	0.68883575
Q04					
	L	-0.56967372	0.00000000	-0.55952130	0.04678283
	M	0.45872641	0.00000000	0.69571514	0.00827215
	O	0.01942902	0.77144141	-0.00029911	0.99922622
	PS	0.84035010	0.00000000	0.81356491	0.00071614
Q05					
	H	0.60318666	0.00000000	0.62983991	0.00000000
	L	-0.01555631	0.39585532	-0.01885915	0.85605443
	M	-0.08317103	0.00000544	-0.23830955	0.02003895
	O	0.33901234	0.00000000	0.34609011	0.00059163
	PB	0.04860168	0.00795353	0.23778107	0.02032427
	PS	-0.21903811	0.00000000	-0.23336413	0.02284824

Figure 4-50: Table showing correlation results for Q01–Q05. H = harmony, O = onset frequency, L = loudness, M = melodic expectation, PS = pitch height of soprano line, PB = pitch height of bass line, and T = tempo. The first two columns of r and p-values are for all data points (50 points sampled per second). The second two columns are for data sampled at every beat or note onset. Negative r-values indicate inverse correlation.

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. Figure 4-49 shows the normalized subject data, mean subject data, and feature data graphs for Q09 (cadence with *crescendo* and *ritardando*). The mean of the subject data is important because it was used in the subsequent analyses.

	FEATURE	*r*	*p*	*r*	*p*
Q06					
	L	0.12251220	0.10133820	0.11579722	0.65807578
	O	0.47023323	0.00000000	0.40735346	0.10460470
	T	0.63220668	0.00000000	0.56879954	0.01718548
Q07					
	H	0.86605972	0.00000000	0.87550465	0.00000005
	L	0.29724263	0.00000000	0.54700111	0.00691031
	M	0.59865184	0.00000000	0.58420386	0.00342108
	O	0.00825482	0.82693068	0.02839480	0.89766780
	PS	-0.76008177	0.00000000	-0.76858978	0.00001836
Q08					
	H	0.35387302	0.00000000	0.54308325	0.00740763
	L	0.69999235	0.00000000	0.66950768	0.00047573
	M	0.11774979	0.00063677	0.14294902	0.51525112
	O	0.66002918	0.00000000	0.70905500	0.00015199
	PB	0.61870943	0.00000000	0.67365974	0.00042538
	PS	0.15661050	0.00000524	-0.32528319	0.12988818
Q09					
	H	-0.70343061	0.00000000	-0.71111435	0.07320188
	L	0.96800235	0.00000000	0.95582008	0.00076941
	M	0.72414543	0.00000000	0.72889239	0.06311496
	O	-0.86898707	0.00000000	-0.84906249	0.01564570
	PS	-0.78338548	0.00000000	-0.78189104	0.03779292
	T	-0.74667546	0.00000000	-0.74942791	0.05246286
Q10					
	H	-0.20143758	0.00000000	-0.20610314	0.09685901
	L	0.79220361	0.00000000	0.77912673	0.00000000
	M	0.01799613	0.39384982	0.08909404	0.47684226
	O	0.18491712	0.00000000	0.14831473	0.23464471
	PB	0.20922164	0.00000000	0.23286939	0.05988333
	PS	-0.21371263	0.00000000	-0.15297703	0.22009385

Figure 4-51: Table showing correlation results for Q06–Q10. H = harmony, O = onset frequency, L = loudness, M = melodic expectation, PS = pitch height of soprano line, PB = pitch height of bass line, and T = tempo. The first two columns of r and p-values are for all data points (50 points sampled per second). The second two columns are for data sampled at every beat or note onset. Negative r-values indicate inverse correlation.

In general, it was difficult to correlate the feature graphs with the subject data because of the jagged edges of the former were at odds with the smooth curves of the slider movements.[8]

The r (Pearson's correlation coefficient) and p-values resulting from the correlation are shown in Figures 4-50 and 4-51. The first two columns of values are

[8]To simplify the analysis, all of the inner voice feature graphs were removed, and only the bass and soprano lines were considered. Derivatives of the feature graphs (e.g. *change* in loudness in addition to absolute loudness) were considered but thrown out in the end.

correlations done with *all* the points in the graphs (50 samples per second). The second pair of values correspond to a correlation done with only points sampled every beat (or if an example does not have beats, onset values). Negative values indicate inverse correlation.

The reason why the p-values are so low in the original set is because of the large number of samples constituting the feature graphs and subject data, all of which are continuous functions. In other words, the samples are not independent of each other; any two adjacent values (at time $t = n$ and $t = n+1$) are very similar because they change very little. The definition of p requires that these values are independent, and this requirement is not satisfied. Even if the correlation values are valid, the p-values are not.

The correlation values represent the relative importance of each feature for each excerpt. For example, the loudness graph does not correlate at all with the subject data for Q05 (Bach-Vivaldi) while it does significantly for Q3 (Beethoven). This makes sense because there are only very subtle and not clearly defined fluctuations in dynamics for Q05 while there are very obvious and clear trends in Q03. 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.

4.2.3 Musical experience

There appears to be some evidence corroborating the results from Experiment 1 regarding differences between musically experienced and inexperienced subjects and sensitivity to harmony. Subjects were placed in one or the other category much in same way as they were in Experiment 1: those categorized as musically experienced rated themselves a 4 or higher on a scale of 1 to 5 describing their overall level of musical training. Given this criteria, 15 out of 33 subjects were placed in the musically experienced category.

Figure 4-52 shows the subject data mean for Q01, a simple example where the harmony resolves at the end but the pitch contour rises. The responses from the musically inexperienced subjects followed the rise of the pitch contour at the end, while the musically experienced subjects appeared to respond more strongly to the harmonic resolution. This supports Experiment 1 results that show musicians respond more strongly to harmony than non-musicians in examples with simple chord progressions.

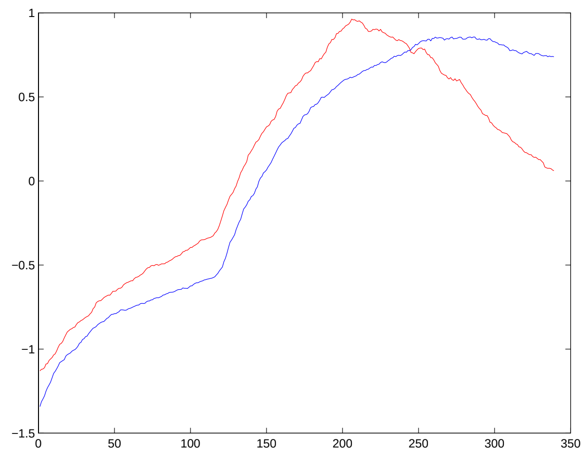


Figure 4-52: Subject data mean for Q01 (simple chord progression with rising melodic line). The red line represents the average response of musically experienced subjects and the blue line represents responses of musically inexperienced subjects.

4.2.4 The goal: 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.

Linear and nonlinear 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, resulting in the following basic formula:

$$P(t) = F(o, h, p, l, t, m) \quad (4.1)$$

$$= F(\hat{x}) \quad (4.2)$$

where o = onset frequency, h = harmony, p = pitch height, l = loudness, t = tempo, and m = melodic expectation.

The goal was to approximate F so that it matched the subject data as accurately as possible. Initially, three types of models were considered:

(A) Linear models

$$F(\hat{x}) = b_0 + b_1x_1 + b_2x_2 + ... \tag{4.3}$$

Linear models are easiest to handle and best understood in machine learning. Coefficients can be identified through a simple matrix inversion. Given the small number of dimensions required in this case, computational resources are not issue. However, linear models only work properly for linear systems. Any kind of nonlinear interaction between the parameters cannot be represented with this type of model.

(B) Polynomial models and generalized linear models
Generalized linear models, also referred to as a linear coefficient models have the following form:

$$F(\hat{x}) = \sum_i b_i f_i(\hat{x}) \tag{4.4}$$

The most popular form of a generalized linear model is a polynomial model:

$$F(\hat{x}) = \sum_i b_i x_1^{a_{1,i}} x_2^{a_{2,i}} ... \tag{4.5}$$

Generalized linear models are nonlinear models that retain some of the properties of linear models. Most importantly, they can be trained by a simple matrix inversion. These models are likely to be a good compromise between descriptive power and model complexity for the task at hand.

(C) Nonlinear coefficient models

$$F(\hat{x}) = \sum_i f_i(\hat{b}_i, \hat{x}) \tag{4.6}$$

Nonlinear coefficient models allow for any imaginable basis function. The descriptive power of these models is limitless as is the computational effort to train them.

In preparation for the regression analysis, feature graphs were smoothed with a raised-cosine (moving-average) filter to remove the sharp edges and discontinuities. This was necessary because while much of the feature data had very sharp edges (particularly in the case of step functions), the subject data was smooth because people tended to use gradual motions to indicate tension changes. Given this disconnect between the features and subject data, it would be very difficult for a model (linear model in particular) to estimate the sudden changes.

Both linear regression (also known as multiple regression) and polynomial regression were performed on the first part (training data) of each example. The

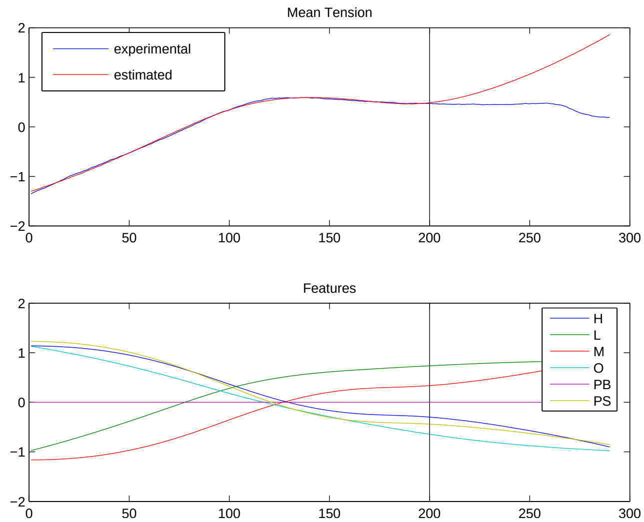


Figure 4-53: The first window shows the results of linear regression for Q09 (cadence with *crescendo* and *ritardando*). The blue line represents subject data and the red line represents the model. The vertical line indicates the division between training data for the model and out of sample data used for prediction. The second window shows the smoothed feature graphs.

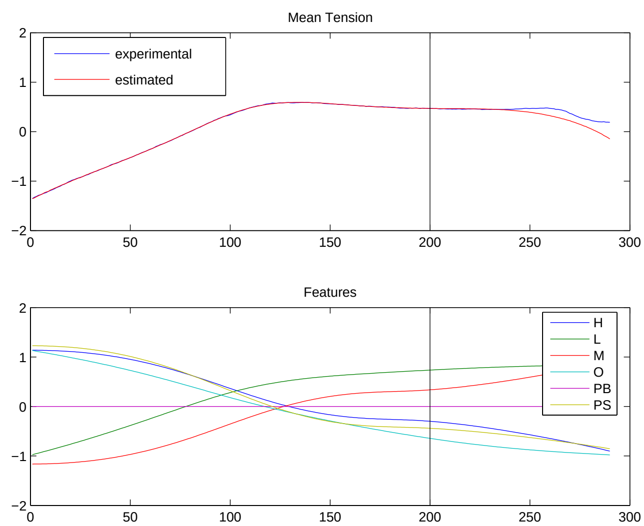


Figure 4-54: The first window shows the results of polynomial (quadratic in this case) regression for Q09 (cadence with *crescendo* and *ritardando*). The blue line represents subject data and the red line represents the model. The vertical line indicates the division between training data for the model and out of sample data used for prediction. The second window shows the smoothed feature graphs.

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

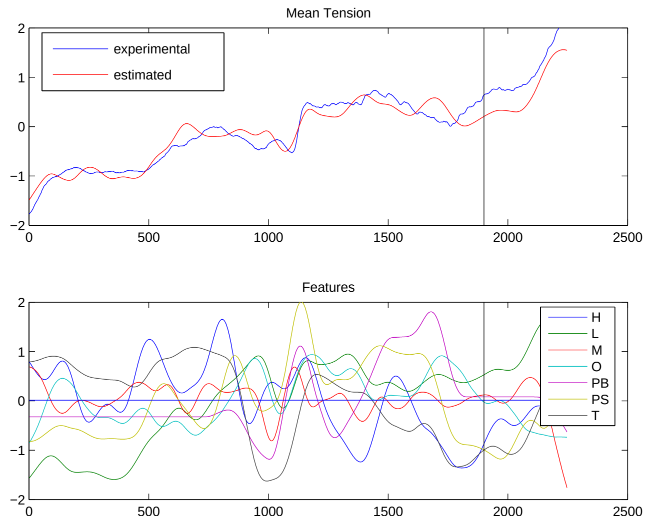


Figure 4-55: The first window shows the results of linear regression for Q10 (Brahms). The blue line represents subject data and the red line represents the model. The vertical line indicates the division between training data for the model and out of sample data used for prediction. The second window shows the smoothed feature graphs.

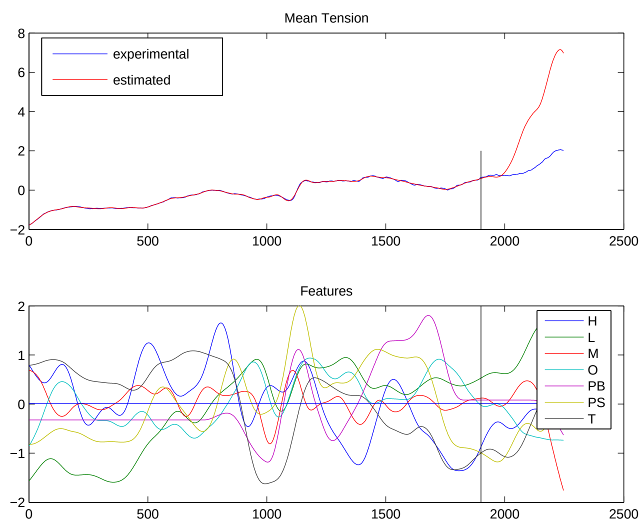


Figure 4-56: The first window shows the results of polynomial (quadratic in this case) regression for Q10 (Brahms). The blue line represents subject data and the red line represents the model. The vertical line indicates the division between training data for the model and out of sample data used for prediction. The second window shows the smoothed feature graphs.

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.

Figures 4-53 and 4-54 show the results of regression analysis performed on Q09. The results to the left of the vertical line show the model's fit of the training data. Results to the right are predictions by the model for the out of sample data. The quadratic model in this case is clearly better—it appears to capture the complexity of feature interactions in a way a linear model cannot.

Figures 4-55 and 4-56 show a case where a linear model works better. The quadratic model fits the data very well for the first part of Q10, but then diverges in second part. This example and others indicate that while it is possible to have an almost perfect fit of the training data by using higher-order models, the divergence for the out of sample data clearly shows that there is a danger of overfitting.

In conclusion, it appears that a fairly simple nonlinear model is sufficient to capture the complexity of the problem. For the most part, a linear model was adequate, however, there were some cases where polynomial models provided a better fit for both the training data and out of sample data. A general issue that needs to be considered is the fact that the training data in some examples was not sufficient enough 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.

4.3 Note on implementation of analysis tools

Scripts for preparing and processing data from both experiments were written in Perl. Additional Perl scripts and Excel spreadsheets were implemented for the statistical analysis of Experiment 1. Programs for analyzing data from Experiment 2 and building the model were written in Matlab using the CWM toolkit [Schoner, 2000]. In addition, an application was written in C++ and OpenGL (see Figure 4-57) to aid in visualizing and comparing subject data and feature graphs. The web-based application for Experiment 1 was written in Flash and PHP and the interface for Experiment 2 was written in C++ using the Win32 API.

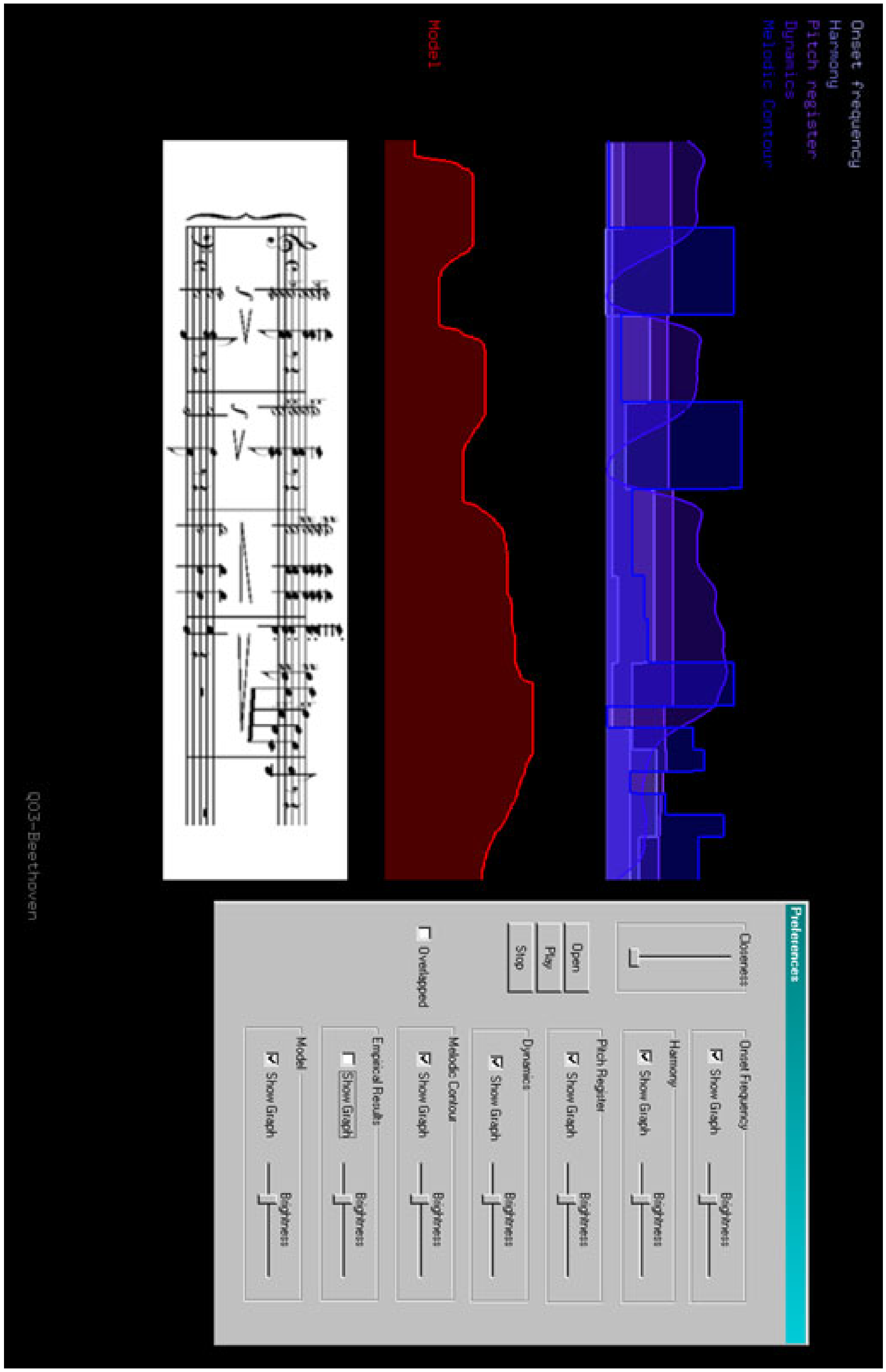


Figure 4-57: Screenshot of application written in OpenGL for visualizing data.

CHAPTER FIVE

Applications

"*Longum iter est per praecepta, breve et efficax per exempla.*"

– Seneca

Implementing an automated or computer-assisted composition system that generates music based on aesthetic or structural criteria is a difficult task. There are many steps required to define how an overarching description is translated to the note level. This process could be aided by a tension model, a tool especially relevant to the design of a musically "intelligent" system.

While existing computer-based composition environments might be able to generate music that sounds plausible or grammatically correct (at least on a small scale), the criteria for generating the material are usually purely mechanical and not aesthetically motivated at a higher level. Composing music is not just about putting together chords that follow logically, or patterns that are employed systematically. It is essential for these systems to have an understanding of music that goes beyond pitch sets, harmony, motivic patterns, phrases, and forms—they need to have some idea of the *meaning* of music. If a system has the ability to identify and measure tension, it might be possible for it to behave more creatively.

One such computer-assisted composition system that could benefit from a tension model is Hyperscore. Hyperscore is an application that facilitates composition by associating musical features with graphical abstractions. Lines drawn in the Hyperscore are interpreted according to shape, color, and position and converted into pitches and rhythmic values. There are two key creative aspects that are entirely in the hands of the users: composing short melodies or motivic material and describing (visually) the large-scale shape of a piece. Providing graphical means to engage in these two activities form the basis for Hyperscore's functionality [Farbood et al., 2004]. As mentioned earlier, the methodology in Experiment 1 was influence by the idea of line shapes in Hyperscore.

5.1 The original tension line in Hyperscore

The very first version of Hyperscore (Figure 5-1) was almost entirely automated.[1] The user drew a *tension line*, and the program then interpreted the shape of the line and generated a piece of music according to the shape and texture of the line. The line was parsed by the program into parabolic sections which were then interpreted as tension-release sections. The user selected all or a subset of nine pre-composed motives which were used to generate the piece. The generating function created five independent musical lines for each piece, all of which were generated sequentially. The main voice was generated in a different manner than the other four voices. The computer randomly chose one of the user-selected motives and added it to the main voice. The main voice was intended to serve as the anchor for the piece. It insured that at any point in time, there was some active motivic material playing. The other voices served as elaborations of the main voice. These elaborations were generated based on the tension value of the user-drawn curve. Depending on this value, the motive itself, a counter-motive, or rests were inserted. The tension value was determined by the sectioning of the line and the average bumpiness measure for each section. For example, if the curve was very smooth, the rhythmic texture would be less dense because there was a higher chance of rests being generated.

There were no deterministic thresholds in making these decisions. The tension value could only influence the overall effect produced by the combination of voices. Finally, the volume for each note was scaled to match the y-value of the curve. The algorithm used in Version 1 was primitive but surprisingly effective in producing musical textures that appeared to match the tension curve of the line. However, it lacked many important features that make music interesting—there was no sense of phrasing and no harmonic movement. Although there was some feeling of tonality, it was mainly a result of the general consonance created by a largely diatonic set of motives. Another feature lacking was the ability to annotate the tension curve in order to specify exactly what motivic material should be used.

5.2 Later versions

Version 2 (Figure 5-2) was the first attempt to experiment with annotations to the tension line. The graphical interface of Hyperscore completely redesigned[2] and the concept of colored "pens" introduced. This interface allowed users

[1] All versions of Hyperscore were written in C++ for the Windows platform.
[2] Egon Pasztor joined the project at this time and implemented the subsequent interfaces.

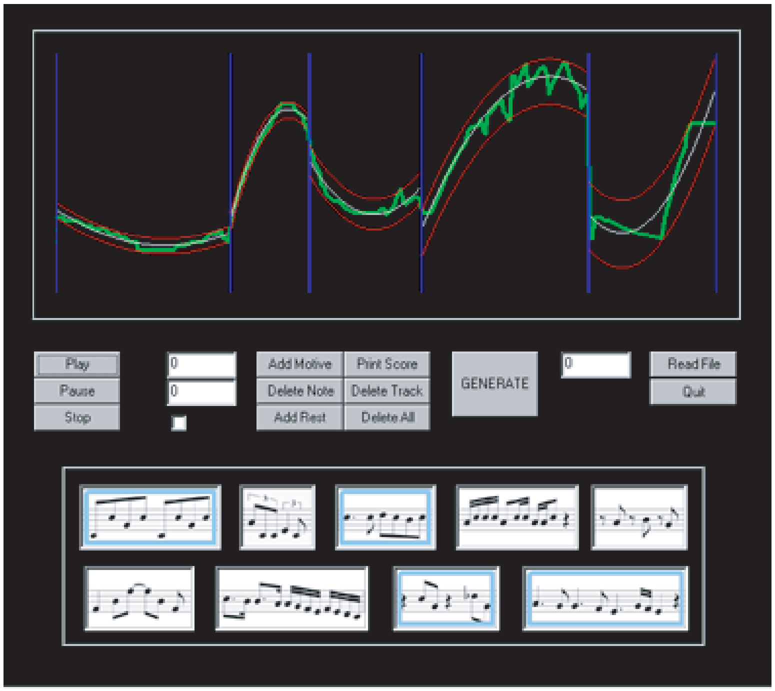


Figure 5-1: Screenshot of the first version of Hyperscore. The green line is drawn by the user, the blue, red, and white lines are added by the computer. The blue lines show the computed section divisions, the red lines show the min and max boundaries of the section, the white lines represent the estimated curves.

to indicate where and what kinds of melodic material were used by selecting and drawing with a color that was mapped to the motive. The annotation's proximity to the tension curve influenced what motive was selected by the generation algorithm. Since the tension curve was now a free-form line that could be drawn anywhere on the screen, a new localized texture measure was applied to determine the tension value.

Another new feature of Version 2 was the harmony generator. This algorithm was implemented using hierarchical Markov chains to handle different layers of organization. One set of Markov chains was used to generate a series of higher-level harmonic functions, and another set was used to generate the actual chords. The chord functions were simple, consisting only of three categories: tonic, dominant, subdominant. Chord function transition probabilities were selected based on the time at which the chord occurred and the function of the chord preceding it. The chords themselves were chosen according to time and relative frequency at which the chord would appear regardless of the circumstances (i.e. not dependent at all on the preceding chord).

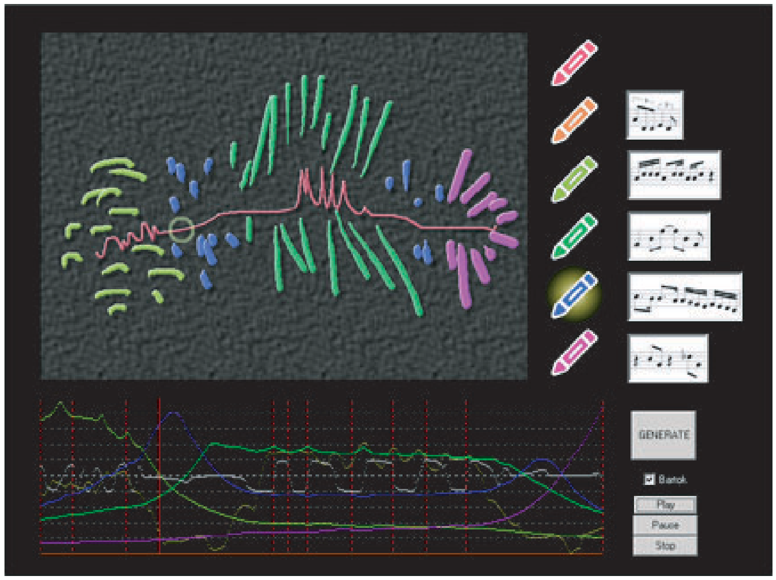


Figure 5-2: Screenshot of Hyperscore Version 2. The red pen, which has no motive attached to it, is used to draw the main curve. The other pens are for making annotations to the red line. The graphs below the drawing window indicate the prominence of each motive (colored lines) as well as the curviness of the tension line (white graph line) at any given point in time. The halo is a snapshot of a blinking cursor indicating playback position.

While the harmony generator did add more depth to the music composed, it was too simplistic to work with anything longer than half a minute of music since the progressions tended to meander without any sense of phrasing or direction. Nevertheless, the algorithm was able to produce some satisfactory progressions on occasion. It also became clear, after some user-testing, that people were often confused by what the computer was doing in response to the drawing. The bumpiness to rhythmic activity map in Version 1 was clearer because the segmentation algorithm provided a better visualization of the texture in a given area. The way annotations were interpreted in Version 2 was also unclear to users, particularly those with no musical training.

After considering the problems encountered in Versions 1 and 2, a different approach was devised for associating motives with annotations which involved more human decision-making and less automatic generation. Instead of influencing the decision-making process of the computer, the annotations deterministically dictated them. One significant result of this change was that the tension line ceased to function in an algorithmically interesting way and was reduced to a timeline and volume scale.[3]

[3]For more detailed information on Hyperscore Versions 1-4, see [Farbood, 2001].

Figure 5-3: Screenshot of Hyperscore Version 4. This version allowed users to enter their own motives instead of using pre-composed ones.

5.3 Final Hyperscore version

Hyperscore's development at this point (Version TS) took a more practical turn due to its role as the primary vehicle for composition activities in Tod Machover's Toy Symphony, a large project bringing together children and professional orchestras through the aid of technology. The goal of Toy Symphony was to introduce children to creative music-making with specially designed hardware and software. These tools allowed children to perform on stage with musicians as well as compose music that was performed by orchestras. It was essential to have a version of Hyperscore for Toy Symphony that was not just experimental in nature, but developed enough for users to compose original pieces of high quality. During the course of the Toy Symphony project (2002-2005) children from all over the world worked with the software to compose pieces for string orchestra, some of which were performed in concert by professional orchestras such as the BBC Scottish Symphony and the Deutsches Symphonie-Orchester Berlin [Machover, 2003].

Version TS incorporated 3D DirectX graphics to expand the visual environment. The sketch window evolved from a static, finite space to an expansive zoomable canvas where users could create any number of motives and pieces. These two types of musical material were encapsulated in individual windows that could be positioned anywhere on the canvas and viewed at four different zoom levels for ease of editing.

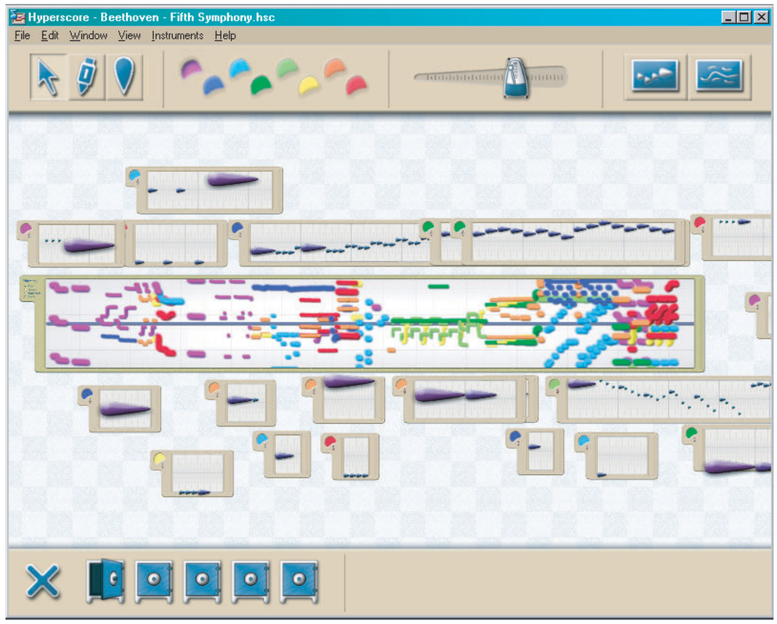


Figure 5-4: A Version TS Hyperscore rendering of the exposition of the first movement of Beethoven's Fifth Symphony.

The first step in composing a piece was to create some melodic material in the motive windows (Figure 5-5). The window's vertical axis represented pitch (spanning two octaves), and the horizontal axis represented time. The windows could be stretched or shortened depending on the length of the motive. Purple droplets represented notes, and users added them by clicking on the grid. Blank spaces were interpreted as rests.

The user could then choose a color for each motive and compose a piece by selecting a pen color and drawing into a sketch window. Every time the user drew a line of a particular color, Hyperscore would insert the motive mapped to that color into the piece. The start and end points of the line determined how many times a motive repeated, and a fixed pixel-to-duration metric calculated the length of time a line would play. If the length of a line did not divide evenly into whole repetitions of a motive, then a fragment of the motive was used for the last iteration.

Drawing a straight line would make the motive repeat with the precise melodic intervals of the original motivic material. The vertical position determined how much the motive was transposed up or down. Curves and bends in the line imposed a pitch envelope on the motive's repetitions but did not alter the melodic contour to the point that the new material was unrecognizable from the original motive (Figure 5-6). Lines could be reshaped by right-clicking and

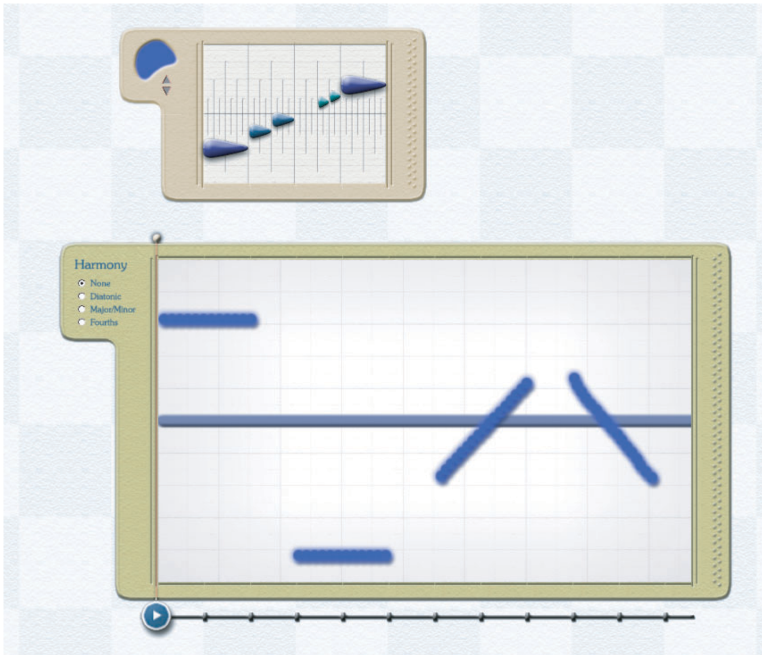


Figure 5-5: A motive window and sketch window with lines drawn in the color associated with the motive.

then dragging. Other editing features included cutting and pasting, changing instrumentation, and increasing or decreasing playback volume.

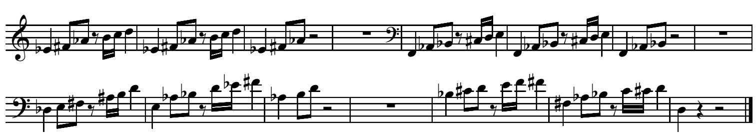


Figure 5-6: Musical realization of the Hyperscore piece shown in Figure 5-5.

The lines drawn into a sketch window combined to form larger, multi-voiced segments of music. These lines could communicate a musical gesture when effectively interwoven and overlapped. Hyperscore was able to facilitate composition by providing a visual representation of the large-scale structure of a piece and simplifying the process of integrating musical material. This representation provided high-level control over the dramatic arc of the piece as a whole as well as the placement of individual motivic elements.

As in previous versions, all sound output was in MIDI format, and either the computer's sound card or an external MIDI synthesizer acted as the output device. Users could also save Hyperscore pieces as MIDI files, a standard format that could be read into any notation program such as Finale or Sibelius. This made it straightforward to go from Hyperscore format to musician-readable format, giving a composer the option of sketching out a composition in Hyperscore and then editing in standard notation.

Version TS addressed harmony in two different ways. In the simplest case, harmony could be a single chord without a reference point and without regard to what preceded or followed it. Users could add individual chords consisting of three simultaneous voices to the sketch window. They were displayed as colored droplets, with each color representing a different harmony type: major, minor, augmented, diminished, and so forth (Figure 5-7). The second type of harmony control utilized what used to be the tension line in previous versions. While the idea of an all-purpose tension line was never fully realized in Version TS, a tension line that focused on a single parameter—harmony—was implemented instead.

Figure 5-7: Colored droplets representing individual chords. The colors were mapped to different chord qualities.

5.3.1 Harmonic tension line

One reason for having a graphical notation system in the form of freehand drawing was to provide the user with an expressive means of shaping musical direction. Drawing a contour is a simple and intuitive way to depict areas of harmonic tension and resolution.

The algorithm for this new "harmony line" was based in part on David Cope's Experiments in Musical Intelligence (EMI). EMI takes existing works in a given style, segments them into musical fragments and then reconstitutes them in an intelligent way to form new pieces in the same style. Cope likens it to a version of *Musikalishes Würfelspiel* (musical dice game), an eighteenth century piece attributed to Mozart consisting of sixty-four bars of musical material that are randomly put together to form a coherent whole. EMI creates a database from existing music by performing functional analysis based on ideas from Schenkerian analysis, and then generates new music using this database.

EMI's musical input consists of events that describe the note attributes of pitch, timing, duration, dynamic, and channel (MIDI events). A database of musical fragments is created by analyzing and segmenting the music. The analysis process uses a classification system of functional identifiers called SPEAC (for Statement, Preparation, Antecedent, and Consequent). Pattern matching is used to determine what recurring signatures should *not* be segmented; it is important that certain signatures remain intact because they are necessary for the stylistic identity of the music. The segments are then placed in a lexicon according their SPEAC meaning. New music is generated by using an augmented transition network (ATN) to recombine musical segments from the database.[4]

Some of the musical works generated by EMI are extremely convincing. The very best examples are not easy to differentiate from the representative works they are intended to emulate. Cope believes his EMI system paralleled what takes place at some level in composers' minds, whether consciously or subconsciously: "The genius of a great composer, I believe, lies not in inventing previously unimagined music but in their ability to effectively reorder and refine what already exists." [Cope, 1996]

Cope's idea of classifying functional identifiers directly influenced the algorithm for interpreting Hyperscore's harmony line. In Hyperscore, users described harmonic progressions by shaping the harmony line. It was parsed into sections [Pasztor, 2002] which were then mapped to functional identifiers that resembled SPEAC. Hyperscore's identifiers had been modified from Cope's, and consisted of four categories: Statement, Antecedent, Consequent, and Modulation. The harmony line running through the center of each sketch window could be modified by clicking and dragging. Color bands would appear to indicate the line's parsing (Figure 5-8). Sections were classified as one of four visual types, each corresponding to a functional identifier:

- **Statement** - flat section, colored white. Musically defined as a statement or prolongation of the tonic.
- **Antecedent** - upward-sloping section, colored green. Musically defined as a combination of chords that need resolution (e.g. dominant chords or combinations of subdominant and dominant chords).
- **Consequent** - downward-sloping section, colored blue. Resolution of preceding Antecedent section. If not preceded by an Antecedent, then restates the tonic.
- **Modulation** - defined by a sharp pointed region or spike, colored yellow. Progression toward a new key.

[4]ATNs are used in natural language processing to construct grammatical sentences. They are context free grammars with an extension that defines constituents by a set of features, allowing aspects of natural language such as agreement and subcategorization to be handled in an intuitive and concise way [Allen, 1995].

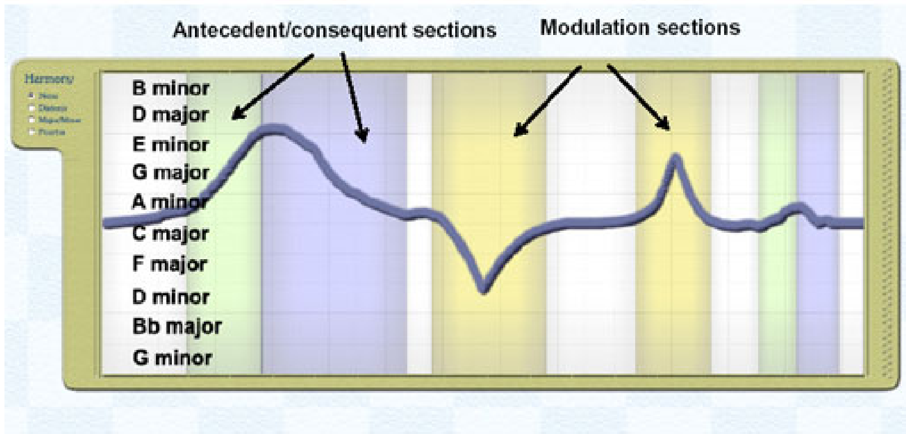


Figure 5-8: An empty Hyperscore sketch window showing the harmony line. The height or depth of the point indicates the key to which the section modulates (indicated by the text overlay).

After the line was parsed, chords were assigned to each section based on its functional identifier, how many beats it spanned, and how textured the section was (i.e. how bumpy it was). The instability of the chords assigned to a section was directly proportional to the amount of texture. The chords chosen were taken from a database that returned either single chords or small progressions based on the selection criteria. The database consisted of chord progressions commonly found in Bach chorales.

When the chords were chosen for the entire piece, the notes generated from the sketch were altered so that they matched either the currently assigned chord or a scale tone in the current key. For minor keys, there were special provisions for inserting a raised $\hat{6}$ or $\hat{7}$ depending on the chord and context. There were several criteria used in deciding how and in what direction a pitch was altered:

- **Beat** - If a pitch fell on a beat or was longer than a sixteenth note in duration, it would be harmonized as a chord tone. If it was short in duration and did not fall on a beat, it would be harmonized as a scale tone.[5]

- **Contour** - Notes were moved up or down as minimally as possible while attempting to preserve the contour of the original melodic material. Even if the original pitch was a valid chord tone before being harmonized, it might still be altered if it distorted the overall melodic contour.

[5] This metric heavily favors minimization of nonharmonic tones and was a deliberate decision given the requirements of the Toy Symphony project.

- **Voice** - The voice determined to be the bass line did not have the strict melodic contour requirements and could be altered radically to fit not just the nearest chord tone, but the bass note of the chord (root or inversion). This did not apply in the case when there was only a single active line (a solo voice).

Users chose from four harmony styles: none, diatonic, major-minor, and fourths. "None" meant that no automatic harmonization was applied. Diatonic mode changed all chromatic pitches into diatonic ones in the current key (defined by the presence of Modulation sections in the harmony line). Major-minor was eighteenth-century-style tonal harmony. Fourths mode was based on chords constructed from fourths rather than thirds. Although fourths mode used the same database as major-minor mode, some of the chord root notes were altered to fit the functional identifiers more closely. For example, the fourths-mode equivalent to a dominant seventh chord was a chord built on the leading tone, giving it a stronger pull toward the tonic.

Figures 5-11 and 5-12 show different harmonic realizations of the first section of the Hyperscore piece shown in Figure 5-9. Aside from a complete harmonization done with regard to a harmonic progression generated from the harmony line, there was an additional option of selecting any subset of the lines drawn in the sketch window to be unharmonized within the current tonal context. This selection was indicated visually by giving the line color a darker tint. The effect of unharmonizing individual lines did not revert the line to its original chromatic form—it altered all necessary pitches to correspond to scale tones in the current key rather than chord tones (Figure 5-13).

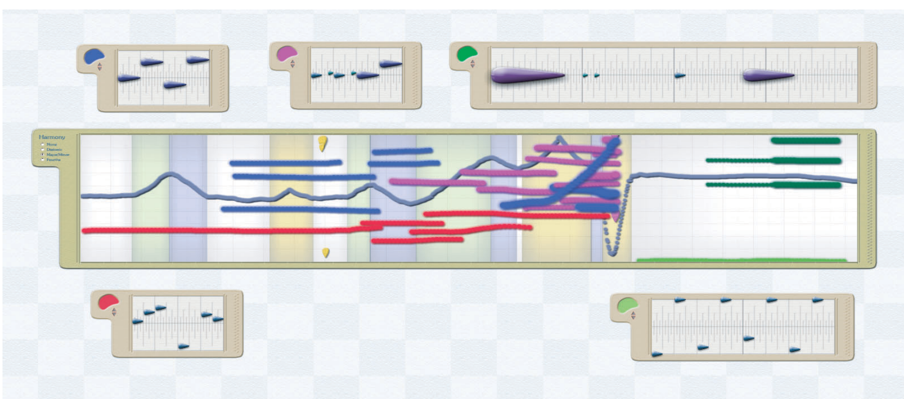
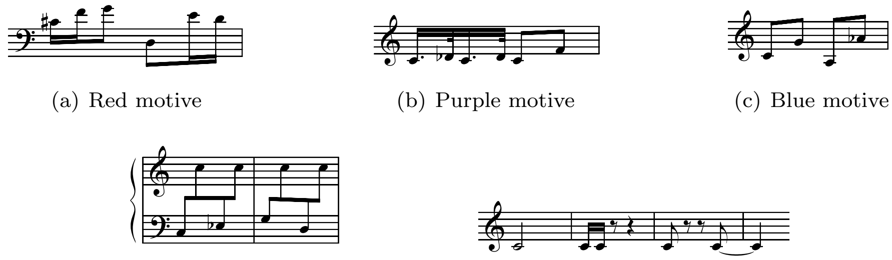


Figure 5-9: Example of a Hyperscore piece that has been harmonized.



(a) Red motive (b) Purple motive (c) Blue motive

(d) Light green motive (e) Dark green motive

Figure 5-10: Musical realizations of motives in Figure 5-9.

5.4 The future of Hyperscore

There are many additional features that should be implemented in the future if Hyperscore is to realize its full potential as a tool to aid both novices and professionals in composing music. One major change would be to allow direct editing at the individual note level within the sketch window as opposed to permitting such changes only through altering motives or inaccurately bending lines. The algorithm for parsing the harmony line needs to be improved in order to allow more precise control. Adding external methods of inputting motivic material—both in audio and MIDI formats—would also be useful.

Another idea is a reverse Hyperscore, where the input is a piece of music (in MIDI format, for example) and the output is a Hyperscore rendering. This would be a far more difficult task than the current graph-to-music approach. There would need to be some concrete method of breaking down a piece into basic motivic elements, perhaps by doing a statistical analysis of recurring rhythmic, melodic, and harmonic patterns. This process would be greatly assisted by a special type of line (perhaps colored a neutral gray) that would allow the addition of musical material that is not associated with a motive. After all, while much of music consists of recurring motives, not all of it does. This is a flaw in the current Hyperscore paradigm that needs to be fixed.

5.5 A new tension line

Harkening back to the original idea of a general tension line, such a feature incorporated into Hyperscore could be a powerful tool for helping users analyze and compose music. It would be difficult to implement, but having a model like the one presented in this thesis is a significant start. The greatest barrier would be the implementation of the theoretical background necessary to quantify individual parameters such as harmony. For example, it would be an extensive project to implement Lerdahl and Jackendoff's generative theory of music (particularly the prolongational reduction) in order to correctly quantify

Figure 5-11: The unharmonized musical realization of the Hyperscore piece shown in Figure 5-9.


C: I
V7
I
11
ii7
V7/vi
vi a: i
VI7
V7
i
i6
i

Figure 5-12: Musical realization of the harmonized Hyperscore piece shown in Figure 5-9. Roman numerals indicate the chords generated by the harmony line.

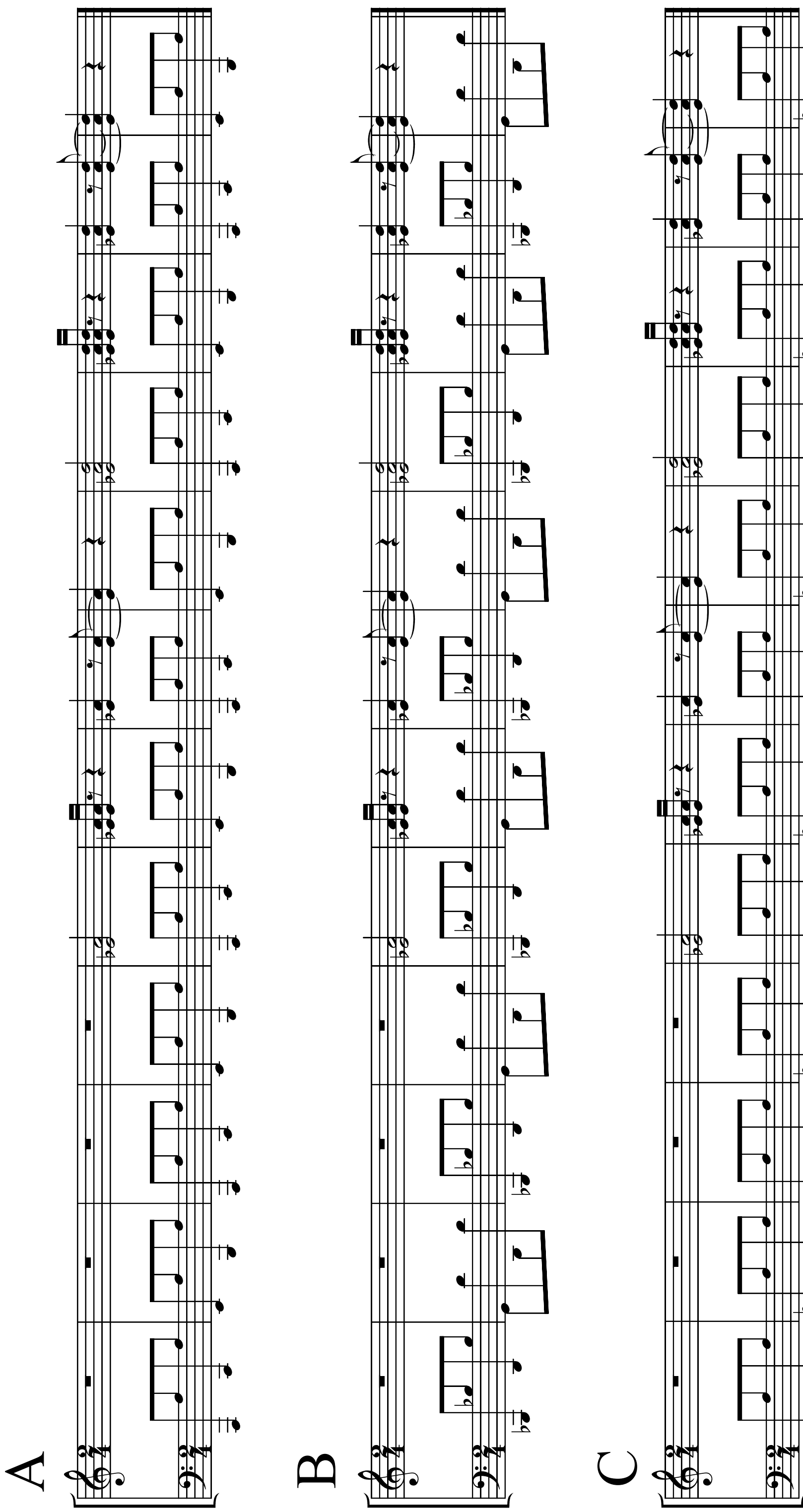


Figure 5-13: Three versions of the latter part of the Hypscore piece shown in Figure 5-9. (A) does not have any harmonization applied. (B) is harmonized in the standard way. (C) Shows the effects of another harmonization option that allows users to selectively unharmonize lines within the current tonal context.

chord distances according to Lerdahl's tonal tension model. Assuming this part of the analysis has been implemented or some acceptable substitutes have been found, the actual functionality of the tension model within Hyperscore would be manifested in not just a single tension line, but multiple tension *lines*.

These tension lines would be present not in the center of the sketch window, but rather as separate detachable windows that could be applied as a global modifier to any Hyperscore sketch window. These lines would either be viewed individually or as a single line representing the output of the global tension model given the individual parameters of harmony, melodic contour, loudness, pitch height, tempo, and onset frequency. This line could at any point be separated into its component parts again. The individual lines as well as the global line could be modified, resulting in Hyperscore recomposing or adding new material based on the material the user has already produced.

As in Version 1, the result of this process would not be deterministic, but produce a result that can either be kept by the user or thrown out. The user could then request more new material without changing any of the current parameters. One practical application of this would be, for example, creative assistance—perhaps a user is looking for some new ideas or is dissatisfied with a particular section and cannot see a clear way to improve it. Being able to see the musical material objectively analyzed and presented in a high-level manner might clarify any problems. The musical parameters (either individually or in any combination) could then be modified to fit the composer's vision more accurately. If that weren't enough, the program would be able to step in and produce interesting new musical material within the given framework.

CHAPTER SIX

Conclusions

"*Et le chemin est long du projet à la chose.*"

– Jean-Baptiste Poquelin Molière

6.1 Summary

A quantitative, parametric model for describing how listeners perceive changes in musical tension was presented. The model takes into account multiple, disparate musical features and is based directly on empirical evidence. As a quantitative model, it provides a foundation for implementation and incorporation into computer-based applications like Hyperscore.

6.1.1 Experimental results

Previous studies have shown that there are numerous parameters which contribute to a listener's response to music. Some of these are expressive features like dynamics and tempo; others are learned from listening to music specific to one's culture. In the latter category are ordered hierarchical relationships involving pitch, harmony, and tonality. The model presented in this thesis takes into account features in both categories and formalizes how they reinforce or detract from a listener's perception of tension. Due to the complex and subjective nature of tension, the model, by necessity, is based on empirical data.

Two experiments with contrasting approaches were described. Experiment 1 was a web-based study designed to collect data from thousands of subjects from all over the world. Subjects were asked to listen to musical excerpts and then indicate how they felt the tension was changing by choosing from nine possible tension curves depicted by line graphs. Experiment 2 collected subject responses to tension in real time. Subjects were asked to move a slider on a computer interface to indicate how they felt tension was changing as they listened.

Most of the musical excerpts used in the experiments were composed expressly for the studies. In addition to these examples, several excerpts from the classical repertoire were also selected (from pieces by Bach-Vivaldi, Beethoven, Brahms, and Schönberg). Musical excerpts composed specifically for the studies were designed to isolate and combine changes in harmony, melodic expectation, pitch height, tempo variation, onset frequency, dynamics, and rhythmic irregularity. Some examples consisted of a single feature changing over time in a continuous and straightforward manner, while others included two or more features either in concert or opposition to one another. The examples were carefully composed so that interference from other variables would be easily detected if not completely absent.

Analysis of data from Experiment 1 clearly demonstrated that listeners' perception of tension was influenced by all features considered with the exception of rhythmic irregularity. While pitch height appeared to have the strongest effect, that might have been a consequence of the obvious mapping of curve shape to melodic contour. Onset frequency appeared to be the weakest factor, particularly when opposed by other features. However, in general, it was difficult to provide a precise ordering for the relative influence of each individual feature on tension. To make it absolutely clear that musical features were changing in one "direction" or another, the examples were short and exhibited exaggerated changes. It is appears that the *obviousness* of how a feature was changing (salience) directly contributed to its overall influence. Although pitch height seems to have had the strongest effect, subjects' responses might have been significantly different had the changes been more subtle. Nonetheless, the results of Experiment 1 confirmed that the parameters considered were in fact legitimate and that changes in these parameters resulted in changes in tension. The analysis of data collected in Experiment 2 proceeded from this confirmation.

6.1.2 The model

All of the excerpts used in Experiment 2 were described quantitatively in terms of how the individual musical features were changing. All of the features confirmed in Experiment 1 were included. Onset frequency and tempo were treated as separate parameters, and one new feature, melodic expectation, was also added. Three theoretical models—Lerdahl's tonal tension model, Jehan's psychoacoustic loudness model, and Margulis' melodic expectation model—were utilized to describe the multidimensional features of harmonic tension, loudness, and melodic expectation. These models were assumed to accurately represent their respective parameters. Tempo, onset frequency, and pitch height are one-dimensional features that were straightforward to quantify and did not require external sources of description.

A new model was implemented that mathematically described and predicted how listeners perceived tension in excerpts from Experiment 2 given the descriptions of the individual musical parameters. Linear and polynomial regression was performed using the first half to three-quarters of each excerpt as training data. The last part of the excerpt was left to test the predictive capabilities of the model. The degree to which the model fit this last part (the out-of-sample data) was a strong indicator of its accuracy.

Results of the regression analysis showed that in many cases, a linear model was adequate. However, there were some cases where a polynomial model provided a better fit, though there was some risk of overfitting. Given these results, it appears that a simple nonlinear model is sufficient to capture the complexities of the tension model.

One general issue that needs to be considered is the fact that the training data was probably not sufficient to thoroughly "learn" all of the necessary interactions between the parameters. This problem was more acute for shorter examples, where the brief time-span of musical events covered by the training data did not contain the amount of information needed to predict future changes.

6.1.3 Musicians vs. non-musicans

The results of comparing responses of musically inexperienced and musically experienced subjects in both experiments indicated that experienced musicians have a greater sensitivity to harmony in particular. This was clearly evident in Experiment 1, where there was a significant statistical difference between how musicians and non-musicians responded. Results from the first experiment also indicated that musicians are more responsive to onset frequency and changes in tempo than non-musicians.

6.1.4 Applications

Hyperscore, a computer-assisted composition system, was described and discussed. The incorporation of a tension model into future versions of Hyperscore would add powerful new capabilities to the application. Such a model could provide users with high-level feedback about their music and could be used in a generative format, allowing composers to make high-level changes that either affect individual musical parameters directly or affect all of them together through a global tension line.

6.2 Discussion

In the course of examining and analyzing the data from Experiments 1 and 2, there were a number of problems 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 there, 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 completely different note, there was so much data collected in Experiment 1, that not all of it could be analyzed. There still lies 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.

In Experiment 2, one of the most problematic issues was the considerable freedom given to the subjects in how they moved the slider in response to the stimuli. In retrospect, there should have been stronger constraints on where the subjects started on the slider and perhaps even where they ended. Although normalization methods were employed to correct these problems to some extent, it goes without saying that more consistent responses result in more accurate analyses.

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 is possible, if difficult, from a technical point of view 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.

Assuming there were significantly more subjects for a study like Experiment 2 (maybe not thousands, but at least hundreds), the possibility of coming up with a model that is truly universal could be possible. More subjects would make it possible to add more test examples. There are so many parameters being considered, that 10 excerpts are insufficient to cover enough changes and interactions between permutations of features required for a successful prediction.

Given a large number of examples designed to cover as many varying situations as possible, subjects would respond to only a subset of the possible examples, as in Experiment 1. In retrospect, even with only 35 subjects, it might have been better to use this type of method for Experiment 2.

6.3 Contributions

In conclusion, the contributions offered by this thesis include the following:

- A web-based study that successfully collected data from thousands of subjects and offers a new paradigm for gathering large amounts of data in music cognition experiments.
- A model of musical tension that takes into account expressive features as well as purely musical ones.
- An approach that considers both linear and nonlinear models derived entirely from empirical data.
- A computer-based application (Hyperscore) that could be used as a platform for the model.

6.4 Future directions and final remarks

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 (Figure 6-1 shows one such example). 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. Given that there is a large body of literature on perception of meter and other rhythmic structures, there might already be existing research beyond the scope of this thesis that sheds light on these issues (see [Hasty, 1997] for references).

Figure 6-1: Example from Experiment 1 showing meter changes.

As discussed, the implementation of the tension model within Hyperscore is a clear next step; it would, in a sense, employ the model in a "real-world" situation. The greatest difficulty would lie in implementing the theoretical background necessary to quantify individual parameters like harmony. Regardless, the model could still be implemented with a simpler (or even user-defined) measure of harmonic tension.

In conclusion, it is doubtful that our understanding of music perception will ever be complete. The work described here is just a small piece of the greater puzzle—a piece already complex and difficult to assess in itself. Nevertheless, it is at least a small step in the right direction toward understanding our perception and enjoyment of music.

APPENDIX A
Musical excerpts from Experiment 2

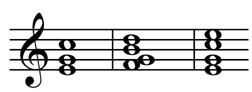


Figure A-1: Q01 (Composed for the experiment by the author.)

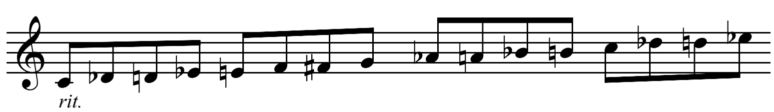


Figure A-2: Q02 (Composed for the experiment by the author.)

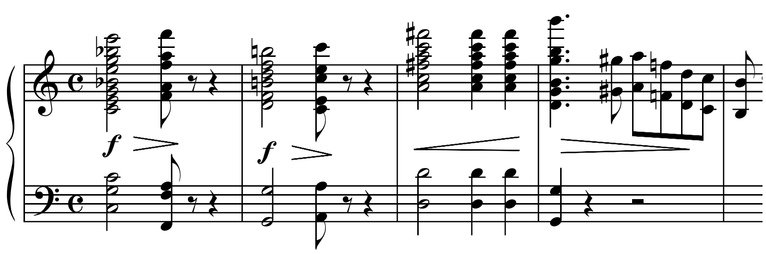


Figure A-3: Q03 (Excerpt from Beethoven's First Symphony.)

Figure A-4: Q04 (Composed for the experiment by the author.)

Figure A-5: Q05 (Excerpt from Bach's organ transcription of Vivaldi's C major concerto.)

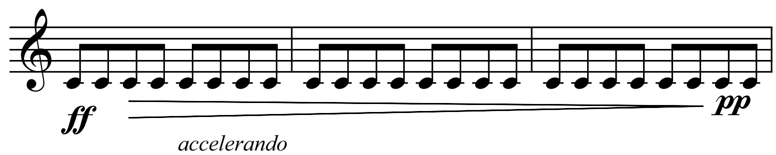


Figure A-6: Q06 (Composed for the experiment by the author.)

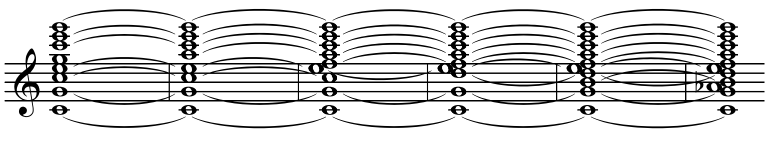


Figure A-7: Q07 (Composed for the experiment by the author.)

Figure A-8: Q08 (Excerpt from Schönberg's Klavierstück, Op. 11, No. 1.)

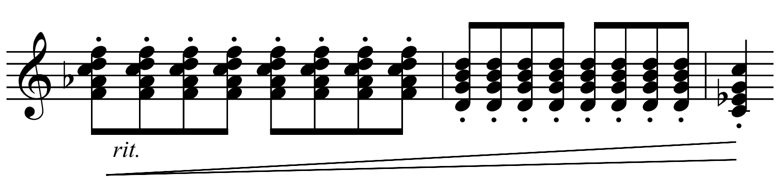


Figure A-9: Q09 (Composed for the experiment by the author.)

Figure A-10: Q10 (Excerpt from Brahm's Piano Concerto No. 2.)

APPENDIX B
Feature analysis from Experiment 2

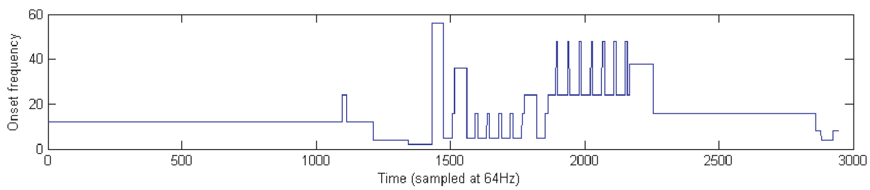


Figure B-1: Graph showing onset frequency values over time of Q10 (Brahms excerpt).

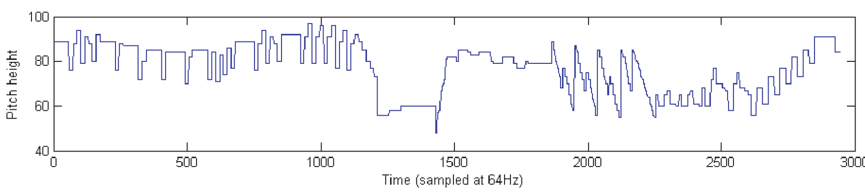


Figure B-2: Graph showing pitch height graph of Q10 (Brahms excerpt).

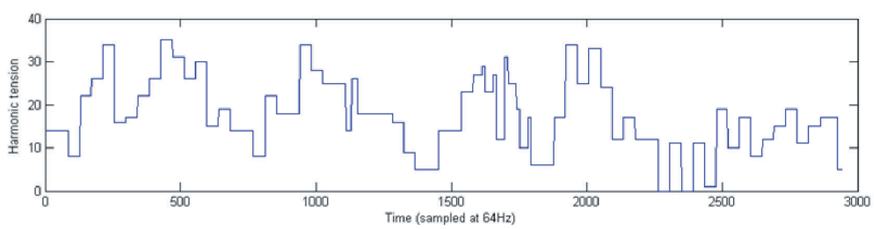


Figure B-3: Harmonic tension graph of Q10 (Brahms excerpt). The x-axis represents time, and the y-axis, tension values.

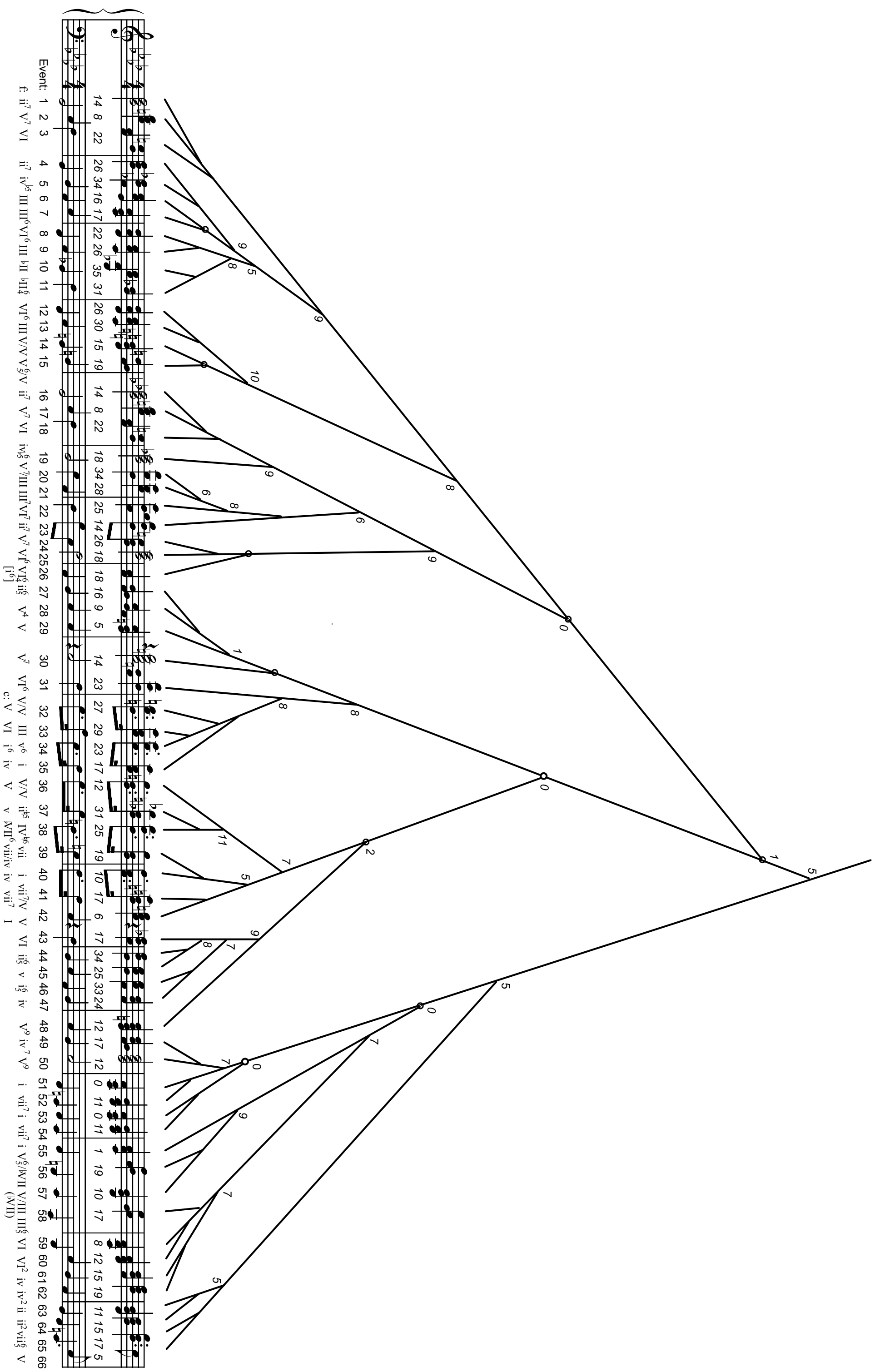


Figure B-4: Prolongational reduction of excerpt of Q10 (Brahms excerpt).

	Chords	i	j	k	δ	Inherited	Scale degree	Inversion	Non-harmonic tones	TOTAL	
d(1,2)	ii7-V7	0	1	5	6	6	1	0	1	14	d(1,2)
d(2,29)	V7-V	0	0	1	1	5	1	0	1	8	d(2,29)
d(3,2)	VI-V7	0	2	7	9	6	1	0	6	22	d(3,2)
d(4,6)	ii7-III	0	2	7	9	15	1	0	1	26	d(4,6)
d(5,4)	ivb5-ii7	0	3	6	9	24	1	0	0	34	d(5,4)
d(6,2)	III-V7	0	3	6	9	6	1	0	0	16	d(6,2)
d(7,6)	III6-III	0	0	0	0	15	0	2	0	17	d(7,6)
d(8,6)	VI6-III	0	1	4	5	15	0	2	0	22	d(8,6)
d(9,8)	III-VI6	0	1	4	5	20	1	0	0	26	d(9,8)
d(10,11)	bII-bII64	0	0	0	0	28	1	0	6	35	d(10,11)
d(11,8)	bII64-VI6	0	1	7	8	20	1	2	0	31	d(11,8)
d(12,14)	VI6-V/V	1	1	8	10	14	0	2	0	26	d(12,14)
d(13,12)	III-VI6	0	1	4	5	24	1	0	0	30	d(13,12)
d(14,2)	V/V-V7	1	1	6	8	6	1	0	0	15	d(14,2)
d(15,14)	V65/V-V/V	0	0	1	1	14	1	2	1	19	d(15,14)
d(16,17)	ii7-V7	0	1	5	6	6	1	0	1	14	d(16,17)
d(17,2)	V7-V7	0	0	0	0	6	1	0	1	8	d(17,2)
d(18.17)	VI-V7	0	2	7	9	6	1	0	6	22	d(18.17)
d(19.17)	iv6b5-V7	0	2	7	9	6	1	2	0	18	d(19.17)
d(20,21)	V7/III-III7	0	1	5	6	26	1	0	1	34	d(20,21)
d(21,22)	III7-VI7	0	1	5	6	20	1	0	1	28	d(21,22)
d(22,23)	VI7-ii7	0	1	7	8	12	1	0	4	25	d(22,23)
d(23,17)	ii7-V7	0	1	5	6	6	1	0	1	14	d(23,17)
d(24,25)	V7-VI6	0	2	7	9	15	1	0	1	26	d(24,25)
d(25,17)	VI6-V7	0	2	7	9	6	1	2	0	18	d(25,17)
d(26,25)	VI64-VI6	0	0	0	0	15	1	2	0	18	d(26,25)
d(27,28)	ii65-V4	0	2	4	6	6	1	2	1	16	d(27,28)
d(28,29)	V4-V	0	0	1	1	5	0	0	3	9	d(28,29)
d(29,51)	V-i	0	1	4	5	0	0	0	0	5	d(29,51)
d(30,29)	V7-V	0	0	1	1	5	1	0	7	14	d(30,29)
d(31,29)	VI6-V	0	2	6	8	13	0	2	0	23	d(31,29)
d(32,34)	V-i6 (c min)	0	1	4	5	21	1	0	0	27	d(32,34)
d(33,34)	VI-i6 (c min)	0	3	4	7	21	1	0	0	29	d(33,34)
d(34,31)	v6-VI6	0	2	6	8	13	0	2	0	23	d(34,31)
d(35,34)	iv-V (c)	0	1	4	5	21	1	0	0	27	d(35,34)
d(36,42)	V/V-V	1	1	5	7	5	0	0	0	12	d(36,42)
d(37,38)	i-VI6 (g)	0	3	4	7	23	1	0	0	31	d(37,38)
d(38,36)	I/Bb-V/cm	1	3	7	11	12	0	2	0	25	d(38,36)
d(39,40)	vii-i	0	2	6	8	10	1	0	0	19	d(39,40)
d(40,42)	i-i	0	1	4	5	5	0	0	0	10	d(40,42)
d(41,42)	vii7/V-V (f)	0	2	8	10	5	1	0	1	17	d(41,42)
d(42,29)	V-V	0	0	0	0	5	1	0	0	6	d(42,29)
d(43,48)	VI-V9	0	2	7	9	7	1	0	0	17	d(43,48)
d(44,45)	vii65-iii (Ab)	0	1	5	6	24	1	2	1	34	d(44,45)
d(45,43)	iii-IV (Ab)	0	2	6	8	16	1	0	0	25	d(45,43)
d(46,47)	i65-iv	0	1	5	6	23	1	2	1	33	d(46,47)
d(47,43)	iv-VI	0	3	4	7	16	1	0	0	24	d(47,43)
d(48,42)	V9-V9	0	0	2	2	5	1	0	4	12	d(48,42)
d(49,50)	iv7-V9	0	2	7	9	7	1	0	0	17	d(49,50)
d(50,51)	V9-i	0	1	6	7	0	1	0	4	12	d(50,51)
d(51,0)	i	0	0	0	0	0	0	0	0	0	d(51,0)
d(52,51)	vii7-i	0	2	7	9	0	1	0	1	11	d(52,51)
d(53,51)	i-i	0	0	0	0	0	0	0	0	0	d(53,51)
d(54,53)	vii7-i	0	2	7	9	0	1	0	1	11	d(54,53)
d(55,51)	i-i	0	0	0	0	0	1	0	0	1	d(55,51)
d(56,57)	V65-I	0	1	5	6	9	1	2	1	19	d(56,57)
d(57,55)	V/III-i	0	2	7	9	0	1	0	0	10	d(57,55)
d(58,59)	III65-VI	0	1	5	6	7	1	2	1	17	d(58,59)
d(59,55)	VI-i	0	3	4	7	0	1	0	0	8	d(59,55)
d(60,59)	VI2-VI	0	0	1	1	7	1	2	1	12	d(60,59)
d(61,59)	iv-VI	0	3	4	7	7	1	0	0	15	d(61,59)
d(62,61)	iv2-iv	0	0	1	1	14	1	2	1	19	d(62,61)
d(63,66)	ii-V	0	1	4	5	5	1	0	0	11	d(63,66)
d(64,63)	ii2-ii	0	0	1	1	10	1	2	1	15	d(64,63)
d(65,66)	vii65-V	0	3	5	8	5	1	2	1	17	d(65,66)
d(66,51)	V-i	0	1	4	5	0	0	0	0	5	d(66,51)

Figure B-5: Chart showing harmonic tension calculations of Q10 (Brahms excerpt). δ is the sum of i, j, and k.

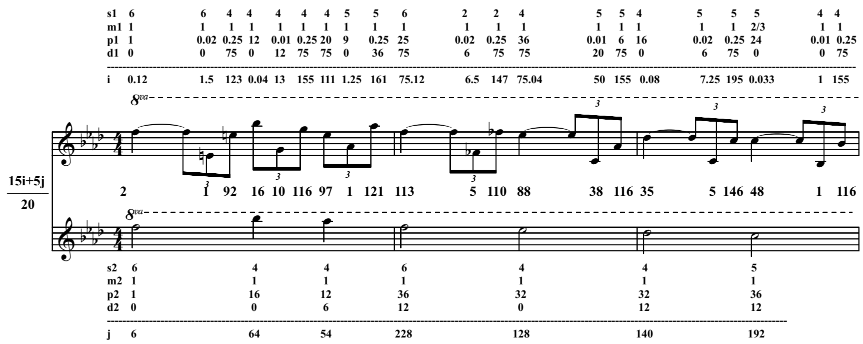


Figure B-6: First page of analysis showing how melodic expectation values are calculated for Q10 (Brahms excerpt). i values consist of direct note-to-note level expectations, and j values consist of high-level patterns based on salience and metrical placement. s represents stability ratings, m represents mobility ratings, p represents proximity ratings, and d represents direction ratings. The values between the two staves are the final melodic expectations ratings for each note.

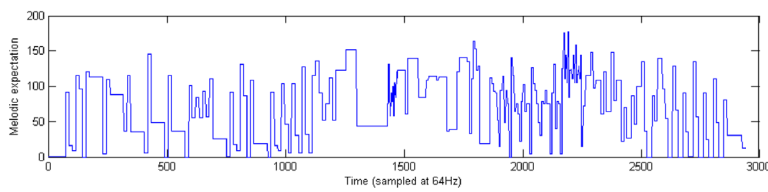


Figure B-7: Melodic expectation graph for Q10 (Brahms excerpt).

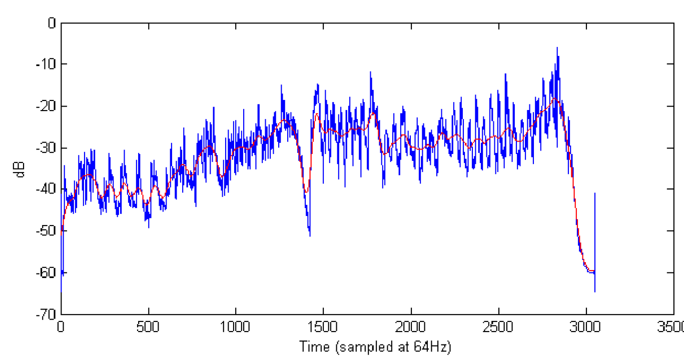


Figure B-8: Loudness curve of Q10 (Brahms excerpt). The blue line is the perceived loudness in dB produced by Jehan's psychoacoustic model, and the red line is the smoothed version used for the feature graph in the analysis of Experiment 2.

Bibliography

[Abe and Hoshino, 1990] Abe, J. and Hoshino, E. (1990). Schema driven properties in melody cognition: Experiments on final tone extrapolation by music experts. *Psychomusicology*, 9:161–172.

[Allen, 1995] Allen, J. (1995). *Natural Language Understanding*. Addison-Wesley, Redwood City, California.

[Bent, 1987] Bent, I. (1987). *Analysis*. W. W. Norton, New York.

[Berger, 2001] Berger, J. (2001). *Musical Expectations*. Draft accepted for publication by MIT Press, Cambridge.

[Bharucha, 1984] Bharucha, J. (1984). Anchoring effects in music. *Cognitive Psychology*, 16:485–518.

[Bharucha, 1987] Bharucha, J. (1987). Music cognition and perceptual facilitation: A connectionist framework. *Music Perception*, 5(1).

[Bharucha, 1996] Bharucha, J. (1996). Melodic anchoring. *Music Perception*, 13.

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

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

[Bigand, 1993] Bigand, E. (1993). Contributions of music to research on human auditory cognition. In McAdams, S. and Bigand, E., editors, *Thinking in Sound: The Cognitive Psychology of Human Audition*. Clarendon Press, Oxford. Translated by D. Dusinberre.

[Bigand and Parncutt, 1999] Bigand, E. and Parncutt, R. (1999). Perceiving music tension in long chord sequences. *Psychological Research*, 62.

[Bigand et al., 1996] Bigand, E., Parncutt, R., and Lerdahl, F. (1996). Perception of musical tension in short chord sequences: The influence of harmonic function, sensory dissonance, horizontal motion, and musical training. *Perception & Psychophysics*, 58(1):125–141.

[Birkhoff, 1933] Birkhoff, G. D. (1933). *Aesthetic Measure*. Harvard University Press, Cambridge, Massachusetts.

[Boltz, 1993] Boltz, M. G. (1993). The generation of temporal and melodic expectancies during musical listening. *Perception & Psychophysics*, 53(6):585–600.

[Boltz and Jones, 1986] Boltz, M. G. and Jones, M. R. (1986). Does rule recursion make easier to reproduce? if not, what does? *Cognitive Psychology*, 18:389–431.

[Carlsen, 1981] Carlsen, J. C. (1981). Some factors which influence melodic expectancy. *Psychomusicology*, 1:12–29.

[Cope, 1996] Cope, D. (1996). *Experiments in Musical Intelligence*. A-R Editions, Madison, Wisconsin.

[Cross, 1995] Cross, I. (1995). Review of the Analysis and Cognition of Melodic Complexity: The Implication-Realization Model, by E. Narmour, Univ. of Chicago Press, Chicago, 1992. *Music Perception*, 12(4):486–509.

[Cuddy and Cohen, 1976] Cuddy, L. L. and Cohen, A. J. (1976). Recognition of transposed melodic sequences. *Quarterly Journal of Experimental Psychology*, 28:255–270.

[Cuddy et al., 1981] Cuddy, L. L., Cohen, A. J., and Mewhort, D. J. (1981). Perception of structures in short melodic sequences. *Experimental Psychology: Human Perception & Performance*, 7:869–883.

[Cuddy et al., 1979] Cuddy, L. L., Cohen, A. J., and Miller, J. (1979). Melody recognition: The experimental application of musical rules. *Canadian Journal of Psychology*, 33:148–157.

[Cuddy and Lunney, 1995] Cuddy, L. L. and Lunney, C. A. (1995). Expectancies generated by melodic intervals: Perceptual judgments of melodic continuity. *Perception & Psychopysics*, 57:451–462.

[Cuddy and Lyons, 1981] Cuddy, L. L. and Lyons, H. I. (1981). Music pattern recognition: A comparison of listening to and studying tonal structure and tonal ambiguities. *Psychomusicology*, 1:15–33.

[Dewar et al., 1977] Dewar, K. M., Cuddy, L. L., and Mewhort, D. J. (1977). Recognition memory for single tones with and without context. *Journal of Experimental Psychology: Human Learning and Memory*, 3:60–67.

[Dibben, 1999] Dibben, N. (1999). The perception of structural stability in atonal music: The influence of salience, stability, horizontal motion, pitch commonality, and dissonance. *Music Perception*, 16(3).

[Dowling, 1978] Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, 85:341–354.

[Dowling and Fujitani, 1971] Dowling, W. J. and Fujitani, D. S. (1971). Contour, interval, and pitch recognition in memory for melodies. *Journal of the Acoustical Society of America*, 49:524–531.

[Dowling and Harwood, 1986] Dowling, W. J. and Harwood, D. (1986). *Music cognition*. Academic Press, New York.

[Ebcioğlu, 1988] Ebcioğlu, K. (1988). An expert system for harmonizing four-part chorales. *Computer Music Journal*, 12(3):43–51.

[Eerola et al., 2002] Eerola, T., Toiviainen, P., and Krumhansl, C. L. (2002). Real-time prediction of melodies: Continuous predictability judgments and dynamic models. In *Proceedings of the 7th International Conference on Music Perception and Cognition*, Sydney.

[Farbood, 2001] Farbood, M. (2001). Hyperscore: A new approach to interactive, computer-generated music. Master's thesis, MIT Media Lab.

[Farbood et al., 2004] Farbood, M., Pasztor, E., and Jennings, K. (2004). Hyperscore: A Graphical Sketchpad for Novice Composers. *IEEE Computer Graphics and Applications*, 24(1).

[Fletcher and Munson, 1937] Fletcher, H. and Munson, W. (July 1937). Relation between loudness and masking . *Journal of Acoustic Society of America*, 9(1):1–10.

[Forte and Gilbert, 1982] Forte, A. and Gilbert, S. E. (1982). *Introduction to Schenkerian Analysis*. Norton, New York.

[Francès, 1958] Francès, R. (1958). *La perception de la musique*. Erlbaum, Hillsdale, NJ. Transl. by W. J. Dowling, 1988.

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

[Hasty, 1997] Hasty, C. (1997). *Meter as Rhythm*. Oxford University Press, Oxford.

[Hauser et al., 2005] Hauser, M. D., Cushman, F., Young, L. L., Jin, K.-X., Mikhail, J., and Salceda, A. (2005). Moral sense test. http://moral.wjh.harvard.edu.

[Jackendoff, 1987] Jackendoff, R. (1987). *Consciousness and the Computation Mind*. MIT Press, Cambridge.

[Jackendoff, 1991] Jackendoff, R. (1991). Musical parsing and musical affect. *Music Perception*, 9:199–230.

[Jehan, 2005] Jehan, T. (2005). *Creating Music by Listening*. PhD thesis, MIT Media Lab.

[Jones, 1987] Jones, M. R. (1987). Dynamic pattern structures in music: Recent theory and research. *Perception & Psychophysics*, 41:621–634.

[Jones and Boltz, 1982] Jones, M. R. and Boltz, M. B. (1982). Controlled attending as a function of melodic and temporal context. *Perception & Psychophysics*, 32:211–218.

[Jones et al., 1993] Jones, M. R., Boltz, M. G., and Klein, J. M. (1993). Expected endings and judged duration. *Memory & Cognition*, 21:646–665.

[Justus and Bharucha, 2001] Justus, T. C. and Bharucha, J. J. (2001). Modularity in musical processing: The automaticity of harmonic priming. *Journal of Experimental Psychology: Human Perception and Performance*, 27:1000–1011.

[Justus and Bharucha, 2002] Justus, T. C. and Bharucha, J. J. (2002). Music Perception and Cognition. In Yantis, S. and Pashler, H., editors, *Stevens' Handbook of Experimental Psychology, Volume 1: Sensation and Perception*. Wiley.

[Krumhansl, 1990] Krumhansl, C. L. (1990). *Cognitive Foundations of Musical Pitch*. Oxford University Press, New York.

[Krumhansl, 1995] Krumhansl, C. L. (1995). Music psychology and music theory: Problems and prospects. *Music Theory Spectrum*, 17:53–80.

[Krumhansl, 1996] Krumhansl, C. L. (1996). A perceptual analysis of mozart's piano sonata k. 282: Segmentation, tension, and musical ideas. *Music Perception*, 13(3):401–432.

[Krumhansl, 2002] Krumhansl, C. L. (April 2002). Music: A link between cognition and emotion. *Current Directions in Psychological Science*, 11(2):45–50.

[Krumhansl and Kessler, 1982] Krumhansl, C. L. and Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review*, 89:334–368.

[Krumhansl and Shepard, 1979] Krumhansl, C. L. and Shepard, R. N. (1979). Quantification of the hierarchy of tonal functions within a diatonic context. *Journal of Experimental Psychology*, 8:24–36.

[Lerdahl, 1989] Lerdahl, F. (1989). Atonal prolongational structure. *Contemporary Music Review*, 4:64–87.

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

[Lerdahl and Jackendoff, 1983] Lerdahl, F. and Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. MIT Press, Cambridge.

[Lerdahl and Krumhansl, 2003] Lerdahl, F. and Krumhansl, C. (2003). La teoría de la tensión tonal y sus consecuencias para la investigacin musical (the theory of tonal tension and its implications for musical research). In

Galán, J. M. and Villar-Taboada, C., editors, *Los últimos diez años en la investigación musical*. Servicio de Publicaciones de la Universidad de Valladolid.

[Machover, 2003] Machover, T. (2003). Toy symphony project. http://www.toysymphony.org.

[Manaris et al., 2005] Manaris, B., Romero, J., Machado, P., Krehbiel, D., Hirzel, T., Pharr, W., and Davis, R. (2005). Zipf's Law, Music Classification, and Aesthetics. *Computer Music Journal*, 29(1):55–69.

[Margulis, 2005] Margulis, E. H. (2005). A Model of Melodic Expectation. *Music Perception*, 22(4):663–714.

[Masson, 1986] Masson, M. E. (1986). Comprehension of rapidly presented sentences: The mind is quicker than the eye. *Journal of Memory and Language*, 25:588–604.

[Mathews et al., 1987] Mathews, M. V., Pierce, J. R., and Roberts, L. A. (1987). Harmony and new scales. In Sundberg, J., editor, *Harmony and tonality*. Royal Swedish Academy of Music, Stockholm.

[McAdams, 1989] McAdams, S. (1989). Psychological constraints on form-bearing dimensions in music. *Contemporary Music Review*.

[Meyer, 1956] Meyer, L. (1956). *Emotion and Meaning in Music*. University of Chicago Press, Chicago.

[Minsky, 1981] Minsky, M. (1981). Music, Mind, and Meaning. *Computer Music Journal*, 5(3).

[Mitchell, 1997] Mitchell, T. (1997). *Machine Learning*. McGraw Hill.

[Moore and Glasberg, 1995] Moore, B. C. J. and Glasberg, B. R. (1995). A revision of Zwicker's loudness model. *Acta Acustica*, 82:335–345.

[Narmour, 1977] Narmour, E. (1977). *Beyond Schenkerism. The Need for Alternatives in Music Analysis*. Chicago University Press, Chicago.

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

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

[Palisca and Bent, 2003] Palisca, C. and Bent, I. (2003). Theory, theorists. Accessed 12 June 2003, http://www.grovemusic.com.

[Palmer, 1996] Palmer, C. (1996). Anatomy of a performance: Sources of musical expression. *Music Perception*, 13(3):433–453.

[Palmer and Holleran, 1994] Palmer, C. and Holleran, S. (1994). Harmonic, melodic, and frequency height influences in the perception of multivoiced music. *Perception & Psychophysics*, 56:301–312.

[Papadopoulos and Wiggins, 1999] Papadopoulos, G. and Wiggins, G. (1999). AI Methods for Algorithmic Composition: A Survey, A Critical View, and Future Prospects. In *AISB Symposium on Musical Creativity*.

[Pardo and Birmingham, 2002] Pardo, B. and Birmingham, W. P. (2002). Algorithms for Chordal Analysis. *Computer Music Journal*, 26(2).

[Parncutt, 1989] Parncutt, R. (1989). *Harmony: A psychoacoustical approach*. Springer-Verlag, Berlin.

[Parncutt, 1994] Parncutt, R. (1994). Template-Matching Models of Musical Pitch and Rhythm Perception. *Journal of New Music Research*, 23.

[Pasztor, 2002] Pasztor, E. (2002). A graphical environment for gestural computer-aided composition. Master's thesis, MIT Media Lab.

[Patel, 2003] Patel, A. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, 6(7).

[Pearce et al., 2001] Pearce, M., Smaill, A., and Wiggins, G. (2001). Towards a framework for the evaluation of machine compositions. In *AISB Symposium on Artificial Intelligence and Creativity in the Arts and Sciences*, Brighton, UK.

[Posner and Snyder, 1975] Posner, M. I. and Snyder, C. R. R. (1975). Attention and cognitive control. In Solso, R. L., editor, *Information processing and cognition: The Loyola Symposium*. Lawrence Erlbaum, Hillsdale, New Jersey.

[Pressnitzer et al., 2000] Pressnitzer, D., McAdams, S., Winsberg, S., and Fineberg, J. (2000). Perception of music tension for nontonal orchestral timbres and its relation to psychoacoustic roughness. *Perception and Psychophysics*, 62(1).

[Roads, 1996] Roads, C. (1996). *Computer Music Tutorial*. MIT Press, Cambridge, Massachusetts.

[Roberts and Shaw, 1984] Roberts, L. A. and Shaw, M. L. (1984). Perceived structure of triads. *Music Perception*, 2:95–124.

[Rowe, 1993] Rowe, R. (1993). *Interactive Music Systems: Machine Listening and Composing*. MIT Press, Cambridge, Massachusetts.

[Schellenberg, 1996] Schellenberg, E. G. (1996). Expectancy in melody: Tests of the implication-realization model. *Cognition*, 58:75–125.

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

[Schenker, 1906] Schenker, H. (1906). *Harmony*. University of Chicago Press, Chicago. Tr. E. M. Borghese, 1980.

[Schenker, 1935] Schenker, H. (1935). *Free Composition*. Longman, New York. Tr. E. Oster, 1979.

[Schmuckler, 1989] Schmuckler, M. A. (1989). Expectation in music: Investigation of melodic and harmonic processes. *Music Perception*, 7:109–150.

[Schmuckler, 1990] Schmuckler, M. A. (1990). The performance of global expectations. *Psychomusicology*, 9:112–147.

[Schmuckler and Boltz, 1994] Schmuckler, M. A. and Boltz, M. G. (1994). Harmonic and rhythmic influences on muiscal expectancy. *Perception & Psychophysics*, 56(3):313–325.

[Schoner, 2000] Schoner, B. (2000). *Probabilistic characterization and synthesis of complex driven systems*. PhD thesis, MIT Media Lab.

[Schroeder, 1991] Schroeder, M. (1991). *Fractals, chaos, power laws: minutes from an infinite paradise*. W. H. Freeman, New York.

[Sloboda, 2003] Sloboda, J. (2003). Psychology of music. Accessed 11 June 2003, http://www.grovemusic.com.

[Temperley, 2001] Temperley, D. (2001). *The Cognition of Basic Musical Structures*. MIT Press, Cambridge, MA.

[Temperley and Sleator, 1999] Temperley, D. and Sleator, D. (1999). Modeling Meter and Harmony: A Preference-Rule Approach. *Computer Music Journal*, 23(1).

[Tenney and Polansky, 1980] Tenney, J. and Polansky, L. (1980). Temporal Gestalt Perception in Music. *Journal of Music Theory*.

[Thompson and Cuddy, 1989] Thompson, W. F. and Cuddy, L. L. (1989). Sensitivity to key change in chorale sequences: A comparison of single voices and four-voice harmony. *Music Perception*, 7:151–168.

[Thompson et al., 1997] Thompson, W. F., Cuddy, L. L., and Plaus, C. (1997). Expectancies generated by melodic intervals: Evaluation of principles of melodic implication in a melody-completion task. *.Perception & Psychophysics*, 59.

[Tillmann and Bigand, 2004a] Tillmann, B. and Bigand, E. (2004a). Musical priming: Schematic expectations resist repetition priming. In *Proceedings of the Eighth International Conference on Music Perception and Cognition*, pages 674–676, Adelaide, Australia.

[Tillmann and Bigand, 2004b] Tillmann, B. and Bigand, E. (2004b). The relative importance of local and global structures in music. *Journal of Aesthetics and Art Criticsm*, 62(2).

[Toiviainen and Krumhansl, 2003] 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.

[Unyk and Carlsen, 1987] Unyk, A. M. and Carlsen, J. C. (1987). The influence of expectancy on melodic perception. *Psychomusicology*, 7:3–23.

[von Helmholtz, 1877] von Helmholtz, H. L. F. (1877). *Die Lehre von der Tonempfindungen als physiologische Grundlage für die Theorie der Musik*. Dover, New York. Transl. by A. J. Ellis, 1954.

[Voss and Clarke, 1975] Voss, R. F. and Clarke, J. (1975). 1/f noise in music and speech. *Nature*, 258:317–318.

[Voss and Clarke, 1978] Voss, R. F. and Clarke, J. (1978). 1/f noise in music: Music from 1/f noise. *Journal of the Acoustical Society of America*, 63(1):258–263.

[Widmer, 1992] Widmer, G. (1992). The Importance of Basic Musical Knowledge for Effective Learning. In Mira Balaban, K. E. and Laske, O., editors, *Understanding Music with AI: Perspective on Music Cognition*. MIT Press.

[Widmer, 1998] Widmer, G. (1998). Applications of Machine Learning to Music Research: Empirical Investigations into the Phenomenon of Musical Expression. In Michalski, B. and Kubat, editors, *Machine Learning and Data Mining: Methods and Applications*. Wiley, Chichester, UK.

[Widmer, 2000] Widmer, G. (2000). On the Potential of Machine Learning for Music Research. In Miranda, E., editor, *Readings in Music and Artificial Intelligence*. Harwood Academic Publishers, London.

[Wiggins, 1998] Wiggins, G. (1998). Music, Syntax, and the Meaning of "Meaning". In *Symposium on Music and Computers*, Corfu, Greece.

[Zwicker and Fastl, 1999] Zwicker, E. and Fastl, H. (1999). *Psychoacoustics: Facts and Models*. Springer Verlag, Berlin.