A Derivation of the Tonal Hierarchy from Basic Perceptual Processes

David Smey
Graduate Center, City University of New York

How does access to this work benefit you? Let us know!

Follow this and additional works at: http://academicworks.cuny.edu/gc_etds

Part of the Cognitive Psychology Commons, and the Music Commons

Recommended Citation
Smey, David, "A Derivation of the Tonal Hierarchy from Basic Perceptual Processes" (2014). CUNY Academic Works.
http://academicworks.cuny.edu/gc_etds/312

This Dissertation is brought to you by CUNY Academic Works. It has been accepted for inclusion in All Graduate Works by Year: Dissertations, Theses, and Capstone Projects by an authorized administrator of CUNY Academic Works. For more information, please contact deposit@gc.cuny.edu.
A DERIVATION OF THE TONAL HIERARCHY FROM BASIC PERCEPTUAL PROCESSES

by

David Smey

A dissertation submitted to the Graduate Faculty in Music in partial fulfillment of the requirements for the degree of Doctor of Philosophy,
The City University of New York

2014
This manuscript has been read and accepted for the Graduate Faculty in Music in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy

________________     _______________________________
Date          Stephen Blum
Chair of Examining Committee

________________     _______________________________
Date          Norman Carey
Executive Officer

Supervisory Committee
Joseph N. Straus, advisor
Philip Rupprecht, first reader
Stephen Blum
Fred Lerdahl

THE CITY UNIVERSITY OF NEW YORK
Abstract

A DERIVATION OF THE TONAL HIERARCHY FROM BASIC PERCEPTUAL PROCESSES

by

David Smey

Advisor: Professor Joseph N. Straus

In recent decades music psychologists have explained the functioning of tonal music in terms of the tonal hierarchy, a stable schema of relative structural importance that helps us interpret the events in a passage of tonal music. This idea has been most influentially disseminated by Carol Krumhansl in her 1990 monograph *Cognitive Foundations of Musical Pitch*. Krumhansl hypothesized that this sense of the importance or centrality of certain tones of a key is learned through exposure to tonal music, in particular by learning the relative frequency of appearance of the various pitch classes in tonal passages. The correlation of pitch-class quantity and structural status has been the subject of a number of successful studies, leading to the general acceptance of the pitch-distributional account of tonal hierarchy in the field of music psychology.

This study argues that the correlation of pitch-class quantity with structural status is a byproduct of other, more fundamental perceptual properties, all of which are derived from aspects of everyday listening. Individual chapters consider the phenomena of consonance and dissonance, intervallic rootedness, the short-term memory for pitch collection, and the interaction of temporal ordering and voice-leading that Jamshed Bharucha calls melodic anchoring. The study concludes with an elaborate self-experiment that observes the interaction of these
properties in a pool of 275 stimuli, each of which is constructed from a single dyad plus one subsequent tone.
Acknowledgements

This project is the result of a rather extended period of introspection and exploration, and I am grateful to the faculty of the CUNY Graduate Center for the generous patience, scholarly latitude and unflagging encouragement that allowed me to undertake such a journey. In particular I am deeply indebted to my advisors Philip Rupprecht and Joseph Straus, who have provided incisive feedback and moral support over the years.

I would also like to express my gratitude for the many scholars who have done groundbreaking work in the area of music psychology and perceptually-oriented music theory over the last few decades – specific works by Albert Bregman, Christopher Hasty, David Huron, Ray Jackendoff, Carol Krumhansl, Fred Lerdahl, and David Lewin were particularly energizing and inspirational for me and I cannot imagine working in this field without these invaluable precedents. The influence of some of these figures on the current study will be rather obvious to an informed reader, and there are, of course, many other individuals who have contributed work that has proven essential to what follows.

Finally I must thank my wonderful wife, Melissa Smey, for her love and support during this rather long and occasionally difficult process. I could not have completed this work without her, and I hope that it lays the foundation for a long and fruitful partnership in the field of music.
Contents

Abstract ................................................................................................................................................ iv
Acknowledgements .............................................................................................................................. vi
Multimedia and Computer Programs ................................................................................................. ix
List of Examples ..................................................................................................................................... x
I. Introduction and Overview ............................................................................................................. 1
  General principles for a perceptual model of tonality ................................................................. 2
  Recent theories of the tonal hierarchy ......................................................................................... 5
    Carol Krumhansl’s *Cognitive Foundations of Musical Pitch* .............................................. 7
    Fred Lerdahl’s *Tonal Pitch Space* ......................................................................................... 17
    Other suggestive precedents ..................................................................................................... 27
  The overall approach .................................................................................................................... 28
II. Vertical Relations, Part One: Consonance and Dissonance ....................................................... 32
  The overtone series ..................................................................................................................... 32
  The overtone series and tonality – historical precedents ............................................................. 34
  Consonance and dissonance ......................................................................................................... 36
  Is the triad the "most consonant" sonority? ................................................................................. 44
  Relative consonance as tonic-making ....................................................................................... 48
  The problem of defining consonance ....................................................................................... 49
III. Vertical Relations, Part Two: Intervallic Rootedness ................................................................. 58
  The residue pitch ......................................................................................................................... 59
  A precedent: Hindemith’s *Craft of Musical Composition* ..................................................... 63
  First thesis: Applying Terhardt’s pitch-finding algorithm to rooted intervals ....................... 65
  Second thesis: Rootedness as "next-best" fit ............................................................................. 69
  Third thesis: Rootedness in pitch-class space .......................................................................... 71
  The acoustic root in context ......................................................................................................... 89
Non-tonic harmonies as a shift in perspective ....................................................... 95
The global significance of tonic tones ................................................................. 98
Can any sonority serve as the center of the tonal hierarchy? ......................... 100

IV. The Underlying Scale .................................................................................................. 102
   Annabel Cohen’s Well-Tempered Clavier experiment ......................................... 104
   Pitch-location memory: The trace ................................................................. 105
   Scales consist of steps ...................................................................................... 113
   The consecutive semitone constraint ........................................................... 119
   The "deep scale" property ............................................................................... 122
   The diatonic collection and consonance ....................................................... 124
   Non-traditional scales and scalar perception: The Ondine experiment .......... 127
   The scalar vs. chromatic distinction and musical experience ...................... 136

V. Simple Diachronic Relations ................................................................................... 155
   The hierarchical structure of a single tone .................................................... 155
   Two events ....................................................................................................... 160
   Beginners’ advantage ...................................................................................... 163
   Arpeggiations ................................................................................................. 164
   The power of voice leading and the finality effect .......................................... 167

VI. The Dyad Plus Monad Experiment .................................................................... 176
   Designing the experiment ............................................................................. 181
   Presenting the data ........................................................................................ 189
   Findings ......................................................................................................... 190
      Stepwise motion vs. arpeggiations ............................................................... 190
      Voice-leading and the “finality effect” ....................................................... 201
      Intervallic rootedness ............................................................................... 204
      Rooted intervals and “winning percentage” ............................................. 211
Multimedia and Computer Programs

Multimedia examples and computer programs are available at http://davesmey.com/dissertation

Chapter 2

HutchinsonAndKnopoff.exe – A dissonance calculator for Windows PC

Chapter 3

PitchClassRootedness.exe – Calculates rootedness of a simultaneity, for Windows PC
ex3-2residuepitch.wav – Audio example of residue pitch

Chapter 4

OndineExperiment.zip – Program for Windows PC, simulates chapter 4 experiment
ex4-2frogs.wav – Audio example, simulated environmental stimulus
ex4-4ckingclip.wav – Audio example, Martin Luther King’s “I Have a Dream” speech
ex4-4dkingtranscription.wav – Audio example, interleaved pure pitch transcription
ex4-6trillthreshold.mp4 – Video demonstration, trill threshold
ex4-6trillthreshold.mp3 – Example 4.6, audio only
ex4-7fissionboundary.mp4 – Video demonstration, fission boundary
ex4-7fissionboundary.mp3 – Example 4.7, audio only

Chapter 5

ex5-1rightbranchtones.mp4 – Video example, tones create right-branching hierarchy
ex5-1rightbranchtones.mp3 – Example 5.1, audio only
ex5-2softattacks.mp4 – Video example, soft attacks create left-branching structure
ex5-2softattacks.mp3 – Example 5.2, audio only
ex5-3toneswithsubdetails.mp4 – Video example, tones with sub-details
ex5-3toneswithsubdetails.mp3 – Example 5.3, audio only

Chapter 6

List of Examples

1.1: Krumhansl and Kessler 1982 experiment design ................................................................. 8
1.2: Krumhansl and Kessler 1982 results ............................................................................. 9
1.3: 40-note string from Temperley-West Marvin experiment .............................................. 15
1.4: Lerdahl's basic pitch-class space (for C major) ................................................................. 18
1.5: Lerdahl's measure of chord distance (for I to V in C major) .......................................... 20
1.6: Chord distance for IV to I in G major ........................................................................... 21
1.7: Finding the “shortest path” within the GTTM framework ............................................. 23
2.1a: Harmonics of 110hz ...................................................................................................... 33
2.1b: Graph of harmonics on logarithmic scale ................................................................. 34
2.2: Overtone interactions in a diminished fifth ................................................................. 38
2.3: Critical band from about C2 (65.41hz) to C6 (1046.50hz) ........................................ 40
2.4: Calculation of roughness rating for \{C4, B\#4\} ............................................................. 42
2.5: Selected trichords and their Hutchinson and Knopoff ratings ...................................... 45
2.6: Roughness ratings for simultaneities between C4 and C6, organized by set-class ... 46
2.7: Roughness ratings for simultaneities between C4 and C6, organized by set-class ... 47
2.8: Minor-mode progression with relatively rough tonics ...................................................... 49
2.9: Roughness of intervals arranged symmetrically around A440 ................................... 51
2.10: Periodicity in a perfect fifth ........................................................................................... 53
2.11: Roughness of equally-tempered intervals (symmetrical around A440) .................... 55
3.1: Four rooted stimuli ......................................................................................................... 59
3.2: Design for three tones at A3 ........................................................................................ 60
3.3: Modeling a perfect fifth with masking and overtone coincidence ................................ 67
3.4: Rooted intervals as partially-fitting overtone patterns .................................................. 71
3.5: Frequency perception as a spiral, representation of A2 and its overtones ................... 73
3.6: Parncutt's root-finding template in pitch-class space ................................................... 76
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.7</td>
<td>Rootedness calculation with registrally weighted PCs</td>
<td>79</td>
</tr>
<tr>
<td>3.8</td>
<td>Sonorities tested in this study</td>
<td>80</td>
</tr>
<tr>
<td>3.9</td>
<td>Perfect fifth illustrates the function of a unisons / octaves value</td>
<td>82</td>
</tr>
<tr>
<td>3.10</td>
<td>Minor sixth versus minor thirteenth</td>
<td>83</td>
</tr>
<tr>
<td>3.11</td>
<td>A &quot;fragile&quot; first inversion minor triad</td>
<td>85</td>
</tr>
<tr>
<td>3.12</td>
<td>Adjustment for octave doubling</td>
<td>86</td>
</tr>
<tr>
<td>3.13</td>
<td>Major sixth</td>
<td>87</td>
</tr>
<tr>
<td>3.14</td>
<td>Possible continuations of stimuli</td>
<td>89</td>
</tr>
<tr>
<td>3.15</td>
<td>Hypothetical C minor passage</td>
<td>91</td>
</tr>
<tr>
<td>3.16</td>
<td>Half-cadence reinforces tonic</td>
<td>92</td>
</tr>
<tr>
<td>3.17</td>
<td>A dominant seventh and auralized tonic</td>
<td>93</td>
</tr>
<tr>
<td>3.19</td>
<td>Fauré &quot;Clair de lune&quot; mm. 5-8</td>
<td>96</td>
</tr>
<tr>
<td>3.20</td>
<td>Tonic tones emerge as structural in a French folk melody</td>
<td>99</td>
</tr>
<tr>
<td>4.1</td>
<td>Krumhansl and Kessler's results for the major and minor modes, revisited</td>
<td>102</td>
</tr>
<tr>
<td>4.2</td>
<td>Simulated environmental stimulus with croaking frogs</td>
<td>107</td>
</tr>
<tr>
<td>4.3</td>
<td>Event-heirarchical interpretation of a single croak, abstraction of the trace</td>
<td>108</td>
</tr>
<tr>
<td>4.4</td>
<td>Passage from the &quot;I Have a Dream&quot; speech with pitch transcription</td>
<td>110</td>
</tr>
<tr>
<td>4.5</td>
<td>Hypothetical melodic fragment with pitch traces and voice-leading</td>
<td>112</td>
</tr>
<tr>
<td>4.6</td>
<td>Trill threshold, frame from demonstration video</td>
<td>114</td>
</tr>
<tr>
<td>4.7a</td>
<td>The fission boundary, diagram from demonstration video</td>
<td>117</td>
</tr>
<tr>
<td>4.7b</td>
<td>Fission boundary values in different registers</td>
<td>117</td>
</tr>
<tr>
<td>4.8</td>
<td>Stepwise connection relegates middle element to chromatic status</td>
<td>121</td>
</tr>
<tr>
<td>4.9</td>
<td>A hypothetical passage with successive diatonic semitones</td>
<td>122</td>
</tr>
<tr>
<td>4.10</td>
<td>Diatonic expectations and relative consonance</td>
<td>127</td>
</tr>
<tr>
<td>4.11</td>
<td>Segmentation of Ravel's &quot;Ondine,&quot; from <em>Gaspard de la nuit</em></td>
<td>129</td>
</tr>
<tr>
<td>4.12</td>
<td>Ondine experiment graphical interface</td>
<td>131</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>4.13</td>
<td>Ondine experiment results</td>
<td>132</td>
</tr>
<tr>
<td>4.14</td>
<td>Influence of event hierarchy on scalar response</td>
<td>135</td>
</tr>
<tr>
<td>4.15</td>
<td>Mozart's String Quartet in G major, K. 387, first movement, mm. 1-4</td>
<td>138</td>
</tr>
<tr>
<td>4.16</td>
<td>Mozart K. 387 excerpt with abrupt cessation after D</td>
<td>139</td>
</tr>
<tr>
<td>4.17</td>
<td>Modeling mm. 1-2 as three simple triads</td>
<td>141</td>
</tr>
<tr>
<td>4.18</td>
<td>Measures 1-2 with some metric normalization</td>
<td>143</td>
</tr>
<tr>
<td>4.19</td>
<td>Measures 3-4, truncated, with metric normalization</td>
<td>145</td>
</tr>
<tr>
<td>4.20</td>
<td>Measures 3-4, truncated at A#</td>
<td>146</td>
</tr>
<tr>
<td>4.21</td>
<td>Schubert's &quot;Heidenröslein,&quot; mm. 1-12</td>
<td>148</td>
</tr>
<tr>
<td>4.22</td>
<td>The introduction of C#, hypermetric context</td>
<td>149</td>
</tr>
<tr>
<td>4.23</td>
<td>Schubert's &quot;Ständchen,&quot; end of first stanza (mm. 17-28)</td>
<td>151</td>
</tr>
<tr>
<td>4.24</td>
<td>&quot;Ständchen&quot;, postlude to first stanza (mm. 27-36)</td>
<td>153</td>
</tr>
<tr>
<td>5.1</td>
<td>Tones create right-branching hierarchy</td>
<td>156</td>
</tr>
<tr>
<td>5.2</td>
<td>Soft attacks create left-branching structure</td>
<td>158</td>
</tr>
<tr>
<td>5.3</td>
<td>Tones with sub-details</td>
<td>159</td>
</tr>
<tr>
<td>5.4</td>
<td>Right- and left-branching structures</td>
<td>161</td>
</tr>
<tr>
<td>5.5</td>
<td>Metric and tonal interpretations tend to align</td>
<td>162</td>
</tr>
<tr>
<td>5.6</td>
<td>Right-branching hierarchy as default interpretation</td>
<td>163</td>
</tr>
<tr>
<td>5.7</td>
<td>Arpeggiation tends to be right-branching</td>
<td>165</td>
</tr>
<tr>
<td>5.8</td>
<td>Arpeggiations of rooted intervals</td>
<td>166</td>
</tr>
<tr>
<td>5.9</td>
<td>Pitch-space voice-leading as a rotation of axes</td>
<td>168</td>
</tr>
<tr>
<td>5.10</td>
<td>Voice-leading as structurally neutral motion</td>
<td>169</td>
</tr>
<tr>
<td>5.11</td>
<td>Lingering trace creates agogic accent, finality effect</td>
<td>170</td>
</tr>
<tr>
<td>5.12</td>
<td>All stepwise relationships and their hierarchical effect</td>
<td>171</td>
</tr>
<tr>
<td>5.13</td>
<td>Repeated stepwise motion is neutral</td>
<td>172</td>
</tr>
<tr>
<td>5.14</td>
<td>Bharucha’s figures and the accumulation of traces</td>
<td>173</td>
</tr>
<tr>
<td>6.1</td>
<td>J. S. Bach’s Partita for Solo Violin in B minor, BWV 1002, Presto</td>
<td>177</td>
</tr>
</tbody>
</table>
6.2: Original conception of the experiment ................................................................. 178
6.3: Interpretations of the dyad + monad stimuli ......................................................... 179
6.4: Generating the set of 275 stimuli ........................................................................... 181
6.5: Checkboxes represent a figure with implied tone, sustain, and voice-leading
connection ..................................................................................................................... 183
6.6: Chromatic pitch-selector buttons .......................................................................... 184
6.7: Two-event tree interface ....................................................................................... 186
6.8: Extending the two-event trees to account for three events ................................. 187
6.9: Three-event trees that disrupt the two-event structure .......................................... 188
6.10: Presenting a specific hearing of a stimulus ......................................................... 190
6.11: Figures that evoked pitch-class voice-leading connections .............................. 192
6.12: Frequency of marked voice-leading connections .............................................. 195
6.13: Summary of voice-leading in arpeggiations and progressions .......................... 197
6.14: Some progressions that lack voice-leading intervals (ICs 1 and 2) ................... 199
6.15: Voice-leading interval collisions ........................................................................ 200
6.16: Intervallic winners and losers ............................................................................. 202
6.17: Winning percentage for the various interval classes .......................................... 203
6.18: Selected figures influenced by the “finality effect” ........................................... 204
6.19a: Sample figure with three interval / tonic combinations ................................. 206
6.19b: Method of calculation ....................................................................................... 207
6.20: Roots chosen with the various intervals ............................................................... 207
6.21: Finality effect vs. rooted interval effect .............................................................. 212
6.22: Intervallic rootedness precedes event hierarchy ................................................. 213
6.23: Winning percentages for rooted intervals ......................................................... 215
6.24: Various effects of interval 5 ................................................................................ 216
6.25: Interval 4 in second position fails to establish root ............................................. 217
6.26: Intervals associated with various roots, separated by position and event-hierarchical
status  ......................................................................................................................  218

6.27:  Minor third evokes missing root .................................................................  224

6.28:  Some figures with non-tonic resolutions .......................................................  225

6.29:  Intervals and non-tonic interpretations ..........................................................  228

6.30:  The predictive model .....................................................................................  233

6.31:  Arpeggiation and progression percentages for stimuli with voice-leading intervals  234

6.32:  Arpeggiation and progression in non-VL figures ...........................................  235

6.33:  Ranked interval-root associations for arpeggiations .......................................  236

6.34:  Root predictions for a major-triad figure .......................................................  237

6.35:  Top 3 results for Stimulus No. 70 .................................................................  238

6.36:  Interval-root lists for progressions ...............................................................  240

6.37:  Root predictions for Figure No. 91 ...............................................................  242

6.38:  Branching predictions for the stacked-fourth figure .....................................  243

6.39:  Top 3 results for the stacked-fourth figure ....................................................  244

6.40:  Actual results .................................................................................................  244
Chapter 1

Introduction and Overview

The experience of most music involves a sense of tonality or key, the impression that, at any given moment, all of the tones being sounded relate back to one central or hierarchically most important pitch level. This idea has been central to discourse about music since the 18th century. One might expect, then, that the question of how listeners determine and experience these key centers would be more-or-less settled amongst music theorists.

In the most basic levels of musical training we have analytic practices that take tonality as a given. The musical surface is parsed in terms of harmonies and non-harmonic elaborations, and examining the way these harmonies interrelate (or "function") can determine what key is in effect during any particular passage. In the music of the 17th to 19th centuries this determination is rarely difficult, thanks to compositional norms that seem to be generally well-understood.

However, when one encounters the music of composers who either precede the era of Common Practice or who sought to modify or even refute these conventions, questions begin to arise. How tonal is a Josquin mass, or the Stravinsky Octet? How does a listener perceive these works? If we simply apply our traditional battery of analytic symbols and techniques to these new objects, what will that explain? We cannot answer these questions without asking, simply, "What is tonality?"

Many in the field of music psychology would answer this question in terms of the tonal hierarchy, the subjective organization of all sounding tones into a well-defined and stable
scheme of relative structural weight. Experiments have been conducted to observe such perceptions in a variety of subjects, and theoretical works have sought to comprehensively map these relations in terms of scale-degrees, harmonies, and keys. However, it is my contention that the existing literature on tonal hierarchy has not yet achieved a satisfactory account of the phenomenon that is tonality. We still don't know how the tonal hierarchy is elicited, how it shapes and is shaped by the moment-to-moment experience of music, and what we as listeners do in order to perceive it. Although the perception of the tonal hierarchy is sometimes implied to be the product a dedicated cognitive apparatus, the origins, structure and functioning of such a mental facility have not been worked out.

The current project seeks to isolate and examine the component parts of the tonal hierarchy and reassemble them into a model that is more detailed and more plausibly authentic than those that are currently available. It is hoped that a better account of the tonal hierarchy will go a long way towards understanding the experience of music that is organized around a key.

**General principles for a perceptual model of tonality**

I think that a successful account of how listeners experience tonality would have several essential qualities. It would describe events from a diachronic perspective, relate to processes of everyday hearing, and include the shaping influence of attention and top-down processing.
Synchronic perception vs. diachronic

Western “Classical” music usually depends on the production, perusal, and performance of notated scores. The use of a score can give an analyst a God-like perspective of the piece, offering a synchronic view of every note that will ever sound in a performance of the work. This explicitness and out-of-time random access to the score can be quite powerful, allowing one to follow one’s analytic train of thought at will, comparing any moment to any other in search of large-scale properties. It makes discourse about music corrigible, as one can test assertions about the piece against the score and be certain that all relevant facts have been accounted for.¹ The use of a score or transcription allows one to develop and communicate sophisticated ideas about a piece that might be difficult or even impossible to formulate through listening alone. It is not surprising, then, that most contemporary analysis avails itself of score-knowledge to deliver a sweeping yet detailed account of a work.

The actual perception of a piece, on the other hand, occurs in a relatively impoverished epistemological environment. Events happen one after another. Simultaneous events must be parsed and separated. Previous events are gone, accessible only through memory, and forthcoming events are unknown or, at best, anticipated. Even the existence of individual “notes” is not a given, but must be extracted and reconstructed from the sound signal. An accurate model of musical perception must begin in this world and actively resist slipping into the omniscient perspective of the score reader.

Many contemporary perceptual theories are themselves insufficiently diachronic in orientation. Creating a moment-to-moment account is not only desirable for the purposes of

authenticity (in that we can be more confident that our model reflects the way people experience music), but it can also lead to new descriptions of the music itself, with analyses that pinpoint specific moments in time and describe what is happening with precision.

Everyday hearing

A plausibly authentic model of musical perception should also be rooted in processes of everyday hearing. The more an account relies on structures and procedures that are wholly specific to the domain of music, the more it raises the question of "why" - why would humans evolve such a complex apparatus that has no known utility for survival?2 It makes more sense to view the performance and perception of music as a repurposing of cognitive facilities that serve other functions, such as understanding our environment with our ears (what Albert Bregman calls “auditory scene analysis”3) or communicating via spoken language. Thus, we should resist explanations that are too “music specific.”

Attention

One facet of experience that is relevant in all domains is that of attention. Whether we are dealing with hearing, vision, smell, touch or taste, we have the ability to focus on different aspects of a stimulus and get varying results. The senses are, of course, also open to peripheral information that is not actively being focused on - otherwise, nothing could ever “come to” our attention. Studies have even suggested that we are susceptible to subliminal perceptions that do


not engage conscious awareness at all. However, attention and the “top-down” expectations that drive it are a constant influence on perception. “Inattentional blindness” describes situations in which a distracted viewer fails to register objects that are plainly accessible in his or her field of vision. Experimenters who have studied IB have gone so far as to make the somewhat radical claim that “there is no perception without attention.” Attention is a particularly crucial element in the experience of the complex and potentially ambiguous stimulus that is music. Any account that does not consider its influence is likely to be incomplete.

As I will argue below, I believe that current theories of the tonal hierarchy lack these highly desirable qualities.

Recent theories of the tonal hierarchy

In the experience of a passage of tonal music, a single pitch class tends to emerge as central or referential. It seems to be the most stable or restful tone, and other notes seem to be heard in relation to it. This is the tonic note after which a key is named (i.e. the D of D Major), and we can say that it is at the “top” or “center” of a hierarchy that also involves all other notes. Other tones of the tonic triad (scale-degrees 3 and 5) are similarly important in the hierarchy but are not as restful and definitive as the tonic tone itself – thus they constitute a second layer that is somehow dependent on the tonic. On the next tier of importance would be other tones within the

---


underlying scalar collection that are not in the tonic triad. Finally, the most “remote” pitches would be the chromatic tones, those that do not belong to the prevailing scalar collection being used. This constellation of relationships is known as the tonal hierarchy, and as long as a sounding passage does not modulate to a new key these general categories of relative stability or centrality will remain fairly consistent.

Some influential music psychologists and perceptually-oriented music theorists think that the tonal hierarchy is the result of a specific cognitive apparatus that can orient itself around a particular pitch-class and subsequently parse all sounding tones. Thus, the hierarchy is not merely a passive effect of listening to tonal music but also a cause of our interpretation of it – it is the engine that makes contemporary theories of tonal perception work. The most prominent and thorough models of the tonal hierarchy to date have been advanced by Carol Krumhansl and Fred Lerdahl.6

Krumhansl’s *Cognitive Foundations of Musical Pitch*

Carol Krumhansl has conducted a series of psychological experiments designed to show the experiential reality of the tonal hierarchy. These studies focus not only on the interrelations of individual tones as described above, but also on harmonies and modulation between keys.

A 1982 study co-authored with Edward Kessler is typical of her approach.⁷ 10 subjects with moderate-to-high musical experience and little theory training were presented with stimuli meant to evoke either a major or minor key center. Each stimulus was followed by a randomized “probe tone.” (This configuration is summarized in example 1.1. Krumhansl and Kessler's stimuli and probes were made with "Shepard tones," electronic sounds which contain octave-related sinewaves in all possible registers in an attempt to avoid the influence of textural top and bottom or fixed voice-leading choices. I've represented this sound design with staves filled with octaves and an arbitrary top and bottom.) After each stimulus and probe-tone combination, subjects were asked to rate how well the probe "fit into" or "went with" the stimulus on a scale of one to seven. The results for major and minor keys were summarized in the profiles of example 1.2. Krumhansl and Kessler managed to elicit the four-level hierarchy of tonic, tonic triad tones, scalar tones, and chromatic tones described above for both major and minor keys.

---

⁷ Krumhansl and Kessler, “Tracing the Dynamic Changes…”
Example 1.1: Krumhansl and Kessler 1982 experiment design

stimulus + probe tone

response: goodness-of-fit rating

1 2 3 4 5 6 7

Stimuli

(Includes only those used in final results. Others were tested but discarded. Test figures were transposed to various pitch levels - here they are arbitrarily represented in C major and A minor.)

major-key stimuli

\[
\text{IV} \quad \text{V} \quad \text{I}
\]

\[
\text{ii} \quad \text{V} \quad \text{I}
\]

\[
\text{vi} \quad \text{V} \quad \text{I}
\]

minor-key stimuli

\[
\text{iv} \quad \text{V} \quad \text{i}
\]

\[
\text{ii}^\# \quad \text{V} \quad \text{i}
\]

\[
\text{VI} \quad \text{V} \quad \text{I}
\]
Example 1.2: Krumhansl and Kessler 1982 results

Major Key Profile

tonic

tonic triad tones

scalar tones

chromatic tones

C C#/Db D D#/Eb E F F#/Gb G G#/Ab A A#/Bb B
Having achieved consistent results that suggested the existence of an internal sense of key, Krumhansl then sought to account for how this hierarchy is developed and engaged. She surveyed studies that counted the relative frequency of appearance of the twelve pitch-classes in various small collections of tonal works.\(^8\) (This task can be accomplished either by simply counting notes or by adding up the durations of notes, so that longer tones count more. Some of the studies analyzed only one melodic part and ignored the accompaniment, while others counted tones within the entire texture.) She found that the relative preponderance of each pitch-class corresponded to its ranking in the tonal hierarchy experiment -- the tonic was the most common pitch-class, followed by scale-degrees three and five, the other scalar notes, and finally the chromatic notes. On the strength of this correlation, she hypothesized that the sense of tonality is essentially a prediction of the probability of occurrence of any particular tone, which is

developed through a lifetime of exposure to tonal music. The tonal hierarchy orients itself as soon as a stimulus shows a propensity to favor pitches around a certain tonic. If another pitch-level becomes locally prominent it can disrupt and reorient the hierarchy. Krumhansl thinks that these probability judgments eventually solidify into “an internal representation specifying various degrees of tonal stability.”

Despite the successful correlations in Krumhansl’s data and her project's compatibility with the overall goals of tonal theory, some potential flaws must be acknowledged. There are two broad issues that one might consider – whether the probe-tone profiles are a good measure of one’s sense of key, and whether listeners actually determine key based on the relative quantity of tones in a passage.

One problem with the probe-tone technique is that it appears to elicit the tonal hierarchy only from trained musicians. An early experiment did not achieve robust results from the untrained portion of the subject pool.10 Here the experimental design presented an ascending or descending seven-note scale followed by a randomized probe tone. Subjects were asked to rate how well the randomized tone "completed" the musical figure. While those with the most musical experience produced results similar to example 1.2, rating tones from the tonic triad highly and so forth, inexperienced listeners tended to rate the probe tones according to their pitch-space proximity to the final tones of the scalar stimulus. If the probe tone made a relatively small interval with the final scale tones, it received a higher rating, regardless of tonal identity. Thus, C♯ was judged to be a better completion of the ascending C-major figure than G. While these results could be explained as a reasonable response to a different experimental design

---

9 *Cognitive Foundations of Musical Pitch*, 76.

10 Krumhansl and Shepard, “Quantification of the Hierarchy of Tonal Functions within a Diatonic Context.”
(including a different verbal cue), it is worth noting that Krumhansl and her collaborators tended to avoid using non-musicians in subsequent experiments, including the 1982 experiment that produced the chart of example 1.2.11

Richard Parncutt and Albert Bregman revisited this issue by conducting a probe-tone experiment in 1998 with equally-sized groups of musicians and non-musicians, and found that the non-musician group failed to make the distinctions that are central to the tonal hierarchy model.12 An initial round of trials collected probe-tone profiles for single triads and dyads, again composed entirely out of Shepard tones. Probe tones were restricted to members of a diatonic collection, so the most remote layer of the hierarchy (the chromatic) was not tested. The verbal prompt was somewhat different that Krumhansl and Kessler’s, asking subjects to rate how “similar” the probe was to the stimulus. Untrained subjects generally failed to rate chord and dyad members significantly higher than non-members, and failed to rate the triadic or intervallic roots significantly higher than non-roots. The authors suggest that the task might be “simply too difficult” for a layperson to execute. Untrained subjects would presumably still be exposed to the statistical regularities of tonal music throughout their lifetimes, but there seems to be something about the experience of actually making music that sensitizes one to the tonal hierarchy to the extent that it can be measured by the probe-tone format.13


12 “Tone Profiles Following Short Chord Progressions: Top-Down or Bottom-Up?” *Music Perception* 18 (2000), 25-57. The musician and non-musician groups each had fifteen members.

13 It is possible that a different experimental design might be more effective at capturing musical intuitions about the tonal hierarchy from untrained subjects. Carol Krumhansl and Frank C. Keil’s “Acquisition of the Hierarchy of Tonal Functions in Music,” *Memory and Cognition* 10 (1982): 243-51 successfully demonstrated that very young
David Butler has pointed out that the probe-tone experiments may not be measuring a robust, internalized structure, instead reflecting a more passive response to relatively superficial aspects of the materials being presented. The stimuli that are ultimately included in the 1982 results, after all, all conclude with the tonic triad. Thus, high rankings for the tonic tones might simply reflect the immediacy of them having been heard most recently. (Indeed, I'll refer to this phenomenon in my Chapter 5 as the “finality effect,” and argue that it is, in fact, a strong factor in tonal orientation.) In addition, the Krumhansl-Kessler cadential stimuli tend to cover all scalar tones (save for the "subtonic" in minor keys, i.e. B♭ in the key of C minor), possibly making the distinction between scalar tones and chromatic notes simply one of short-term familiarity.

The measurement of the statistical distribution of pitch classes in musical works has become quite popular within the field of music psychology and perceptually-oriented music theory. Dozens of authors have proposed modifications to Krumhansl’s original method of correlating the aggregate proportions of pitch classes with perceived key. One of the issues that has been discussed is whether the technique should sample the content of wide swaths of music, possibly taking in complete pieces, or be restricted to narrower segments. Krumhansl’s original study proposed a simple segmentation of a work into individual measures to view the effect of tonicization and modulation – others have experimented with a temporal window of decay that favors recently sounded tones and gradually deprecates tones that lie further in the past. Some authors have noted that it makes more sense to correlate the commonness of pitches in a specific subjects prefer diatonic melodies to chromatic ones, and that older listeners make increasingly fine tonal distinctions.


work to actual distributional data derived from a large body of works—these distributional profiles are subtly different from the probe-tone ratings derived from Krumhansl and Kessler’s experimental trials. Some have also developed mathematical methods of correlation that are different from Krumhansl’s. Thus, while Krumhansl’s original work achieved a somewhat limited accuracy in making key judgments, subsequent studies have reported progressively better results. A recent corpus study boasted a 91% success rate across a collection of 492 works.

The question remains, however, whether listeners actually use the aggregate distribution of tones to determine the key of a work. As is apparent in the Krumhansl-Kessler experiment, a minimal cue such as a single triad is sufficient to invoke the major-minor key profiles. A proponent of the pitch-class distribution model might note that such a figure already has a strong correlation with one of the 24 key profiles (namely .83 for a major triad with its corresponding major key and .87 for a minor triad), but surely this is a different phenomenon from a preponderance of tones over time. An alternate hypothesis (which I personally favor) would be that there is something about the acoustic properties of the triad itself that induces this key center. As we will see in my chapter 6 experiment, certain cues can even evoke tonics that are not sounded in the stimulus, creating an orientation that would not be predictable with the key-

---


18 Albrecht and Shanahan, “The Use of Large Corpora…” Applying the algorithm to entire pieces achieved 91% accuracy, but arbitrarily restricting statistical samples to the first and last eight measures of each work boosted performance to 93%.
profile correlation technique. Thus, it seems that a statistical distribution of tones that favors a tonic triad is not necessary to actually evoke a key.

In addition, an experiment conducted by David Temperley and Elizabeth West Marvin concluded that the relative frequency of tones is not always sufficient to establish a key.\textsuperscript{19} Temperley and West Marvin stochastically generated 40-note melodic strings which conformed to the pitch-class proportions of various key profiles but otherwise contained no intentional structuring. Example 1.3 reproduces one of these figures, which has a pitch distribution that is close to the $E_b$-minor key profile. (The authors use key profiles derived from the Essen Folksong Collection, which are slightly different from Krumhansl’s profiles. This ideal distribution is graphed in the lower half of the example.) Music students asked to identify the tonic and mode of these passages showed very low intersubjective agreement, and favored “wrong” keys about 48\% of the time. The melody in example 1.3 is one of these problematic strings – although it most closely matches the $E_b$-minor profile it evoked $B_b$ major, B major, and $B_b$ minor more often than $E_b$ minor.

\textbf{Example 1.3: 40-note string from Temperley-West Marvin experiment}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{example13.png}
\end{figure}

\textsuperscript{19} “Pitch-Class Distribution and the Identification of Key,” \textit{Music Perception} 25 (2008), 193-212.
Since the overall 52% success rate is much higher than chance, Temperley and West Marvin conclude that “pitch-class distribution is clearly one component of key identification.” However, there is an alternate explanation that could potentially rule out the role of pitch-class distribution entirely – that the statistical regularities observed in tonal music are the mere by-product of other specific processes that are necessary to maintain a sense of key.

The statistical approach to tonal analysis is anathema to many traditional music theorists because it essentially observes unordered aggregates of tones. In the experience of performance, analysis, and composition the ordering of tones often seems critical in the creation of certain effects. It would seem to matter greatly if, say, a passage were arbitrarily reversed or if its tones were scrambled around. Subsequent chapters in the current study will argue that tonality is indeed created by a handful of properties that arise from specific temporal configurations of tones. However, if the sense of key is wholly contingent on fine details, this might also mean that it is less robust and more “fragile” than one might like to imagine. Passages with the “wrong” details (say, an inordinate emphasis on a remote melodic tone or harmony, or the improper resolution of an embellishment) could easily destroy the established tonal center and create an unintentionally ambiguous or complex effect. The statistical regularity of tonal music
can thus be reimagined as a reflection of these constraints, a byproduct of the actual tonal processes at work.

In a recent keynote address to the Society for Music Perception and Cognition, Carol Krumhansl discussed the relationship between statistics, structure, and style analysis, at one point quoting Leonard Meyer:

> It may be possible to distinguish Wagner's operas from Verdi's in terms of the relative frequency of deceptive cadences. But unless one can explain how the frequency of deceptive cadences is related to other features of Wagner's style the trait will be no more than a marker.\(^{20}\)

Meyer’s notion of a “marker,” a piece of information that can effectively categorize a work but remains otherwise superficial is perhaps instructive in evaluating the distributional analysis of tonality. It seems possible that the key profiles are a mere marker of more essential properties, elements which the current study is intended to illuminate.

**Fred Lerdahl's Tonal Pitch Space**

Fred Lerdahl's work with linguist Ray Jackendoff has provided what is perhaps the most comprehensive theory of music perception of the last few decades. Their *Generative Theory of Tonal Music* (GTTM) offers a framework that accounts for the organization of heard music into a coherent and syntactically rich internal representation which is somewhat analogous to the

---

The theory is built out of the complex interplay of several perceptual domains which are carefully accounted for in terms of “preference rules” which assign a particular structure to a given stimulus. However, while the influence of tonality is an important component in the GTTM model, the authors declined to account for it in rigorous terms, stating only that they took traditional notions of harmonic organization as given.22

Lerdahl's *Tonal Pitch Space* (TPS) was an attempt to rectify this omission.23 It presents a framework that can assign a tonal interpretation to any stimulus that is based on triads. Its simplest level looks much like Krumhansl's probe tone ratings, as pitches within the key are arranged into hierarchical tiers (example 1.4). The only real deviation from Krumhansl's results is that the fifth scale-degree is decisively elevated to a higher rank than the third - in the experimental studies this configuration was only obtained with major-key stimuli but was reversed in minor keys.

**Example 1.4: Lerdahl's basic pitch-class space (for C major)**

<table>
<thead>
<tr>
<th>tonic</th>
<th>C</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>fifth</td>
<td>C</td>
<td>G</td>
</tr>
<tr>
<td>tonic triad</td>
<td>C</td>
<td>E</td>
</tr>
<tr>
<td>diatonic</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>chromatic</td>
<td>C</td>
<td>C# / D♭</td>
</tr>
</tbody>
</table>


22 GTTM, 117.

Lerdahl uses this basic pitch-class hierarchy to model interchordal and interkey relationships. The primary property ascribed to harmonies and keys is “distance.” Psychological distance is a familiar concept in cognitive theory and a basic element in the metaphor of a space. Entities that are “closer” together are more similar or more directly related. The mental effort needed to perceive this relationship can be relatively small. More complex relationships are “distant,” and distant entities could also be thought of as more dissimilar or distinct.

Lerdahl assumes, for the most part, that tonal passages are parsed into familiar triads, seventh chords, and other conventional harmonic entities. The challenge of assigning a key or keys to a passage involves determining the best fit for these harmonies. His model functions by following "the principle of the shortest path," selecting the interpretation that involves the smallest attributed distance from chord to chord. This key-selection algorithm functions within the rigorously structured grouping-and-meter grid of GTTM, and its determinations go on to influence the event hierarchy, the sense that certain events elaborate other events in a constant structural ebb and flow.

Let us look more closely at how Lerdahl calculates the distance between chords within a key. His chord distance measure adds together two factors. The first is a measure of root distance on a circle of fifths. Harmonies related by fourth or fifth would score a 1, those related by whole step a 2 (since they are two fifths away), and mediant relations score a 3 or 4. A second measure evaluates the pitch-class commonality of the two chords, weighted according to the hierarchical tiers of the basic pitch-class space (example 1.4). Each chord is assumed to have

---

24 A third element $i$ accounts for distance between underlying scalar collections, which allows the theory to account for modulations and other chromatic phenomena. For the sake of simplicity I have omitted this variable from the following discussion. In all examples $i$ would equal zero.
its own hierarchy of root, fifth, and implied scalar collection tones. If highly weighted tones in
the first chord overlap with highly weighted tones of the second, the calculated distance is low.
Example 1.5 demonstrates the procedure for finding the degree of commonality between the C
major and G major triads, a putative I and V in C. As we evaluate the overlap of the two triads,
the common-tone of 5 provides only 1 point worth of divergence, because it is the highest ranked
tone in the G major triad and the second-most ranked tone of C major. However, the other tones
of the V are mere scalar tones in C, and thus the distance score is increased by 3 for a total of 5.

Example 1.5: Lerdahl's measure of chord distance (for I to V in C major)

Lerdahl’s distance rule (intrakey) is \( \delta(x \rightarrow y) = j + k \), where

\[ j \] is distance on a diatonic circle of fifths (in this case, 1), and

\[ \begin{array}{c}
I (C) \\
\end{array} \xrightarrow{1} \begin{array}{c}
V (G)
\end{array} \]

\[ k \] is the degree of uncommonality between the two chords’ hierarchies
(in this case, 4).

hierarchy for C:I

(cont.)
It is worth noting at this point that the model would also consider an interpretation of the same triads as IV and I in G major to be equally plausible, because the model also calculates a distance of 5 between IV and I (example 1.6).

**Example 1.6: Chord distance for IV to I in G major**

\[ j \text{ is distance on a diatonic circle of fifths (in this case, 1)} \]

\[
\begin{align*}
\text{IV (C)} & \quad 1 \\
\text{I (G)} &
\end{align*}
\]

(cont.)
Example 1.7 shows how the principle of the shortest path is considered within the GTTM framework. Here the slightly simplified first phrase of the Bach chorale “Christus, der ist mein Leben” is segmented into time-spans according to the Lerdahl and Jackendoff grouping and metric preference rules. Within each segment the model seeks the closest possible relationship between harmonies. This can create ambiguity in some places, as we cannot rule out two equidistant possibilities such as V-I and I-IV. I list these alternate interpretations below each time-span in example 1.7. The ascribed distance of each pairing is noted in parentheses. As the time-spans recursively combine into larger segments, however, this ambiguity is ironed out in

\[ k \] is the degree of uncommonality between the two chords’ hierarchies (in this case, 4).

\[
\begin{align*}
G & \quad G \\
G & \quad D \\
G & \quad B & D & G \\
G & \quad A & B & C & D & E & F & G \\
G & \quad G/G & A & A/B & B & C & C/D & D & E & E & F & F & G
\end{align*}
\]

**hierarchy for G:1**

\[
\begin{align*}
G & \quad C \\
G & \quad C \\
G & \quad C \\
G & \quad C \\
G & \quad G & \quad E & \quad F & \quad G \\
G & \quad G & \quad A & \quad A/B & \quad B & \quad C & \quad C/D & \quad D & \quad E & \quad E & \quad F & \quad F & \quad G
\end{align*}
\]

**hierarchy for G:IV**

3 points for the new root

1 point for the new third

---

25 after TPS, 75.
favor of a globally tenable analysis in F major. In addition to this bottom-up process Lerdahl posits a simultaneous top-down sense of prolongational structure that is sensitive to issues of large-scale coherence, which would presumably help to reduce this ambiguity in real time.

Example 1.7: Finding the “shortest path” within the GTTM framework

Thus, Lerdahl’s model of tonality interacts with many aspects of musical structure that a simple statistical dragnet does not. It is sensitive to phrase boundaries, meter, rhythmic effects such as the agogic accent, the creation of long-range linear connections, and the importance of cadences. However, this complexity makes it difficult, if not impossible, to automate the model with a computer program and create a head-to-head comparison with the results of distributional
analyses. Since the many decisions involved in a Lerdahl-and-Jackendoff analysis require the intuitive balancing of preference rules, the results are inescapably subjective and lack easy corrigibility. This absence of quantification and automaticity have led some proponents of tonal distributional models to discount Lerdahl’s work as an alternate theory. David Temperley, for example, has asserted that “there is, as yet, no proposal as to how…ordering cues might be integrated into a more general model of key identification.” By considering tonal relations within the meter-and-grouping grid of GTTM, the Lerdahl theory is indeed sensitive to the order of events.

The lack of a complete quantification of the GTTM theory may not be an actual problem – it may well be true than an authentic theory of tonal perception must leave the interaction of its various components open to the subjective influence of attention and top-down expectations. However, the model does lack some of the highly desirable qualities I outlined at the outset of this discussion – it is insufficiently diachronic in orientation, it is not grounded in everyday hearing processes, and (despite the reliance on an intuitive balancing of rules) it does not explicitly consider the influence of attention. As a result, the theory may appear to some as excessively rigid or inauthentic. It is my view that the essential content of GTTM and TPS is, in fact, largely correct and adequate, but that certain stylistic and organizational modifications may serve to enhance it.

From the outset the co-authors of GTTM concede that their model is not an account of real-time tonal perception, but rather a means of predicting the “final state” of what is perceived. Elsewhere Ray Jackendoff has made a useful distinction between an “information-

---

27 GTTM, 4.
processing” approach to cognitive theorizing and a “structural” approach. In support of the latter, he asserts that “…a theory of processing, when it gets into any detail at all, must come to grips with the question, What is the form of the information that is being processed, stored, and retrieved?” \(^{28}\) In general the immediate goal of GTTM and TPS appears to be to predict what the perceptual system will do, not necessarily to describe how it does it.

One relatively easy way to make the theory more dynamic is to consider how it would interpret a stimulus as it hypothetically freezes at specific moments in time. Such speculative glimpses of interpretation-in-progress can include the influence of projective expectations as part of the information to be considered. The current study does this, occasionally, without much theoretical overhead.

On a larger scale, however, it appears that the division of the GTTM method into discrete phases which analyze grouping, meter, and event hierarchy in turn is a more formidable barrier to arriving at a truly dynamic theory. Each phase does have a built-in feedback loop that allows a subsequent analytic layer to revise a previous one – Grouping Preference Rule 7 allows the grouping analysis to be modified in order to arrive at a better event-hierarchical interpretation, and Metric Preference Rule 9 does the same for meter. \(^{29}\) The authors believe that the sequence of phases reflects a real cognitive ordering, \(^{30}\) but Ray Jackendoff has also suggested that such a sequence, once initiated, creates a cascading effect of top-down interactions that mitigates the strict bottom-up flow of information. \(^{31}\) Nevertheless, the GTTM theory currently encourages an

\(^{28}\) *Consciousness and the Computational Mind*, 38.

\(^{29}\) GTTM 346, 348.

\(^{30}\) TPS, 4. See also *Consciousness and the Computational Mind*, 239-245.

\(^{31}\) *Consciousness and the Computational Mind*, 100-103.
analyst to assay an entire passage of music in each phase, balancing local details with more global considerations such as symmetry and parallelism before proceeding on to the next stage of analysis. This strictness, which seems designed to arrive at the best end result, tends to obscure any sense of how these different aspects of perception might interact as a passage unfolds. An essay by Jackendoff which sought to explicitly address the dynamic processing of music largely failed to discuss the interaction between the phases, preferring to focus on the dynamic unfolding of grouping, meter and event hierarchy as if they occur in discrete dimensions.\textsuperscript{32} It is my contention that a truly dynamic version of the Lerdahl and Jackendoff theory would require a radical reorganization of its various components into something that is more explicitly integrated and interactive, a task which is beyond the scope of the current project.

As I noted above, the complexity of the theory and the flexibility in the application of preference rules allow for the element of attention and top-down subjective expectations to enter in to any GTTM-style analysis. Thus, with a fairly simple change in emphasis it is possible to use the theory to explore the influence of these crucial factors in tonal perception. The current study occasionally examines the consequences of listener expectations using the tools of GTTM.

Finally, and most importantly, the theory is too “music-specific.” In positing a sophisticated, innate cognitive apparatus with discrete and fully coherent structures for grouping, meter, tonal hierarchy and event hierarchy, perhaps the most central question is not “what” (i.e. what does the perceptual facility do) or “how” but, rather, “why.” Why would humans develop such a complex system for understanding an activity that has no known purpose? Lerdahl has

examined ways in which his model parallels the perception of language, but the individual elements of the theory are generally not grounded in processes of everyday hearing.

The current study is an attempt to lay such a foundation. As it proceeds it will make use of many elements from GTTM, but it will not take the tonal hierarchy as a given, self-evident phenomenon. By rebuilding the tonal hierarchy from simple perceptual processes I think we can eventually construct a theory that feels more authentic and dynamic, thereby accounting for the way that the general consistencies of tonal music arise from specific structural properties that influence our perception on a moment-to-moment basis.

Other suggestive precedents

David Butler has used the term "intervallic rivalry" to suggest a model of tonal induction in several brief articles. In his discussions the term seems to be married to the "rare intervals hypothesis," which posits that the rarest intervals in the diatonic collection (namely the semitone and tritone) are central to the determination of key. While I have misgivings about focusing on the rareness of intervals which I will discuss in chapter 4, the overall approach of the current


The current project could indeed be described as an intervallic rivalry, as sonorities "compete" to establish a key center, and the "winners" are determined largely by intervallic content.

Jamshed Bharucha has contributed another important component to the current model in his work on "melodic anchoring."36 The influence of voice-leading and what I call the "finality effect" becomes important in my chapters 5 and 6.

A reader might also notice that the title of the current study bears a strong resemblance to an article of David Huron’s, which attempts to account for the traditional rules of counterpoint in terms of low-level psychoacoustic properties.37 This patterning is an intentional homage, as I aspire to the thoroughness and clarity of Huron’s inspirational work.

The overall approach

The current project is an attempt to supplant the idea of the tonal hierarchy as a fixed, coherent, and unified cognitive facility that imbues the incoming musical stimulus with a consistent tonal interpretation. I intend to carefully build an alternative account that consists of simple processes that can be observed in isolation. I argue that the perception of key is the result of a few primitive pitch-oriented processes in conjunction with ones that would normally be described as rhythmic, or, perhaps, "event-hierarchical." While this interaction of processes


tends to produce consistent results that can still be generally characterized as a tonal hierarchy, I do not think it is cognitively implemented as a single, self-contained apparatus.

Chapters 2 and 3 begin by examining the vertical aspect of music and the properties of simultaneous intervals in an effort to account for the “heart” of the tonal hierarchy, the tonic triad. Chapter 2 considers what is perhaps the most common explanation for the functioning of tonality, namely the phenomenon of consonance and dissonance. Ultimately, I argue that the relative consonance of the major and minor triads is inadequate to account for their essential role in tonal perception. Chapter 3 discusses an alternate psychoacoustic property that seems to be much more crucial – that of intervalllic rootedness. The process of intervalllic root-finding chooses one pitch of a dyad as primary and assimilates other sounding notes to it. This basic property of certain intervals is not only responsible for the perceived rootedness of harmonies, but it is a crucial factor in the induction of tonal centers as well. In order to understand the origins of intervalllic rootedness in the simpler process of pitch-finding, I undertake my own computerized model of rootedness perception as it is derived from the overall spectral image of a chord.

Chapter 4 explores the idea of a background collection or scale. Most contemporary theories of tonal hierarchy include a distinction between tones that seem to “belong” to a key and the chromatic tones that are somehow “outside” of it, and thus we must examine the phenomenological reality of this distinction, its possible origins, and the properties that govern it. I undertake a few brief analyses of Classic and early Romantic works and argue that, ultimately, the sense of an underlying scale is rather pliable and passive, with little direct influence on our event-hierarchical interpretation.
Chapter 5 discusses how temporal ordering and voice-leading can work together to create hierarchy, a phenomenon Jamshed Bharucha has called “melodic anchoring.”

Finally, chapter 6 recounts an elaborate self-experiment that observed stimuli created from various conjunctions of a dyad plus a single pitch. In the results we can see how horizontal and vertical intervals interact in a crude rhythmic context to create harmonies and tonal centers. Several factors that determine the tonic are identified in the data, including intervallic rootedness and melodic anchoring.

All of these discussions are undertaken with a mixture of approaches. I try to draw on the findings of experimental psychologists when possible, and even design a few computerized self-experiments to develop ideas and test hypotheses. These self-experiments are perhaps preliminary steps towards more carefully designed and conscribed tests with a pool of subjects. In analyzing my results I construct some quantitative models that are similar to Lerdahl's algorithms in an attempt to come up with an account that has the best fit to the data.

As a music theorist I am able to draw on mainstream theoretic ideas to guide my project. Many of the results are highly congruent with the assumptions and practices of Schenkerian analysis - this similarity is not quite intentional, but ultimately it is perhaps not surprising given the ambitions of the Schenkerian theory to thoroughly explain the experience of tonal music.

It strikes me that much cognitive modeling is also "theory" - given the assumption that the mind is a machine-like entity that must engineer every aspect of experience out of neurons and synapses, one is free to observe the parameters of experience and speculate as to what mental organization would be necessary to produce it. Such speculation may lack the empirical
grounding of physiological evidence, but it does have a logical rigor that should not be
discounted.

Ultimately I'd most like to identify with a phenomenological approach, with its careful
observation and description of sensory experience and skeptical resistance to received
knowledge. The present project is driven by a belief that the traditional vocabulary of tonal
analysis is grounded in a formalistic, score-based epistemology which obscures as much as it
illuminates and needs to be carefully rebuilt in experiential terms. I assume that my own direct
observations remain the richest and most reliable source of information, and that even
psychological studies with large pools of subjects can be “wrong” if they are motivated by
incorrect assumptions and poor analysis of stimuli and responses. As I explain in my chapter 6,
my elaborate, computer-mediated self-experiment with dyads and monads was in part inspired
by Husserl’s concept of “eidetic variation.”

While I assume that the experience of tonal perception is somewhat generalized or even
universal, it must be conceded that many of the observations herein reflect my experience alone.
One variable that simply cannot be accounted for within the scope of this project is the influence
of my own personal training, teaching, musical experience, and general acculturation. However,
I will proceed as though the perceptions I describe are shared, a strategy that is, for the time
being, an optimistic idealization.
Chapter Two

Vertical Relations, Part One: Consonance and Dissonance

At the heart of the tonal hierarchy lies the tonic triad. Once a key is established, the tonic and the tones of its triad tend to be perceived as more central or more structurally important than other tones. This is often the case regardless of whether the tonic tones are presented simultaneously, arpeggiated, or intermingled with other intervening tones. This essential anchoring role of the triad (as opposed to some other pitch configuration) seems to be determined by properties it possesses when heard as a simultaneity.

The next two chapters explore two essentially vertical aspects of sound that will be useful in our model of tonality -- the notion of consonance and dissonance and the phenomenon of intervallic rootedness. But in order to sufficiently account for either of these properties, we will first have to consider the structure of the overtone series, which is present in most naturally occurring pitched sound.38

The overtone series

The perception of pitch involves the conversion of vibrations (i.e. regular oscillations of air pressure) into the impression of constant sound. More rapid vibrations create higher pitches

38 One exception would be bell or gong tones, which present frequencies that do not conform to the series.
and slower vibrations create lower pitches. However, most pitched sounds involve not one vibration but multiple oscillations at mathematically-related rates - these vibrations are overtones. Thus, the perceptual system is constantly exposed to vertically layered sensations that are converted into unified pitch percepts. We are generally not conscious of these layers – we tend to perceive these complex stimuli as single pitches with timbres that vary according to the variety and intensity of overtones.

Overtones vibrate at rates that are whole-integer multiples of the base frequency or fundamental. Thus, for a pitch at A2 there will probably be vibrations at 110hz, 220hz, 330hz, 440hz and so on. (Acousticians occasionally speak of such frequencies as "harmonically" related.) Though the overtone series proceeds in regular multiples, these frequencies are not perceived as equidistant in musical pitch-space. The perception of pitch height actually functions on an exponential scale - a pitch one octave above $x$ is $2x$, but the pitch two octaves higher is $4x$, then $8x$, $16x$ and so on. Thus, overtones form an intervallic series in which each subsequent tone ascends by a smaller interval than the previous.

Example 2.1

a. Harmonics of 110hz (A2)
b. Graph of harmonics on logarithmic scale

![Overtones of A2 (110hz)](image)

The overtone series and tonality – historical precedents

Throughout the modern era theorists have had an idea that the primacy of triads had something to do with the physical properties of musical pitches. These claims have been inspired by the overtone series and its unmistakable similarity to the intervallic structure of a major triad. As one can see in example 2.1 (above), the first five overtones of A2 include frequencies that correspond to A3, E4, A4, C#5, and E5 – thus presenting all of the pitch classes of the A major triad within the tone itself.

For Rameau, knowledge of the properties of a vibrating body (the *corps sonore*) led to the assertion that the entire tonal system could be derived from the proportions of one tone. His
acoustically-based thesis reached its most elaborate form in his *Démonstration du principe de l'harmonie* of 1750:

The *corps sonore* - which I rightfully call the *fundamental sound* - this single source, generator, and controller of all music, this immediate cause of all its effects...does not resonate without producing at the same time all the continuous proportions from which are born harmony, melody, modes, and genres, and even the least rules necessary to practice.\(^{39}\)

Heinrich Schenker’s derivation of the prolongational background structure from the "chord of nature" is perhaps closest to the contemporary sense of tonal hierarchy, as the tones of the tonic triad are expected to act as hierarchically important events in any passage of tonal music:

In nature sound is a vertical phenomenon... In this form, however, it cannot be transferred to the human larynx...Therefore art manifests the principle of the harmonic series in a special way, one which still lets the chord of nature shine through. The overtone series, this vertical sound of nature, this chord in which all the tones sound at once, is transformed into a succession, a horizontal arpeggiation...\(^{40}\)

It is certainly understandable that theorists would seek a connection between the overtone series and triads - the similarity between the two structures is perhaps too striking to ignore.


However, no account of pitches and triads as external, objective phenomena can adequately account for their usefulness in tonal music. The musical properties of sonorities are the result of internal psychoacoustic processes -- we cannot understand the significance of the raw materials without looking more closely at what we do with them as listeners. In this chapter, as well as our chapter 3, we will look at the way the perceptual system processes layers of overtones and its consequences for the experience of tonal music.

**Consonance and dissonance**

One popular belief about the significance of the triad posits that its central role in tonality is due to its relative consonance. While it is my position that the interplay of consonance and dissonance has limited power to account for the functioning of the tonal hierarchy, this property is occasionally worth observing.

Speaking generally, dissonance refers to the extent that a tone or combination of tones sounds relatively harsh, unpleasant, noise-like, unclear, or unstable. The term often means different things in different contexts -- David Huron has identified thirteen distinct phenomena that might fall under this conceptual category.\(^{41}\) Consonance, on the other hand, seems to traditionally correspond to a subjective “pleasantness” and stability. For the bulk of the following discussion we will assume that consonance is defined simply as the absence of ____

\(^{41}\) personal communication
dissonance. We will return to the problem of defining consonance in positive terms at the end of the chapter.

The acoustic property that seems most closely related to the music-theoretical notion of dissonance within simultaneities is *acoustic roughness*, the perception of noise or interference that occurs when two tones are close together. Since the 19th century it has been thought that the musical intervals traditionally considered dissonant (such as the minor second, major second, diminished fifth, or major and minor seventh) are perceived as such because of the roughness generated by either proximate fundamental frequencies or proximate overtones. Example 2.2 shows how the intersection of overtones comes about in two pitches forming an augmented fourth or diminished fifth -- it is the roughness generated by these points of intersection that is thought to account for its dissonant status.

---

Example 2.2: Overtone interactions in a diminished fifth

How close do overtones need to be in order to cause roughness? The answer to that question must be expressed in terms of the critical band – a frequency distance thought to correspond to about 1 millimeter of physical distance on the basilar membrane. Simultaneous frequencies that lie within the interval of one critical band are perceived differently than those that are further apart. Relevant to our current discussion, tones within this range will produce some amount of sensory dissonance or roughness.

Experimental psychologists have differed on the exact frequency distance of the critical band, and as they relate it to a variety of phenomena it is possible that they are measuring multiple frequency bands as opposed to a single, unified property of perception. However, there is general consensus that the distance as it impinges on acoustic roughness is neither a constant musical interval (which would require a frequency range that expands exponentially as

---


pitch height increases) or a fixed frequency range (which would decrease like the intervals of the overtone series as pitch height increases.) It is more like the latter, however, in that it affects a larger musical area in low registers and a smaller area in high registers. Example 2.3 plots the size of the critical band as defined in the dissonance model we will use. The lines furthest from center indicate the point at which a second tone would create roughness.\footnote{Reiner Plomp and W. J. M. Levelt, "Tonal Consonance and Critical Bandwidth," \textit{Journal of the Acoustical Society of America} 38 (1965), 548-60. Plomp and Levelt actually thought that tones begin to interact at 1.2 times the width of a critical band as they defined it - that is the distance plotted in example 2.3.}

In a landmark study Reiner Plomp and Pim Levelt determined that perceived roughness follows a curve that peaks at 25\% of a critical band.\footnote{\textit{ibid}.} Their findings were fitted to mathematical equations by Hutchinson and Knopoff,\footnote{William Hutchinson and Leon Knopoff, "The Acoustic Component of Western Consonance," \textit{Interface - Journal of New Music Research} 7 (1978), 1-29.} and encoded into a computer program by Richard Parncutt.\footnote{http://www.uni-graz.at/~parncutt/rough1code.html, accessed 1/23/12} While there are aspects of the Hutchinson and Knopoff model that perhaps need revision, it seems adequate for use in the current study to make observations of relative consonance and dissonance. Plomp and Levelt's points of maximal dissonance are also plotted in example 2.3, as the flanking gray lines close to the center frequency.
Example 2.3

Critical Band from about C2 (65.41hz) to C4 (261.63hz)

- Dotted line: upper limit for interaction
- Light gray line: maximum roughness
- Dark gray line: referential pitch
- Middle gray line: maximum roughness
- Dashed line: lower limit for interaction

~C2 one grid line = one semitone ~C3 ~C4
Example 2.3 (continued)

Hutchinson and Knopoff assume that pitches have a spectral profile of $1/n$ where $n$ is the position in the overtone series. If a fundamental is modeled with an amplitude of $x$, the octave above it has an amplitude of $1/2x$, the fifth above that is $1/3x$, and so on. They consider only the first 10 partials of each pitch. The proximity and relative strength of all sounding frequencies is then evaluated, and any roughness between partials is summed into an overall score. They use the critical band function that I've plotted in example 2.3. Example 2.4 presents additional notes on the model and a graphical representation of one calculation, a roughness score for C4 and B♭4.
### Example 2.4: Calculation of roughness rating for \{C4, B♭4\}

**Notes on calculations**

All frequencies are “rounded” to their equal-tempered equivalents. Overtones that coincide on the same frequency sum together as $\sqrt{\text{amp}x^2 + \text{amp}y^2}$ (where amp\(x\) and amp\(y\) are amplitude of pitches \(x\) and \(y\)).

The width of the critical band is calculated at the mean frequency between \(x\) and \(y\), or \((\text{freq}x + \text{freq}y)/2\). The width at a given mean is $1.72 \times \text{meanfreq}^{0.65}$. Once the bandwidth is determined the distance between \(x\) and \(y\) can be expressed in terms of critical bands (\(\text{CBdist}\)) by calculating $|\text{freq}x - \text{freq}y|/\text{critical bandwidth}$.

The Plomp & Levelt dissonance curve is approximated as $(e \times \text{CBdist} / 0.25 \times e^{\text{CBdist} / 0.25 \times -1})^2$ where \(e\) is the natural log.

Roughness between overtones \(X\) and \(Y\) is $\frac{\text{amp}x \times \text{amp}y \times \text{disvalue}}{\text{amp}x^2 + \text{amp}y^2}$. The denominator normalizes the result by dividing by overall amplitude. Roughness between overtones \(X, Y\) and \(Z\) would be

\[
\frac{\text{amp}x \times \text{amp}y \times \text{disvalue} + \text{amp}y \times \text{amp}z \times \text{disvalue} + \text{amp}x \times \text{amp}z \times \text{disvalue}}{\text{amp}x^2 + \text{amp}y^2 + \text{amp}z^2}
\]

<table>
<thead>
<tr>
<th>overtones of C4 and their modeled amplitudes</th>
<th>overtones of B♭4</th>
<th>dissonance value</th>
<th>× amplitudes =</th>
</tr>
</thead>
<tbody>
<tr>
<td>D8 (4698.63hz)</td>
<td>1</td>
<td>.014</td>
<td>.002</td>
</tr>
<tr>
<td>C8 (4186.0hz)</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B♭7 (3729.31hz)</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A♭7 (3322.43hz)</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E7 (2637.02hz)</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F7 (2793.82hz)</td>
<td>2</td>
<td>.470</td>
<td>.008</td>
</tr>
<tr>
<td>D7 (2349.32hz)</td>
<td>2</td>
<td>.032</td>
<td>.001</td>
</tr>
<tr>
<td>C7 (2093.0hz)</td>
<td>1</td>
<td>.041</td>
<td>.001</td>
</tr>
<tr>
<td>B♭6 (1864.65hz)</td>
<td>4</td>
<td>.052</td>
<td>.002</td>
</tr>
<tr>
<td>G6 (1567.98hz)</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E6 (1318.51hz)</td>
<td>2</td>
<td>.088</td>
<td>.005</td>
</tr>
<tr>
<td>F6 (1396.91hz)</td>
<td>3</td>
<td>.726</td>
<td>.048</td>
</tr>
</tbody>
</table>
Example 2.4 cont.

This quantitative measure of acoustic roughness differs from our usual conception of consonance and dissonance in some important ways. Since the critical band is larger in low registers and smaller in higher ones, the same pitch intervals will have divergent values depending on where they are transposed - a low major second is more dissonant than a higher one. Dissonance values in actual music would vary even more - different combinations of instruments would also have differing levels of roughness due to the spectral profiles involved, and intonational nuance could also have a slight effect. While the model normalizes its dissonance scores by dividing by overall amplitude, the roughness of the actual signal would
increase and decrease as literal loudness does - thus a particularly loud sonority is rougher than the same one at a lower volume.

Musicians seem to abstract away general categories of intervallic consonance and dissonance from these widely varying stimuli. Nevertheless we can use this tool to observe trends when factors such as instrumentation and dynamics are otherwise equal, and it will be useful to determine the relative consonance of very complex sonorities that would be otherwise difficult to evaluate.

**Is the triad the "most consonant" sonority?**

Strictly speaking, the triad is not the least dissonant possible configuration of pitches. Simpler configurations such as the open perfect fifth and even the octave are actually more consonant. Yet subjects who are asked to rate the pleasantness of simultaneities rate fifths and octaves sonorities as less pleasant than thirds and sixths.\(^{49}\) We seem to prefer the sense of "fullness" or euphony that tertian sonorities provide.

It might seem, then, that the major and minor triads occupy a relatively fixed "sweet spot" in the spectrum of simultaneities that is not too plain but not too dissonant, and this position accounts for its status as the heart of the tonal hierarchy. However, the Hutchinson and Knopoff model does not paint such a simple picture. While major triads in certain inversions are indeed the most consonant trichords one can make, the range of possible roughness values

---

overlaps considerably with those of other sonorities considered dissonant. Example 2.5 presents a series of simultaneities ordered from left to right by their Hutchinson and Knopoff scores. I've included a collection of triad shapes built on C4 and also E4, to demonstrate that a modest elevation in register decreases the roughness value significantly. (For example, a root-position major triad on C4 rates .087, but the same triad transposed up a major third is reduced to .065.) Perhaps the most surprising result is that the augmented triad is scarcely more rough than the major. I've also inserted a few quartal and quintal sonorities into the group to show a few other untraditional trichords that can easily intermingle at the same basic dissonance level, and a diminished triad on F4 that is just under the roughness score for one of our major triads.

Example 2.5: Selected trichords and their Hutchinson and Knopoff ratings

Example 2.6 considers every possible two, three, and four-note simultaneity that can be constructed in the range between C4 and C5. The results are grouped by set-type, and the roughness score of each chord is plotted on a graph. One can see that, again, major and minor triads (set class 3-11) have the lowest average roughness value for a trichordal set, but there is significant overlap between its range and that of other set-classes, including the diminished triad, augmented triad, and quartal/quintal trichord.
Example 2.7 considers the same cardinalities distributed between C4 and C6. This wider range allows for more divergent registers to come in to play, as well as comparisons between close spacings (such as those in example 2.5) and more open voicings. With such liberal criteria for juxtaposition most of the set-classes overlap with set-type 3-11.
Example 2.7

Thus, it does not appear that the hierarchy-making power of the major and minor triads can be based on an absolute "sweet spot" of consonance. Within the same range of roughness scores there are other viable candidates for anchor sonorities which rarely if ever serve as tonic.
Relative consonance as tonic-making

It is possible, however, that it is the relative consonance of the triad in contrast to other, more dissonant sonorities that allows it to serve as a tonic. This is another idea that extends back to Rameau. Rameau thought that dissonance is essential to the cadences that establish a key. For him it was the dissonance in the dominant seventh chord that impels the harmony back to tonic -- even when the seventh is absent he argues that it is implied.\footnote{Jean-Philippe Rameau, \textit{Treatise on Harmony [1722]}, trans. Philip Gossett (New York: Dover, 1971), 65, 78-81, 235-9.} Following the same line of reasoning he asserted that the subdominant essentially includes an added sixth (say, an added D to the IV in C major), which is driven upwards (to E) by its dissonant status.\footnote{\textit{Ibid.}, 73-78, 240-245.}

Fred Lerdahl has generalized this principle to suggest that in the absence of other factors a relatively dissonant sonority will be hierarchically subordinate to a more consonant event.\footnote{Fred Lerdahl, \textit{Tonal Pitch Space} (New York: Oxford University Press, 2001), 349.} In the following chapters we will have some opportunity to observe whether this principle is significant in the shaping of tonality.

However, it is easy to imagine contexts where a relatively consonant dominant serves to tonicize a more dissonant tonic - in any minor key most iterations of the tonic are likely to have a higher roughness score than their surrounding dominant triads. Example 2.8 presents a i-V-i progression in G minor with its Hutchinson and Knopoff ratings – there we see that roughness scores for the tonic triads are almost double those of the dominants. Thus, while it remains
possible that contrasting levels of dissonance might be capable of establishing a tonic, it is clearly not a necessary or sufficient condition for the functioning of tonality.

**Example 2.8: Minor-mode progression with relatively rough tonics**

Example of a minor-mode progression with relatively rough tonics.

I do think it is possible that a more pronounced contrast in roughness may have some influence on our tonal interpretation, either as an enculturated cue for specific tonal contexts or a weak psychoacoustic principle. We will have some opportunity to observe this property in the experiment of Chapter 6.

**The problem of defining consonance**

Throughout this chapter we have referred to consonance and dissonance exclusively in terms of roughness, and it appears that modeled roughness scores alone can produce counterintuitive results when applied to some familiar tonal figures. However, since Helmholtz
first published *On the Sensations of Tone* musicians have protested that the concept of consonance cannot be adequately explained by the mere avoidance of clashes between overtones.\(^{53}\) It is thought by many that consonance must have a positive definition, by which certain preferred combinations of tones can be said to create a tangible, definable property, one that is usually assumed to be the underpinning of the entire tonal system.

Theorists from the time of Pythagoras to the present have held that consonance is created by pairs of frequencies with simple, whole-number ratios such as 3:2 and 5:4 (which produce the perfect fifth and major third, respectively.)\(^{54}\) However, numerical simplicity per se is not an explanation as to why an interval might be heard as particularly pleasant or stable -- there must be something about these frequency ratios that produces this desirable quality or qualities.

Since the pattern of overtones proceeds as a set of integer multiples (2x, 3x, 4x and so on) it follows that pairs of pitches in low-integer ratios (particularly those in the form of \((1+x) : x\)) will produce series of overtones in which multiple frequencies coincide. This high rate of overlap produces minimal roughness. As we saw in examples 2.5 through 2.7, above, the way the critical band varies in size with register complicates the way roughness relates to any particular simultaneity (as a higher-register instance is always less rough than a low-register one). We can neutralize the effect of register in our model, however, by considering intervals that are symmetrical around a fixed pitch. Example 2.9 graphs the modeled amount of roughness


in intervals as they expand outward from a fixed pitch of A440.\textsuperscript{55} Also noted as specific points on the graph are the twelve intervals of the octave in low-integer-ratio incarnations. We can see that the intervals of Zarlino’s senario, the octave (2:1), fifth (3:2), fourth (4:3), major third (5:4), minor third (6:5) and major sixth (5:3) all fall on local minima in the roughness curve. This is one explanation, then, for the perceived consonance of these intervals (though the minor sixth, also generally considered consonant, fares somewhat poorly, due to its proximity to the fifth.)

\textbf{Example 2.9: Roughness of intervals arranged symmetrically around A440}

\begin{figure}
\centering
\includegraphics[width=\textwidth]{example2.9}
\caption{Roughness of intervals arranged symmetrically around A440}
\end{figure}

\textsuperscript{55} The Hutchinson and Knopoff model considers the size of the critical band from an arithmetic mean of two frequencies, \((x + y)/2\). This is not the actual pitch-space center of the interval in question, which would be the geometric mean. In keeping with the model example 2.9 uses a constant arithmetic mean.
Another possible explanation for the special status of the low-integer ratios has to do with the periodicity of their combined waveforms. Waveform periodicity is one means of determining the overall pitch of a sound. With a simple sine wave, which presents a fundamental pitch and no overtones, it is obvious that the rate at which the wave repeats its pattern of positive and negative pressure will determine its heard pitch height. A sinewave at A440, for example, will repeat its pattern 440 times per second (example 2.10a). As we add overtones to our A, these create fluctuations in the wave that repeat at faster rates than the fundamental, but the overall waveform still repeats this intricate pattern 440 times per second – its periodicity remains the same (example 2.10b). A higher pitch like E660 would have a shorter period which repeats more cycles per second – in example 2.10c we can see that three of these waveforms take up the same timespan as two A440 waveforms (example 2.10c). Pitch combinations in low-integer ratios like 3:2 thus create a common periodicity that could theoretically be easy to observe – a perfect fifth creates a complex pattern that is twice the length of the lower pitch’s waveform and three times the length of the upper pitch (example 2.10d). Proponents of a periodicity-based model of pitch perception argue that it is the simplicity of this combined pattern that accounts for the special status of these intervals, producing an impression of “smoothness” that is generally called consonance.
Example 2.10: Periodicity in a perfect fifth

a) pure sine wave at A4, 440hz

period = 2.273 ms

b) fundamental + 3 overtones of A4, 440hz
(amplitude of overtones = 1/2, 1/3, 1/4)

period = 2.273 ms
Thus we have (at least) two explanations for the consonance of the low-integer ratios, one which relies on the coincidence or proximity of individual frequencies and another which appeals to a common periodicity amongst complex waveforms. Both models would, however, have to grapple with the variability of intonation in actual music, including the adoption of equal

c) fundamental + 3 overtones at E5, 660hz
(three repetitions fall in the same timespan as two, above)

\[
\text{period} = 1.515 \text{ ms}
\]

\[\text{period} = 4.545 \text{ ms}\]


d) A4 and E5 combined
temperament as a standard tuning system. Equal temperament alters these intervals so that their frequency ratios can no longer be expressed in low integers, yet the resulting sonorities are still generally deemed consonant.\textsuperscript{56}

The theory of coinciding frequencies can accommodate the alterations of equal temperament with ease, since the Plomp and Levelt curve posits a roughness value for the near-coincidences of equal temperament. Example 2.11 lists the ordered roughness values for the twelve equally-tempered intervals when symmetrical around A440, and again the intervals of the senario appear as the least dissonant, in an overall ordering that is mostly what we might expect. However, if lack of roughness is our definition of consonance, we would still need to account for the way roughness varies according to register. It seems possible that we abstract the varied manifestations of the intervals into these commonly accepted general categories of consonances and dissonances. Also somewhat problematic is that fact that listeners generally do not seek out a complete lack of roughness in

\begin{example}[h]
\textbf{Example 2.11: Roughness of equally-tempered intervals (symmetrical around A440)}
\begin{tabular}{ll}
\hline
\textbf{interval} & \textbf{roughness} \\
\hline
unison & 0 \\
P8 & 0.001 \\
P5 & 0.014 \\
P4 & 0.029 \\
M6 & 0.045 \\
M3 & 0.050 \\
m3 & 0.069 \\
m6 & 0.080 \\
TT & 0.081 \\
m7 & 0.083 \\
M2 & 0.175 \\
M7 & 0.213 \\
m2 & 0.452 \\
\hline
\end{tabular}
\end{example}

\textsuperscript{56} Musical performances that are not constrained by an equally-tempered, fixed pitch instrument can, of course, make use of intervals that are closer to their ideal small-integer representations, and many musicians of the modern era have deemed this highly preferable. Hindemith refers to the equally tempered system as “a compromise which is presented to us by the keyboard as an aid in mastering the tonal world, and then pretends to be that world itself” in \textit{The Craft of Musical Composition}, trans. Arthur Mendel (Mainz: Schott Söhne, 1945), 155.
musical sound – subjects actually prefer the euphony of thirds to the more perfect fourths, fifths and octaves.\textsuperscript{57}

A periodicity-based theory also needs to account for the way an altered fifth (or other tempered interval) would cause its two periodic components to drift out of sync, denying us a relatively short and simple common pattern. Peter Cariani’s temporally-based model of pitch perception relies on a complex inventory of all recurrent oscillations that is extracted from the waveform, obviating the need for a single repeating pattern.\textsuperscript{58} Tolerances within this process (which would allow recurrent segments of the waveform to be recognized as “similar enough”) might account for the detection of the ideal periodicities that are not quite present in imperfectly tuned intervals.

In defining consonance we need to clarify not only the underlying physiological theory, but also which experiential qualities we are associating with the term. Again, up to this point we have continued to focus on a “lack of roughness” (or a “smoothness.”) However, many psychoacousticians include other properties in with their definition, perhaps under the reasoning that consonance is an umbrella term that refers to any property of the low-integer ratios that seems essential to the functioning of tonality. Ernst Terhardt, for example, lists three additional phenomena as part of consonance – the “affinity of tones” (e.g. the qualitative similarity of the root and fifth of a harmony, which can lead to confusions between the two), the “compatibility of chords and/or melodic segments” (the notion that some combinations of tones seem to go well

\textsuperscript{57} Plomp and Levelt, “Tonal Consonance and Critical Bandwidth,” 551-2.

together and function interchangeably, and others do not), and the phenomenon of chord rootedness.59

In the current chapter I’ve argued that the level of roughness produced by triads cannot account for their primacy in tonal music. There are, indeed, other forces at work. However, I do think it is useful to carefully separate our terminology, making distinctions between the different observable properties of musical experience for the sake of precision. Thus, it is my opinion that the notions of “consonance” and “dissonance” might be best reserved for the sense that certain combinations of tones create a sense of “friction” or “clash” and others do not. One might further qualify these terms by calling them “sensory” consonance and dissonance.

Overall, we have eliminated several traditional assumptions about the role of consonance and dissonance in the functioning of tonality. The role of the triad as an essential structural sonority is clearly not due to an absolute level of sensory consonance, as there are sonorities with a similar consonance value that rarely if ever serve as tonics. Relative consonance is also not necessary to establish tonic, as more consonant dominants can easily serve to tonicize a relatively dissonant minor triad.

It seems true that, as others have asserted, there are other properties of verticals that are essential to the functioning of tonality. We now turn to one that seems very important indeed, namely the phenomenon of intervalllic rootedness.

---

59 “The Concept of Musical Consonance…,” 279.
As we discussed in the previous chapter, the absence of sensory dissonance cannot adequately explain the centrality of the triad in the tonal hierarchy. There must be something else about the most common materials of tonal music that is particularly conducive to establishing a tonal center. That something is the phenomenon of rootedness.

All trained musicians know that tertian sonorities have roots -- one uses the root to refer to chords and spell them correctly. The phenomenological reality of these roots, however, is perhaps less well understood. Simply put, in the absence of any wider context a rooted sonority will induce the sensation of a central tonic pitch class.

Example 3.1 presents a selection of possible stimuli which are rooted. All of these are likely to suggest that the pitch-class C is the tonal center, whether they present a fully-voiced triad (a), an arpeggio (b), a bare perfect fifth (c), or a diachronically presented fifth (d). I assert that this property is the sine qua non of the tonal hierarchy, quite possibly the origin of the very notion of tonic. To understand more about the sensation it creates, we must also consider its origins in the basic process of pitch perception.
Example 3.1: Four rooted stimuli

<table>
<thead>
<tr>
<th>a.</th>
<th>b.</th>
<th>c.</th>
<th>d.</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image.png" alt="Music Staff" /></td>
<td><img src="image.png" alt="Music Staff" /></td>
<td><img src="image.png" alt="Music Staff" /></td>
<td><img src="image.png" alt="Music Staff" /></td>
</tr>
</tbody>
</table>

The residue pitch

As we discussed in chapter 2, most naturally occurring pitched sound includes overtones, and the auditory system tends to transform the layered frequencies which are present in the sound signal into unified pitch percepts. The phenomenon of residue pitch provides intriguing evidence of how it goes about doing this.60

Residue pitch results from stimuli that contain layered overtones but are missing a vibration at the fundamental frequency. Such a stimulus sounds qualitatively different from a sound with an intact fundamental -- the timbre will be brighter or more "tinny" -- but it will still be assigned to the same pitch level. (We commonly experience such stimuli in using the telephone, which effectively filters out lower frequencies.) Example 3.2 presents a schematic representation of how pitches with missing fundamentals might be designed, and the corresponding audio example compares these three tones.

---

Example 3.2: Design for three tones at A3

<table>
<thead>
<tr>
<th>Tone A (with fundamental)</th>
<th>Tone B (no fundamental)</th>
<th>Tone C (high overtones)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A6 (1760hz)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G6 (1540hz)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E6 (1320hz)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C#6 (1100hz)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A5 (880hz)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E5 (660hz)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A4 (440hz)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A3 (220hz)</td>
<td>⇒ (pitch assigned here)</td>
<td>⇒ (pitch assigned here)</td>
</tr>
</tbody>
</table>

See ex3-2residuepitch.wav for audio

This phenomenon demonstrates that the auditory system is capable of recognizing the extant pattern of frequencies and determining where the missing fundamental would be. The process even works when many expected components of the tone are absent - the effect can be produced by isolated pairs of relatively high overtones (say, the fifth and six) and even by pairs of non-adjacent overtones. There are a few competing models of how this pattern-matching process is accomplished on a cognitive level. Ernst Terhardt has dubbed the assigned fundamental a "virtual pitch" and posits the existence of a learned template for pitch perception that can also account for many other details of audition, such as the stretched octave. Julius Goldstein describes the process as one of probability analysis, which determines the most likely

---


fundamental or fundamentals from a stochastically noisy signal. His approach is sometimes referred to as an “harmonic sieve.” Peter Cariani believes that the process is temporal in nature, accomplished by capturing the recurrent time intervals between spikes in the auditory neural response.

To look at this process in a very general way, the auditory system is detecting a pattern of frequencies and locating the bottom of that pattern, even if there is nothing sounding at that fundamental frequency. The pitch is represented at that level, and all upper partials are assimilated to this unified percept. (They are still audible, in a way -- the timbral representation of these tones is consistent with the way these frequencies would sound by themselves. Direct comparison of differently composed tones, such as in example 3.2, makes the contributions of individual overtones fairly explicit, and with practice one can learn to hear the partials within complex sounds. Leon van Noorden has suggested that we actually utilize two simultaneous auditory pathways to create this unified pitch-and-timbre percept, one of which determines the fundamental pitch level while the other has direct access to spectral information.)

The perception of intervallic rootedness has many intriguing parallels with the perception of pitches. It also involves the recognition of an intervallic pattern and the identification of a "bottom." Other sounding pitches are interpreted as "upper" components in the overall sonority.

---

63 Gerson and Goldstein, “Evidence for a General Template…”


they are audible, to be sure, but somehow hierarchically subordinate to the root. The root may not actually be the lowest sounding pitch, of course. Inverted sonorities present something other than the root in the bass – these harmonies are analogous to the residue pitch stimuli that lack their fundamental. Inverted configurations sound less grounded or less stable, much like the pitches with no fundamentals sound less satisfying than those with the full spectrum present.

Upper tones in a sonority (that is, those that are not judged to be the root) acquire an intervallic color relative to the root. Fifths sound “fifthy,” thirds are “thirdy” and so on. Only the root seems colorless, relating to nothing but itself. This configuration seems similar to the status of overtones in a single tone – the overtones are also subordinated to a bottom pitch location and add “color” to the sound.

Finally, there is a strong objective correlation between the shape of the overtone series and the phenomenon of intervallic rootedness. Intervals that occur lowest in the intervallic pattern of the overtone series are the most strongly rooted. (This trend excludes the octave and its multiples, however, which do not seem to be rooted at all.) The perfect fifth occurs between the second and third overtones, and its inversion, the perfect fourth, is presented between the third and fourth. These intervals have the most tonic-defining power. The major third which occurs next in the series is also somewhat strongly rooted.

Thus, we can confidently hypothesize that intervallic rootedness is a recursion of the process of pitch-finding. What might seem like an entirely “music-specific” phenomenon is grounded in one of the most elemental stages of aural perception. However, fashioning a theoretical tool that can predict the rootedness of a complex sonority while elegantly reflecting
this intuition is somewhat challenging. In the present study several approaches have been tried and ultimately ruled out.

A precedent: Hindemith’s *Craft of Musical Composition*

Paul Hindemith was a strong proponent of the notion that dyadic intervals have roots, and he presented a method of examining the intervalllic content of a complex sonority that could predict what its root would be.\(^67\) He thought that *combination tones* (also known as difference tones) were the acoustic basis of rootedness. Such tones are audible in a stimulus that presents two relatively loud, high-register pitches -- two flutes or a violinist playing double-stops can produce them. Given pitches X and Y, with Y being the higher of the two, the difference tone is a lower pitch that is equivalent to Y-X. A perfect fifth between A\(_5\) (880hz) and E\(_6\) (1320hz in just intonation) might produce the impression of a pitch on A\(_4\) (440hz), which is the difference of 1320hz - 880hz.\(^68\)

There also exists a second-order combination tone which is equivalent to 2\(X-Y\) (the difference between the first overtone of the lower tone X and the fundamental frequency of Y). For the A\(_5\)-E\(_6\) fifth, this will also produce A\(_4\) (440hz), but for intervals other than a fifth the second-order combination tone diverges from the first order. These pitches are also essential to Hindemith's theory.


\(^68\) In the discussion that follows I will continue to cite frequencies that are simple multiples of 440, rather than the equally-tempered frequencies that pitch labels normally indicate.
Combination tones are thought to be the product of a physical fact, what engineers would call intermodulation distortion. They have been demonstrated to be physically real under some conditions, and it is thought that the vibrations we perceive in certain musical situations are possibly formed in the inner ear.69

Hindemith noticed that in some common rooted intervals the difference tone lines up with the root of the interval to provide a sort of octave doubling. (This is true with the A5-E6 fifth described above, in which the first-order combination tone is A4, and the second-order tone is A6.) This, he argued, was the basis of intervalllic rootedness. However, there are some problems with this theory. It is somewhat weakened by appealing to both the first-order and second-order combination tones. He needed the latter to account for the root of the minor sixth, as a sixth containing C#6 (1100hz) and A6 (1760hz) produces a first-order difference tone of E5 (660hz) and a second-order tone of A4 (440hz). E5 is the “wrong” result but A4 doubles what we would expect to be the root. Hindemith also thought that the minor third was rooted on its lower member, but both difference tones present pitch-classes outside of the interval -- he thus appealed to a parallelism with the major third to argue for its rootedness.

Reiner Plomp has found that combination tones are inaudible under normal listening conditions.70 In addition, the theory offers no explanation as to why the doubling provided by combination tones would actually produce the phenomenon of rootedness, a vertical hierarchization in which the non-root is subordinate to the root.

---


70 *ibid.*, 1123.
Despite the problems with Hindemith’s causal explanation, *The Craft of Musical Composition* provides a useful practical method for determining the root of a complex sonority – the lowest, strongest rooted interval is said to determine the overall root. This will usually be the lowest perfect fifth. For instance, in a minor seventh chord containing \{C₄, E♭₄, G₄, B♭₄\} the root would be C -- the C - G fifth would trump the E♭ - B♭ fifth because it is lower in register. In a C major triad in first inversion, \{E₃, C₄, G₄, C₅\}, the C-G fifth would still trump the weaker E-C sixth, despite being higher in register, because fifths are more strongly rooted than minor sixths.

Ultimately, the only objection I would make about Hindemith’s root-finding method is that it seems somewhat ad hoc and inorganic – it would be more persuasive to construct a model that has a more direct relationship to the way the auditory system processes the sound signal.

We will try several variants of this approach below.

**First thesis: Applying Terhardt’s pitch-finding algorithm to rooted intervals**

Ernst Terhardt’s pitch-finding algorithm was first published in 1979 and later realized by Terhardt as computer code.\(^7\) It can use information on the frequencies within a sound signal to predict what pitches are most likely to be perceived. It accounts for many complex, real-world phenomena, such as the way sounds mask one another and the way that overtones that deviate from their ideal tunings can still become part of a unified pitch percept.

Because the perception of intervallic rootedness seems to be so closely related to pitch perception, it seems reasonable to expect that the Terhardt algorithm could be useful to model root-finding as well as pitch-finding. Not only would the model be able to identify, say, C4 and G4 from a sound signal that contains all of their component frequencies, but perhaps C4 (or some other pitch-class C) would emerge as a significantly stronger pitch-percept, a “super pitch” if you will, which might account for its status as root. If a direct application of the same model did not immediately achieve this result, perhaps a reasonable modification could account for the adaptation of pitch-perception to the similar but different process of root perception.

Unfortunately, simply applying the Terhardt model to rooted intervals produces counterintuitive results. In any interval where we would expect the lower tone to emerge as root, the pattern of overtones actually reinforces the upper tone instead, making it the stronger pitch percept. The opposite is true with intervals that are inverted with the would-be root on top - in those cases the coincidence of overtones serves to reinforce the lower tone.

Example 3.3 presents a modeling of an A3-E4 perfect fifth. Input to the algorithm was the two pitches with their first 20 overtones tapering off gradually in amplitude. The overall timbral contour of each pitch was modeled on a sampled flute tone from a commercial recording (example 3.3a-b)
Example 3.3: Modeling a perfect fifth with masking and overtone coincidence

a) Modeling the lower tone as input

Example 3.3c shows Terhardt’s prediction of how thepartials within these two notes would be perceived, with frequencies either combining together to create a stronger impression or attenuated due to the masking effect of other proximate frequencies.\(^2\)

\(^2\) Two modifications to the Terhardt computer code needed to be made to achieve these results. Terhardt’s masking function seems too strong overall – I modeled a new curve after R. H. Ehmer, "Masking Patterns of Tones," *Journal of the Acoustical Society of America* 31 (1959), 1115-20. Also, the program needed to be modified to allow...
c) Terhardt’s predictions of the perfect fifth as perceived

A3-E4 fifth as perceived

d) Lower tone isolated from (c)

A3 as perceived

virtually identical frequencies to combine into a single stronger overtone – otherwise the model predicted (contrary to experience) that they would obliterate each other in an extreme case of masking.
The pattern of reinforced overtones clearly benefits the upper E4, with every other overtone amplified by the combination with A3. (In A3, every third overtone is amplified, and the intervening partials are strongly suppressed by masking.) Thus, it appears that Terhardt’s pitch-finding theory cannot account for the emergence of a root.

Second thesis: Rootedness as "next-best" fit

It is possible that amplitude doesn't actually have much to do with pitch perception -- that, once above a certain threshold, all perceivable frequencies are considered more or less equally as part of an intervallic pattern. Goldstein takes this binary on-or-off approach in his model. Goldstein, Julius L. "An Optimum Processor Theory for the Central Formation of the Pitch of Complex Tones," *Journal of the Acoustical Society of America* 54 (1973), 1498.
frequencies to the overtone template -- if we consider, say, a \{C4, G4\} fifth the perception of these pitches would match two separate templates in those specific locations. The intervallic root, on the other hand, might be the best possible fit that considers all sounding frequencies as a single entity. Example 3.4 presents a few rooted intervals as partially-fitting overtone patterns. The perfect fifth represents a good percentage of the upper partials in the pattern, hitting 7 out of the first 10. From there each less-rooted interval covers fewer overtones, with the perfect fourth covering 5, the major third 4, and the minor sixth a somewhat problematic 3. Of course, even with this observation, we would need a good explanation for why a sub-optimal fit of the pitch template would emerge as such an influential perceptual force.
Example 3.4: Rooted intervals as partially-fitting overtone patterns

Third thesis: Rootedness in pitch-class space

At this point we may need to admit that the process of root-finding does not operate exactly like the process of pitch-finding. Pitch-finding, after all, actually produces a pitch percept, whereas root-finding produces a hierarchization that is pitch-class generalized. Given a C4-G4 fifth, the C4 emerges as root, but so too will any subsequent C in any register. All other pitch classes will also be perceived in consistent ways - all G's will sound like fifths, all E's as thirds, et cetera.
Richard Parncutt has suggested a model of root-finding that occurs in pitch-class space, and I will ultimately use a modified form of his algorithm as our rootedness tool. But considering our interest in grounding the theory in basic perceptual processes, it is perhaps appropriate to first ask, “What is a pitch class?” Is it an authentic dimension of perception or just a convenient theoretical abstraction?

Pitch-class identity (or chroma, as music psychologists often refer to it) is one dimension of the way pitch is represented to consciousness by the mind. As pitch height increases from frequency X, it will eventually reach a point Y at which the two frequencies sound somehow "the same" (even though X is definitely lower and Y is definitely higher.) In example 3.5 I illustrate this two-dimensional scheme as a spiral - as pitch proceeds upwards from the lowest perceivable frequencies it proceeds clockwise and outwards from the origin. However, all points on the same radius share the same chromatic identity - pitch height has been fitted onto an overlapping circle of equivalence.
Example 3.5: Frequency perception as a spiral, representation of A2 and its overtones

The mapping of pitch-class is somewhat analogous to the representation of color in the visual domain. There too, frequency values are converted to a property that is presented to consciousness. (Light is more conventionally described in terms of wavelengths, but these values also represent frequencies as with sound.) The lowest perceivable frequencies of light are represented as the red end of the perceivable spectrum, and the highest correspond to the violet end. However, we perceive no relative “height” in color, only a continuity between a few basic poles. In order to preserve this continuity between the primary colors, the mind represents a color area that bridges the gap between the low end of the spectrum and that of the high end -
this is magenta, which does not correspond to any physical frequency but is perceived when high (violet) and low (red) frequencies mix.\textsuperscript{74} This continuity is useful, because it allows us to view the physical world as continuous, even though our ability to detect light is finite and has fixed boundaries.

Why would we want to represent pitch in a two-dimensional scheme with both absolute pitch height and a circular chromatic identity? Why do pitches related by an octave sound somehow "the same"? The answer seems to have something to do with the challenge of representing pitch to consciousness. Remember that naturally occurring pitches are always layered in overtones that we combine into a unified percep. The aspect of timbre attempts to represent the presence and relative strength of these overtones within the sound.

The representation of timbre seems to be a translation of what these frequencies would sound like by themselves into a single, blended sonic object. A pitch with strong high partials sounds "brighter" than the same pitch with only low overtones. One might also say that the latter sounds "taller" in that we can hear that it occupies a frequency range that the other does not. Individual overtones are occasionally perceivable with careful listening. Ultimately there seems to be a continuity between timbre and pitch - as a partial is amplified it will begin to "stick out" and emerge as a quasi-pitch percep of its own.

The chroma circle makes combining overtones into a unified percep relatively easy. Any layered pitch is going to have octaves - the first ten overtones include five of them, counting three octave-multiples of the fundamental as well as the redundancies of the fifth and third.

\textsuperscript{74} David Kretch, Richard S. Crutchfield, Norman Livson, William A. Wilson, Jr. and Allen Parducci, \textit{Elements of Psychology}, 4th ed. (New York: Alfred A. Knopf, 1983), color plate after p. 112. Since the cones in the eye are exclusively sensitive to the red, green, and blue frequencies it would seem that all mental representations of "in-between" colors are similarly synthetic in nature.
Thus, by wrapping pitch perception onto a circle in which octaves sound like "the same thing" it becomes easier to present a unified sonic object.

Example 3.5 (above) demonstrates how a pitch at A2 (110hz) would be wrapped onto the pitch-height and chroma spiral so that some overtones line up. One might ask why we don't just declare an octave to fall every 110hz, thus making all overtones "the same" - the answer is that this would only work for a perfectly-tuned A and its multiples -- the partials of other pitches would not coincide at all. Combining circularity with a logarithmic scale creates a consistent structure for all pitches that mimics the multiplicative nature of overtone vibrations.

(This octave-related identity may be the result of the earliest organizational processes of the inchoate mind, as octave-related frequencies co-occur most frequently and are therefore integrated into the strongest neural networks. After all, octave-related multiples of 110hz are not exclusive to A-percepts. They also occur in any “subharmonic” of A -- all D’s, F’s and so on. As we are exposed to a variety of pitch levels octave relationships would be the most common recurring element.)

Modeling intervallic rootedness as a transference of the pitch template into pitch-class space, then, seems plausible enough. Richard Parncutt has published a root-finding model that considers all sounding pitches as pitch-classes, and then finds the best fit of an octave-generalized template based on the first ten overtones.\textsuperscript{75} The process is illustrated in example 3.6. The lower (and more numerous) overtones are given a higher weight - the bottom of the overtone

series, representing the fundamental and its octave doublings, is worth 1, fifths are worth 1/2, major thirds 1/3, minor sevenths 1/4 and ninths are worth 1/5. One can imagine this template being rotated around the circle to find the highest-scoring fit - that position will indicate the root of the stimulus.

**Example 3.6: Parncutt's root-finding template in pitch-class space**

Since the pitch-classes taken as input are octave-generalized, this method will account for inverted intervals as easily as "normal" ones -- the fourth G3-C4 looks the same as the fifth C4-G4 before the template is even applied. Either stimulus would be evaluated as in example 3.6, where the best positioning of the template is on C, with a score of 1.5.
The pitch-class-generalized model of intervallic rootedness works fairly well, except that it ignores the fact that registral position does matter in determining the roots of somewhat ambiguous combinations of PCs. The minor-seventh chord (set class \[0358\]) is the classic example, containing two fifths (or fourths) which could potentially define the root. If the chord is stacked as \{C₄, E₄, G₄, B₄\} the C-G fifth will probably define a C root, and thus the chord will be heard as a minor seventh. However, if the C is simply inverted to the top, as in \{E₄, G₄, B₄, C₅\} the E-B fifth will dominate and the sonority will be heard as a major triad with an added sixth. (Note how the modal color of the entire sonority changes from minor to major due the shift in vertical hierarchy.)

Parncutt's method cannot resolve this ambiguity, since both chords look the same in pitch-class space. In a subsequent revision of the model he solved the problem by tacking on a "bonus" score to the lowest note in the voicing, thus tipping the balance when potential root scores were otherwise equal.\(^{76}\) This method seems somewhat inelegant, in that it fails to treat all pitches within the stimulus consistently. We can take Parncutt’s bonus score and turn it into a weighted continuum with a little logarithmic math.

Parncutt's pitch-class circle becomes a pitch spiral if we weight all pitches on a fixed scale of 1.5 to the \((12-n)\text{th}\) power, where \(n\) represents the number of octaves from the bottom. For convenience we'll define the "bottom" as C₀, and thus \(n\) is equivalent to register number. C₁₂ is thus worth a value of exactly 1 (or 1.5⁰), C₁₁ is worth 1.5, C₁₀ is 2.25, and so on all the way down to C₀ which is worth 1.5₁² or about 129. Each octave down is worth 1.5 times the one.

above it. Our exponent $n$ is fragmented in 12ths for each chromatic step down, so B11 is worth $1.5^{1/12}$, Bb11 is $1.5^{2/12}$, and so on. This scale represents an intuition about pitch perception nicely, that lowness of register corresponds to a relative "heaviness" or substantialness. Using a logarithmic scale for registral weightedness necessitates a computer program to calculate results, which I have realized in C++ for a Windows PC.\footnote{"PitchClassRootedness.exe," hosted at http://davesmey.com/dissertation.}

Our model thus rotates a template similar to Parncutt's around the pitch-class circle, but it multiplies the intervallic value in each position by the weight of the pitch being considered. Register matters -- the lower fifth will always beat a higher one because it is weighted more strongly. In testing this algorithm, I found that I had to modify Parncutt's values in the pitch-class intervallic template, increasing the value of roots, fifths, and major thirds. The circular template now awards 1.33 for the root and its octaves, 1.1 for fifths, .42 for thirds, .25 for minor sevenths and .2 for ninths. Example 3.7 illustrates the calculation for a simple major triad on C4.
Example 3.7: Rootedness calculation with registrally weighted PCs

As the template is rotated around the chromatic circle it produces a series of values for each position. The "winning" value is judged to be the best position for the template and thus the best root for the stimulus. We can judge the relative strength of the root by how decisively it beats other candidates. If the best root scores are virtually identical (say, within 2% of each other), we can consider the stimulus to be so ambiguous as to be un-rooted. We'll declare the minor sixth (in which the upper note beats the lower by 7.4%) as the threshold that defines "strongly rooted," and the range between 7.4% and 2% will be "weakly rooted." Example 3.8 lists all the simultaneities tested in the course of this study. All chords are spelled so that a C
root is what I would consider the correct result. (Usually this is the conventional theoretic root, but some tested configurations revealed unconventional roots that I find persuasive.) The majority of the chords are strongly rooted. A "W" indicates that a chord is weakly rooted, and an "A" indicates that the sonority is judged to be ambiguous. An "X" marks chords with undesirable results.

Example 3.8: Sonorities tested in this study
We can look at a few details of the tool to appreciate its inherent tensions. The purpose of the "root and octaves" value in the circular template might initially seem confusing. After all, unlike the other intervallic relationships in the circle, the octave does not seem to be rooted, and awarding all notes a value of 1 1/3 as the template rotates around might seem redundant. What this position in the template does, however, is privilege "something over nothing," giving notes that are actually present in the stimulus more weight than notes that are absent. Example 3.9 illustrates two rotations of the template for a C-G perfect fifth - without some value for a root on C, the model would prefer to interpret both notes as upper partials of an absent F!
Example 3.9: Perfect fifth illustrates the function of a unisons / octaves value

Combined with the registral weighting this unison value allows some lower notes to "break free" from a rooted relationship as they are extended by an octave. The minor sixth, for example, is typically rooted on its upper note, as an inversion of a major third. The minor thirteenth, however, is predicted to be rooted on its lower note, as the unison value times its relatively "heavy" registral weighting trumps the value of the major-third relationship. This prediction seems experientially correct, as an otherwise unadorned E3-C5 gives the impression of an unstable ♯6 on top, which wants to descend to ♯5. Generally speaking, the stronger the values in the intervalllic template are, the more pitch-class generalized the model is, whereas increasing the octave weighting causes register to become a more disruptive factor.
Example 3.10: Minor sixth versus minor thirteenth

minor sixth E4-C5

root / octave
1.33

C root score: \( 1.33 \times 6.63 + .42 \times 8.69 = 12.47 \)
(next best root is E, with 11.56)

minor thirteenth E3-C5

root / octave
1.33

E root score: \( 1.33 \times 13.03 = 17.34 \)
(next best is A with 14.34, then C with 14.29)
Minor triads are handled by the model without much difficulty, as the fifth relationship trumps the major third between upper tones of the triad. However, one particular voicing of the minor triad with a third on the bottom followed by root and fifth proved to be particularly fragile -- as the model was adjusted in various ways the E₅ would occasionally be chosen as root. Again, playing the sonority suggests that this result is somewhat appropriate, as it is a particularly subtle voicing of the minor triad. This example proved to be a useful limit on the relative strength of thirds as opposed to fifths -- thirds could not be nudged above .42 in the template without upsetting this result.
Example 3.11: A "fragile" first inversion minor triad

Our test stimuli followed conventional standards of registral arrangement and doubling for a three or four-voice texture, and for the most part these yielded good results. A doubling of major triads in first inversion, however, revealed an undesirable effect when multiple copies of a non-root (in this case the third) are considered. The model as it stood would add together the weighted value of the top E and bass E to privilege the third an excessive amount, and thus E would emerge as root due to nothing other than its "unison or octave" value times this doubled weight. This result did not match experience, so I needed to add an exception for octave doublings -- if any two notes are the same PC, the weight of the upper note is cut in half (example 3.12.) Now simply doubling the bass of a first inversion triad does not sway the overall result. However, tripling the bass does sway the balance, which seems appropriate, as this unusual configuration can indeed sound like a minor sonority with an added flat thirteenth.
Example 3.12: Adjustment for octave doubling

One particularly surprising result has been the half-diminished seventh chord. One does not expect it to be strongly rooted, as it traditionally would be an inappropriate sonority to establish tonicity. However, the perfect fifth between third and seventh result in a strong rootedness rating for the third of the chord. Audition reveals this to be plausible - if the chord were to serve as tonic harmony, the tone we consider to be the third is the most plausible home note. Thus, from the standpoint of intervalllic rootedness the sonority is "really" a minor triad with added sixth. (See below for more thoughts on the significance of these roots in a tonal context - I am not denying that the chord works well in its more traditional functions.)

C root score: $1.33 \times 6.63 + 1.1 \times 7.85 + .42 \times 13.03 + .42 \times \frac{8.69}{2} = 24.76$

(Next closest is E, with 23.12)
We have one remaining problem to consider, cases where the model predicts an absent tone as root. For example, the model predicts a C root for a \{G, E\} major sixth. This result seems experientially correct!

**Example 3.13: Major sixth**

C root score: \[1.1 \times 7.85 + .42 \times 5.79 = 11.07\]
(next best is G with 10.44, A with 8.33)

However, there are other cases where the prediction of a phantom root seems unjustified. The model predicts a missing A\flat root for a \{C, E\} minor third as well as the \{C, E\flat, G\flat\} diminished triad. A fully diminished seventh chord \{C, E\flat, G\flat, B\flat/\}\ receives a predicted F root (since C can serve as upper fifth, E\flat as minor seventh and B\flat/\ as upper third.) If we exclude these phantom roots the lowest sounding tone (C) scores highest, which seems more plausible in
all three of these cases. But how do we account for the inconsistency between the minor sixth, which seems rooted on an absent tone, and these other sonorities that do not?

I think the explanation for this inconsistency lies in the influence of enculturation and the subjective contexts we impose on these materials. If we actually follow a minor third with its predicted "completion," the result makes sense - a \{C, E_{\flat}\} dyad can be revealed to be the upper part of an A_{\flat}-major triad (example 3.14a.) However, we don't tend to imagine this completion - hearing C as tonic is simply easier and perfectly plausible. With a \{C, E_{\flat}, G_{\flat}\} diminished triad we are primed by experience to auralize a very specific resolution to a D_{\flat} tonic (example 3.14b). However, there seem to be enough prominent uses of a \{G, E\} sixth as °5 and °3 to cue its subjective orientation towards a phantom C tonic.

Also, in general, it must be noted that it becomes easier to imagine sonorities as upper parts of a missing tonic when the tones sound in a relatively high register, and less plausible in low registers. A \{C_2, A_{\flat}2\} dyad is perhaps more likely to be heard as an unstable °1 and °6, not °5 and °3. A high register minor third like \{E_5, G_5\} actually does sound like °3 and °5, whereas a middle or low-register third does not.
Example 3.14: Possible continuations of stimuli

a. minor third  
\[ \text{\textbf{strongly rooted, but unlikely auralization}} \]

b. diminished triad  
\[ \text{\textbf{likely auralization}} \]

c. diminished triad  
\[ \text{\textbf{strongly rooted, but unlikely}} \]

In the instance of the \{G, E\} sixth, the subjective contextualization is actually providing a missing tone - in a sense we are not measuring the psychoacoustic properties of a mere sixth, but rather a subjective \{G, C, E\} triad. Thus, our theory is actually consistent - predicted phantom roots do not exist unless they are actually subjectively auralized. Lacking this subjective context absent tones should be excluded from consideration and the highest-scoring sounding tone should be considered the best root.

The acoustic root in context

We now have some idea of the origin of intervallic rootedness, but what is its significance in a typical tonal context? Do we actually hear a clear series of roots as a passage unfolds? Theorists from Rameau onwards have believed this to be the case. Rameau’s \textit{basse fondamentale} presented an acoustically-derived root for each harmony, and the practice of following these roots throughout a progression gave musicians an extremely useful way to track
the harmonic contents of a passage and think about the way each sonority might interrelate or “function.” The acoustic root of each harmony still remains important for some contemporary theorists, and any analyst that uses roman numeral terminology is engaging this belief that roots constitute an essential reference point for the underlying sonorities in a progression.

However, I would like to suggest that for the typical harmony that is not a tonic, intervallic rootedness may not have a significant influence on our interpretation of the passage or a perceptible impact on our experience. I've identified two experiential consequences of the root - it identifies a tonic, and it organizes all other tones in a hierarchy, characterizing them by their intervallic relation to tonic. (Thirds sound "thirdy", fifths "fifthy" etc.) I assert that, for the most part, a new hierarchy is not created when we move to a non-tonic harmony. There is, however, some room for nuance based on localities of attention – the hierarchy of each harmony represents an alternate perspective that we can (and, to be sure, often do) shift to in the course of a passage. For the sake of reducing theoretical overhead the discussion that follows takes some important principles of voice-leading, grouping, and metric hierarchy as given.

In general, one should view the rootedness of a sonority as a potential, a property that may influence our interpretation of a particular moment in music. The networked view of cognitive organization holds that our minds continuously activate some associations and meanings that are then filtered out as "wrong" - the circuit in question attempts to nudge our conscious perception in a certain direction, but may be overruled by other factors. Ray

---


79 Richard Parncutt argues that the root of a simultaneity emerges as more salient than other tones, and this has consequences for other intrachordal measures in his *Harmony: A Psychoacoustical Approach* (Berlin: Springer-Verlag, 1989). Fred Lerdahl constructs a local hierarchy for each harmony in his Tonal Pitch Space model. See TPS, 53-59.
Jackendoff, for example, frequently appeals to a “selection function” that acts as a gatekeeper to consciousness and suppresses conflicting information, thus allowing higher-level processes to work efficiently.80

In the absence of any other context, the rootedness of a sonority will usually suggest a tonic. Once that tonic is established, however, subsequent stimuli have a much lower chance of supplanting it. Example 3.15 presents an extremely simple scenario, a C-G fifth followed by some E♭-G thirds -- although the thirds are rooted, one would not expect them to dislodge the C orientation that has been established on the first beat. They would, rather, reinforce the tonic. Rooted intervals work diachronically as well as simultaneously, so as the thirds are sounding the upper G reiterates the fifth of C. Understanding these tones in the context of C creates a feedback loop that strengthens the original orientation.

Example 3.15: Hypothetical C minor passage

![Musical notation]

Non-tonic harmonies are understood in a similar way, relative to the tonic. The importance of V as the most useful contrasting pole to tonic seems determined, in part, by the

relationship of its root to $\hat{1}$. As the bass moves to $\hat{5}$ it is easy to hear as an upper fifth, and the returning $\hat{5} - \hat{1}$ motion reinforces the tonic very strongly. But even this perception of $\hat{5}$ in itself reinforces the sense of the absent $\hat{1}$ - that is why half-cadences are so clear and tonally coherent. In this situation I would argue that we aren't really perceiving the root of the dominant triad as such - we are hearing it as the upper fifth of the tonic.

Example 3.16: Half-cadence reinforces tonic

(In addition, due to perceptual properties we will discuss in chapters 4 and 5, we hear the other tones in the V chord (scale-degrees $\hat{2}$ and $\hat{7}$) in a stepwise relationship to $\hat{1}$. These neighbor relationships are also very palpable and hierarchical in nature, and also serve to reinforce our orientation around the tonic. Generally speaking, all harmonies that are heard in a key relate to tonic through some combination of stepwise connections and rooted intervals, and the specific way in which they do this constitutes their function in the key. Unfortunately, however, a thorough discussion of harmonic function lies beyond the scope of the current project.)
There are also moments when we relate tones forward to a tonic that hasn't occurred yet. For instance, if we audition an isolated sonority at the piano, it doesn't always sound like a tonic. We may hear it as a dominant or other non-tonic harmony. This is certainly likely if we consider the major-minor (aka "dominant") seventh chord. In acoustic terms it is very strongly rooted, but we tend to auralize a connection to a tonic that is not present. This imaginary tone is real as far as we are concerned, and once we consider it we engage the same feedback loop -- the 5-1 relation reinforces the orientation, making it plausible and stable. In a recent study experimenters found that their most expert subjects even interpreted isolated major triads as dominants about 17% of the time, a result I suspect was encouraged by mixing dominant sevenths and triads together in the pool of stimuli.81

Example 3.17: A dominant seventh and auralized tonic

Much of in-the-moment tonal perception involves this kind of imagination and projection -- we are not only hearing stimuli in relation to what came before but also deciding where we expect to go next. The opening measures of Schumann's *Carnaval* illustrate the role of our

---

future expectations on our in-the-moment interpretation (example 3.18). The first few chords are thunderously voiced and rhythmically accented, and the A\(^{b}\)\(^{6}\) sonorities form a dominant-tonic relationship with the more accented D\(^{b}\) triads. If the first eight beats were taken out of context and presented to a naive listener, one would expect them to hear D\(^{b}\) as tonic. However, if we are familiar with the piece we know that the D\(^{b}\)'s will fall back down to the C's, and A\(^{b}\) will quickly be revealed as tonic in a final cadential motion. (Schumann clarifies this with a repeat, forcing us to hear opening material in light of this "twist.") An experienced listener will even have trouble hearing the out-of-context opening chords as tonics - we know they are subdominants. Our memory of what happens next completely transforms our perception of the chords.


"Naive" hearing of opening measures

\[\text{Quasi Maestoso} \]

\[
\begin{array}{c}
\text{D}^{b}: \ V^{6} & \ I & \ V^{6} & \ I \\
\end{array}
\]

(continuation)
Non-tonic harmonies as a shift in perspective

Acoustically rooted, non-tonic sonorities do have the potential to create their own hierarchies that are similar in kind to the tonic hierarchy. To think of the most extreme case, we might say that every harmonic root is a would-be tonic which, if given enough time and enough supporting details could become a full-fledged key that essentially makes us forget our original tonal orientation. In the typical harmonic progression we resist this shift of perspective. However, there is room for nuance. A harmony that is elaborated somewhat with rhythmic and melodic figuration can take on a local hierarchization that mimics tonicity.

We can perhaps get a better sense of how we shuttle between perspectives by considering a simple and clear-cut example of a tonicization of the dominant, from the introduction to Fauré's chanson "Clair de lune" Op. 46, No. 2 (example 3.19). The first two phrases begin and end on the tonic of B♭ minor, and would probably be easy to understand on a moment-to-moment basis by tracking each harmonic move as it relates to the home note. The third phrase (mm. 5-8)
moves to the dominant triad and repeats it regularly with a few intervening \{B\flat, D\flat, F, G\sharp\} sonorities - this is the tonic altered to serve as a Rameauian subdominant of V with added sixth! (The added sixth does indeed make the harmony more dissonant than the dominant, whereas substituting another F for the G gives it a lower roughness score. I suspect, however, that the repetition and rhythmic emphasis of V would be enough to tonicize it without the added sixth.)

Example 3.19: Fauré "Clair de lune" mm. 5-8

As we listen to this phrase I think we will tend to shift to an F-oriented interpretation. A's sound like 3 and, most crucially, B\flat is its upper neighbor and a "tendency tone." This is achieved by the time we reach the downbeat of measure 6, as one of the secondary subdominants has resolved to F and another measure-long chunk is beginning on this now-stable harmony. However, we have an ability to choose perspectives - we can "remind ourselves" that the A is the leading-tone, and anticipate its return to B\flat. Likewise was can consciously continue to hear F as
an upper fifth. Essentially this shift involves thinking ahead to the return to the original tonic, to actually anticipate measure 9 while the phrase is still sounding.

Ultimately, I don't think it would be correct to assert that we can hear both hierarchies simultaneously. They are mutually incompatible. They are both available to us, but at best we can consciously choose (or unintentionally wander) between them. Non-tonic rootedness is perceivable to the extent that a passage can distract us and cause us to shift focus, and it coexists with the prevailing key to the extent that we can shift back to the tonic orientation at will, without the music explicitly forcing us to. (To be sure, Fauré does actually force us to return to B♭ minor orientation by destabilizing F in measure 8.)

There does seem to be at least one sense in which the two perspectives can coexist and blend. I've claimed that the intervallic color of tones is dependent on a vertical hierarchy that is centered on a tonic – thirds sound “thirdy” because they are heard relative to tonic and so on. The rootedness of non-tonic harmonies may also be perceivable to the extent that the localized character of tones "bleeds through" into our tonic-centered perceptions, so that the leading-tone can also sound like a third, or perhaps the third scale-degree is stripped of its thirdiness by a move to the mediant harmony. It seems that, in general, we don't conceive of scale-degrees exclusively as they sound relative to tonic, but also hear them in their local, hierarchically weighted intervallic identities. This blurring of intervallic character is not necessarily a perspectival shift like the change in tonic orientation - it is apparently perfectly coherent for a leading tone to sound "thirdy" and yet be strongly subordinate to its tonic.
The global significance of tonic tones

In our previous segment I argued that the root of the V chord interacts with an underlying Ê as it is sounding, creating a feedback loop which serves to strengthen our sense of key. That feedback loop also elevates the importance of V, so that it will tend to emerge as more structurally significant than surrounding events. Throughout a passage of tonal music, all three tones of the tonic triad have a general structural advantage over other tones, and they will tend to emerge as locally important.

A brief French folk melody illustrates this tendency in a very simple context (example 3.20).82 The notes I’ve marked with open noteheads in the lower staff all belong to the tonic triad, and collectively they seem to occupy the highest levels of structure for the passage. Each of these arrivals on a 3 or 5 points back to a previous tonic, and this property creates a particularly “close” or “simple” relationship with the underlying Ê. Thus, these scale degrees make more sense than the surrounding tones as event-hierarchical reference points. (Also, to be sure, rhythm, meter, and grouping strongly emphasize these tones as structural. Every structure-making property at our disposal is “in phase” to produce an exceedingly simple, stable, and coherent result.)

82 This melody appears in Robert W. Ottman’s Music for Sight Singing 5th ed. (Upper Saddle River, NJ: Prentice-Hall, 2001), 78 as excerpt #291.
Example 3.20: Tonic tones emerge as structural in a French folk melody

Note that this general advantage for tonic tones does not mean that every 1, 3 and 5 in a passage will rise to the highest level of structure. As I discussed above, the rootedness of intervals should be understood as a potential that may or may not be engaged in any given instance. In the above passage there are actually two 5’s and one 3 that are subordinate to their surrounding tones due to the structuring forces of meter, implied harmony, and the norms of harmonic progression. (Specifically, the 5’s of measures 2 and 6 are part of a hypermetrically weak dotted-quarter beat, and they serve to fill out dominant harmonies that are subordinate to the flanking tonics. The 3 in measure 10 makes the most sense as a passing tone that connects an underlying subdominant to the underlying dominant of measure 11.)

If intervallic rootedness is the property that globally elevates the tonic tones, we have a familiar problem in accounting for the minor mode, namely that the relationship of 1 and the
minor 3 is not rooted. It would be a potentially interesting project to survey a corpus of monophonic music and examine whether the minor 3 is handled differently than the major 3.

We know that the minor 3 and 5 do have an intervalically rooted relationship which emphasizes 3 as a potential tonic, and many minor-mode melodies do exhibit a tendency to suggest tonicizations of III (or a brief move to the relative major). Even this tendency, while disruptive for our sense of tonic, would serve to elevate the minor 3 relative to its surrounding tones!

Can any sonority serve as the center of the tonal hierarchy?

We’ve seen that the major or minor triad serves to indicate and reinforce a tonic through the property of intervallic rootedness, and that, even when presented diachronically, the tones of 3 and 5 both reinforce the key and are structurally elevated as a result. The question remains as to whether other rooted sonorities (such as a seventh chord, a stack of fifths, or other common 20th-century construction) could serve the same function and thus constitute the second layer of our hierarchy. The answer seems to be “yes and no.”

It is clear that any rooted sonority can have the same immediate effect as the major and minor triads, orienting us around a possible key center. “Extra” tones may make a sonority more ambiguous, but according to our model they will not necessary destroy this effect. A composer could easily choose to consistently return to a non-traditional tonic construction and thus create the impression that all the tones presented therein are an essential part of a referential sonority.

However, it does not follow that the tones from a non-traditional tonic will be globally
elevated in the same manner as \(^3\) and \(^5\), as they may lack an intervallically rooted relationship to tonic. (Indeed, in my chapter 6 experiment the major third and perfect fifth were the only intervals to exhibit a consistent influence on tonal interpretation that can be attributed to rootedness – even the minor 7th failed in this regard.) At best, a composer can hope to create what Joseph Straus calls an “associational” relationship with tonic, to present the tone in such a way as to remind us of its original context and thus evoke a similarly stable and weighted relationship.\(^83\)

Thus, in a sense, the property of intervallic rootedness is sufficient to account for the triad as the second layer of the tonal hierarchy, if we assume that membership in this layer means that the tones will necessarily exhibit a psychoacoustically primed structural advantage over other tones throughout a passage. However, as we have seen in the current discussion, the way our sense of key relates to the musical surface is rather complex, and many other structure-making factors are potentially available to a composer. Even if we accept this account of our tonal sensibilities as essential and unalterable it does not follow that tonal music must necessarily be triadic.

---

Chapter 4

The Underlying Scale

Up until this point we've been primarily concerned with the status of the tonic tone and the members of its triad - in a typical piece of tonal music these are the pitch classes that emerge as most salient or structurally important. Beyond the tonic triad, there is one more level of differentiation between pitches, the distinction between tones that seem to belong to the key’s underlying scale or collection and the chromatic tones that are perceived to lie somehow “outside” of it (example 4.1).

Example 4.1: Krumhansl and Kessler's results for the major and minor modes, revisited

![Major Key Profile Diagram]
The Krumhansl probe-tone experiments elicited the scalar vs. chromatic distinction in response to two different kinds of stimuli. Some prompts included all of the tones in a major or minor scale, which were sounded in the course of a short harmonic progression. Others presented only a single major or minor triad. Thus the experiment could conceivably be measuring two different but related psychological processes. In responding to a passage that contains a complete diatonic collection, the scalar vs. chromatic distinction could be a simple matter of recognizing whether a pitch had been heard recently. However, subjects also rated major-scale tones higher when presented with a single major triad, and likewise preferred minor-mode tones in conjunction with a minor triad. These results suggest some other process at work.

---

-- either an acculturated expectation that certain tones accompany other tones, or some deep structural property of scalar “belongingness” that remains undiscovered.

In both of these results we see that tonal music generally implies an underlying collection that includes some tones as essential or fitting and excludes others as outside or foreign, and we can expect this distinction to have some effect on our experience of this music. The current chapter will explore the psychological processes underlying this phenomenon and its consequences in musical perception.

Annabel Cohen's *Well-Tempered Clavier* experiment

As a passage of music unfolds we seem to have access to a relatively short-term, unordered memory of the tones we have been hearing. A study conducted by Annabel Cohen attempted to demonstrate the reality of this ability in the psychology lab. She measured the ability of undergraduate music majors to accurately sing the underlying scales implied by passages of real music, namely the first twelve preludes of *The Well-Tempered Clavier*, Book I.85 For each prelude she assembled four different excerpts -- a brief incipit of four notes, the first four measures, the first eight measures, and the concluding four measures. Across all stimuli, the students were able to successfully sing the scale and tonic indicated by the key signature 53% of the time. This demonstrates that scalar memory certainly exists, but the success rate seems quite low. Performance seems to have been driven down by the selection of passages of arbitrary

length. Many of the excerpts concluded in tonally awkward places or even mid-modulation, yet only the tonic indicated by the key signature was deemed correct. The picardy third at the end of the minor-key preludes also created modal confusions. If we consider the four-measure excerpts that actually conclude on or near the tonic harmony, a strong majority of subjects were able to perform the appropriate scale on the appropriate tonic. (For example the scores for the first four measures of the C major, C minor, C-sharp major and E minor preludes were 89%, 72%, 89% and 61%, respectively.) The study also showed a strong effect of scalar expectations, in which a partial stimulus elicited the complete correct scale. Presenting the first four notes of a prelude (which tend to indicate the tonic harmony and not much else) actually yielded the best results, with 74% accuracy across the board.

Cohen’s results demonstrate that it is possible for subjects to retain the pitch collections from recently heard passages and assemble them into ordered scalar sequences. My own similar home experiment, discussed below, arrives at the same conclusion. Since this procedure appears to be relatively easy and automatic, we might ask whether it is grounded in more general processes of everyday hearing. The following discussion attempts to account for the short-term memory of pitches as a low-level aspect of aural scene analysis and language perception.

**Pitch-location memory: The trace**

I suspect that the ability to remember recent pitch events is an application of a much more general act of auditory scene analysis, i.e. the everyday act of understanding what is happening
When we hear a sound, we know that something out there in the world caused it by performing some action. We want to be able to keep track of that agent and understand what it is doing over time. In order to do that, we need to be able to quickly associate what we just heard with what we will hear next. We also need to be able to separate this agent's sounds from any other sound source that might be in the environment.

Thus, at any given moment we must hold an aural inventory of recent events. As new information comes in, we want to be able to associate it with what we've already heard, if possible -- as Albert Bregman would put it, we want to assign it to an existing auditory stream. We do this on the basis of similarity - if what we are hearing is similar enough to what we've heard, we know that it probably came from the same sound source. In everyday life we frequently use spatial location and timbre to separate sources, but music tends to neutralize these factors in an effort to create a blended object – it presents a large number of timbrally similar sound events from roughly the same location. However, music does intensely utilize the property of pitch, which is another important criterion for associating or dissociating sources. All other things being equal, we assume that sonic events that are identical or proximate in pitch emanate from the same source, but sounds that are disparate in pitch may not.

In example 4.2 I’ve used a short recording of a croaking frog to simulate an environmental stimulus that would engage pitch for the purposes of an aural inventory. The sound file presents the same sample repeated a few times at two different pitch levels. If we were to hear this signal in the wild and attempt to guess how many frogs were croaking, we’d probably say two. Each event leaves a “trace,” a short-term memory of that event tied to a sonic

---

actor. The trace tells you “something is out there, and this is what it is doing.” Future sounds are compared to that trace – if they are similar, they replace the trace with a new one. (We say to ourselves, “The same thing is still out there, but now this is what it is doing.”)

Example 4.2: Simulated environmental stimulus with croaking frogs

Graphical representation

See file ex4-2frogs.wav for audio.

Our hypothetical stimulus with two frogs seems to present two traces which are abstracted from the more complex structure of each croak. Each instance presents a louder, longer low sound followed by a higher sound. (Using the onomatopoeic language we teach to children we might say that the “rib” is primary and the “-bit” is a subordinate continuation.) In addition, the low sound begins with a brief slide up in pitch followed by a flatter stretch – the perceived pitch of the croak is derived from this more stable part of the sound. Thus, we are

---

putting each complex croak into a small-scale event hierarchy and abstracting out a basic pitch level that is prominent in memory. Each distinct sound source (i.e. each frog) seems to be represented on the pitch level of the lower sound (example 4.3) The secondary "-bit" sound also leaves a trace, but it is generally not engaged within this stimulus - with each new croak we use the primary sound to recognize the hypothetical sound source.

Example 4.3: Event-hierarchical interpretation of a single croak, abstraction of the trace
In our frogs example each trace represents a single sonic actor. This is not our only use for short-term pitch memory, however. A more complex stimulus from a single actor can create multiple traces which emanate from substreams within the sound. These persistent pitch locations allow us to track higher-level contours and changes.

We can observe such complexity in the perception of spoken language. The perception of speech usually involves a stimulus that emanates from a single sound source (i.e. one person). Each vowel sound within a phoneme is assigned to a pitch level in the same manner as our frog croaks – the phoneme may scoop or smear up or down but we have a sense of its main pitch. Between words one often has a larger sense of contour – large-scale rises and falls are often used for various rhetorical effects. And, most relevant for our current discussion, pitch events can create lingering traces that are intentionally left and then reengaged in the course of the phrase.

Examples 4.4a and b present fragments from Reverend Martin Luther King, Jr.’s famous “I Have a Dream” speech\(^88\) with a graph of their pitch contour. The phrase we are examining reads as follows:

…and so we’ve come to cash this check, a check that will give us upon demand the riches of freedom and the security of justice.

Each example presents a pitch-contour graph generated by linguistic transcription software called wavesurfer in a top panel,\(^89\) as well as my fixed-pitch transcription and commentary below. The fixed-pitch transcription draws on the variable undulations of the wavesurfer graph for guidance but was further adjusted by ear - an audio file of the entire King excerpt and its

\(^88\) Audio obtained from \url{http://www.americanrhetoric.com/speeches/mlkihaveadream.htm}, accessed 8/13/12.

\(^89\) \url{http://sourceforge.net/projects/wavesurfer/}, accessed 8/13/12.
pitch transcription are also offered as examples 4.4c and d. There are several moments in this phrase where the interplay of pitch registers becomes apparent, as the stimulus breaks into substreams that each have their own trace.

Example 4.4a shows the pitch contour at the beginning of the phrase. The words “come” and “cash” engage a pitch level around C#4, but King finishes the clause by dipping down to a lower pitch and scooping upwards with “this check.” (“This” is around C4, and “check” picks up this lower trace and pulls it upward. As we connect pitch events that are proximate but slightly higher or lower we are also seeing a crude kind of voice-leading -- we'll see below how pitches within a certain narrow range of proximity tend to be perceived as "the same" despite some fluctuation.)

**Example 4.4a: Passage from the "I Have a Dream" speech with pitch transcription**

![Pitch Diagram](image_url)
4.4b: Passage from "I Have a Dream" speech with pitch transcription

See also audio files ex4.4ckingclip.wav and ex4.4dkingtranscription.wav.

The dip down and quick flip upwards on "this check" serves as a kind of half cadence that indicates more is forthcoming, and King continues by returning to his primary C#4 pitch level (on "...check that will...") Thus we see that King employs a primary register that he can diverge from and return to, utilizing its persistence in short-term memory for rhetorical effect.

As the phrase reaches its climax King uses pitch to emphasize and connect key words (example 4.4b). This emphatic register is established around D#4 with "RICH-es", and connects with "FREE-dom" and "se-CU-ri-ty" which are both around C#4. Throughout this passage King also uses a neutral floor register around B3 to intone less important words like "the" and "of" -- interestingly, as the phrase draws to a close the disparate streams seem to merge and the final
emphatic "JUST-ice" is grounded at the lower pitch level. (This is probably necessary to give the sentence a sense of conclusiveness.)

The perception of an underlying collection of pitch locations in music is often more like the analysis of the local details of speech than our aural inventory of frogs, in that a single sound source leaves many discrete pitch traces. The individual parts or voices in most Western music repeatedly traverse a series of fixed pitch locations (i.e. the underlying scale that the passage is based on.) Even though a series of notes are assigned to a single stream (or part or voice), we retain a memory of those pitches. Example 4.5 illustrates a hypothetical melodic passage that descends through several tones and then returns to reengage a tone. Each stepping-stone along the way is far enough apart to be persistent in memory, producing its own pitch-memory trace, and yet the locations are close enough together to create a larger-scale sense of motion that we would call voice-leading.

**Example 4.5: Hypothetical melodic fragment with pitch traces and voice-leading**
Scales consist of steps

Up until this point I’ve used the terms “scale” and “collection” somewhat interchangeably. However, they denote two slightly different concepts. Let us define a collection as any group of pitches that might be presented in a passage – it does not necessarily include a tonic or have any other structural requirements. Scales are a very specific kind of collection. In common musical experience they usually contain a referential pitch class which is conceived as the “bottom” (and “top”) of the scale. And they seem to have one structural constraint – they are made of steps.

We could define steps as being typically the smallest intervals contained in a passage of music. We think of them as such when they are traversed melodically – small intervals within verticals are generally not considered steps. As they occur, steps tend to be the smallest conceivable melodic intervals within their immediate context – they seem to adequately fill or even exhaust the available musical space.

A classic experiment by George Miller and George Heise demonstrated one important property of steps – they create very strong melodic connections. When a melodic line moves relatively slowly, most listeners have no trouble hearing it as a coherent series of sounds from a single source, regardless of the intervals it contains. However, when that motion accelerates to a fairly rapid rate (like sixteenth notes at 150 BPM, 100 ms per tone) perceiving a single line can become more difficult or even impossible – the signal may seem to fragment into multiple

---

streams emanating from multiple sound sources. Such is the case with relatively large intervals, anyway, but small intervals remain connected in a single stream. Miller and Heise referred to the difference between intervals that retain a sense of connection at high speeds and those that don’t the “trill threshold,” and found it to be roughly equivalent to a 15% difference in frequency between tones. This is a musical interval that lies between a major second (which is a 12.2% difference under equal temperament) and a minor third (18.9%). Example 4.6 presents a stimulus that is somewhat similar to the one used in the Trill Threshold experiments, with audio commentary and a video illustration.

Example 4.6: Trill threshold, frame from demonstration video

See files ex4-6trillthreshold.mp3 and ex4-6trillthreshold.mp4 for audio and video demonstration.
Leon van Noorden has examined how the interval required to maintain a coherent stream systematically becomes larger and eventually becomes all-encompassing as an alternating stimulus becomes slower. The stimulus in example 4.6 represents tones 100ms apart, or sixteenth notes at 150 BPM. Van Noorden found that alternating tones 120ms apart (or sixteenths at 120 BPM) could be heard as coherent up to an interval of about a perfect fourth, and notes 140ms apart (sixteenths at 107 BPM) don't split into irreconcilable streams until they reach a major seventh or so. Van Noorden also found that the effect of stream segregation was susceptible to attention - these wider intervals represent the distance at which an alternating stimulus can be heard as unified. However, when subjects were instructed to try to perceive the pulses as separate for as long as possible they found that the trill threshold (Miller and Heise's 15% distance) represented the interval at which the stimulus necessarily became unified.

Miller and Heise's trill threshold, then, represents a distance at which tones will remain connected regardless of speed or attention. We thus have a practical maximum size for scale steps – in Western classical music, at least, steps tend to be a semitone, whole tone, and, occasionally, an augmented second. Since these intervals seem to make maximally strong connections, there is a sense that they adequately fill musical space.

The minor third's marginal status as a step makes some sense when one considers that the coherence of streams exists on a continuum that is more forgiving at slower tempi. Thus it would seem that the anhemitonic pentatonic scale (which includes three whole-tone steps and two minor thirds) and other scales with minor third or augmented-second steps can serve to divide and exhaust musical space in some musical applications. The pentatonic scale also

---

contains a structural similarity to the diatonic collection in that it has only two scale-step sizes, and, unlike the diatonic scale, one step-size is not divisible by the other.

Subsequent research with Miller and Heise's alternating stimuli has revealed a second, smaller interval that is also crucial for melodic perception - the "fission boundary." If one begins with two tones that are virtually identical in pitch one hears this alternation as a single "warbling" tone, similar to the effect of vibrato in everyday music. (We also saw this in our King excerpt above, as phonemes that were slightly different in pitch seemed to be "the same.") As this interval grows larger one can eventually hear an alternation between discrete tones - this is the fission boundary. This zone in which tones are essentially conflated into a single percept has been found to correlate to 25% of the critical band. It thus varies according to register - at C2 it is equivalent to a musical interval of about a whole step but it dwindles to slightly less than a half-step at C6. Example 4.7a presents a video demonstration of the fission boundary and 4.7b lists its intervallic range at C2, C4, and C6, with common musical intervals included for comparison.

---


Example 4.7a: The fission boundary, diagram from demonstration video

See files ex4-7fissionboundary.mp3 and ex4-7fissionboundary.mp4 for audio and video demonstration.

b. Values in different registers

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>whole step</td>
<td>12.2%</td>
</tr>
<tr>
<td>fission boundary at C2</td>
<td>10.51%</td>
</tr>
<tr>
<td>fission boundary at C4</td>
<td>6.39%</td>
</tr>
<tr>
<td>half step</td>
<td>5.95%</td>
</tr>
<tr>
<td>fission boundary at C6</td>
<td>3.87%</td>
</tr>
<tr>
<td>quarter-tone</td>
<td>2.93%</td>
</tr>
</tbody>
</table>
(One might recall that this same distance, 25% of a critical band, has also been cited as the point of maximal roughness between two simultaneous tones, and simultaneous tones at less than this distance also tend to fuse into a single pitch percept.)

Thus, we also have a practical minimum for step size - the semitone would seem to be close to the smallest interval that can maintain a sense of discrete pitches between steps. Intervals smaller than the fission boundary would cause scale degrees to blur together and inhibit memory for a fixed collection of positions. (Though the fission boundary is larger than a half step in low registers, this is not necessarily a problem, as overtones in the middle and high registers would probably aid in the perception of discrete pitches. A motion from, say, C3 to C#3 would include motion between harmonics G4 to G#4, C5 to C#5, E5 to E#5 and so on, well into the region in which the half-step distance is sufficient. A passage in an extremely low register like C2, however, might have a difficult time communicating its pitch collection if it were to move by half step.)

The fission boundary, like auditory roughness, appears to be an effect that arises from the workings of the inner ear, as its intervalllic distance conforms to a fixed distance on the basilar membrane (i.e. the critical band.) Miller and Heise's trill threshold, however, seems to be a higher-level process concerned with the segregation of streams - the most telling difference is that this boundary for temporal coherence expands as the alternation between tones becomes slower, whereas the fission boundary is not dependant on time.

---


95 van Noorden, “Minimum Differences…,” 1041.
The consecutive semitone constraint

Defining a scale as a series of steps imposes two structural limitations. We can assume that tones will be potentially added to a scale until it forms a series of strong connections that exhaust musical space. Also, scales seem to be limited by what Dmitri Tymoczko calls the “consecutive semitone constraint.” Simply put, we do not expect scales to contain two half-steps in a row. Tymoczko argues that by considering all possible sets that contain only steps (which can be defined strictly as half steps and whole steps or broadened slightly to include augmented seconds / minor thirds) and do not contain consecutive semitones, one arrives at the entire collection of commonly observed scalar collections in twentieth century music. The strict half-steps and whole-steps constraint produces the diatonic scale, the melodic minor (or heptatonia secunda), whole tone and octatonic scales. The looser constraint produces the harmonic minor, "harmonic major" and hexatonic collections, which Tymoczko asserts are also all ecologically valid (i.e. appearing in 20th Century classical music and jazz.) (A third restriction, that the generated scale cannot be a subset of other valid scales, excludes pentatonic collections, though these would intuitively seem to be "real" scales that conform to the criteria we are discussing.)

---


97 p. 138.

98 p. 140-142.
In asserting that these are the scales that tend to appear in actual tonal music, there is a certain subjective tautology being applied. Actual music will present collections that do violate the two-semitone constraint, but in analysis we tend to filter out some of these tones as "chromatic embellishments." We might even say that these more dense collections "aren't scalar." For example, Paul Johnson has identified 24 Stravinsky works that feature the “diatonic octad” (set-type 8-23), a collection that could be described as an eight-note segment of the circle of fifths, thus including three semitones in a row.99 This set certainly appears to be a referential collection for Stravinsky, yet in casual observation we might describe such passages as featuring “variable” or “conflicting” scale degrees.

I think that this distinction has to do with the connecting power of the intervals below the trill threshold. I've said that steps tend to create melodic connections that seem both inexorable and sufficient. If we consider a hypothetical three-tone line that descends by semitones (e.g. C, B, B♭), we can see how one tone might tend to be downgraded to chromatic status while the other two are considered essential to the underlying collection (example 4.8.) C-B creates a strong connection between tones, but as we arrive on B♭ we realize that there is also an inexorable and sufficient connection between C and B♭. The three tones will probably be interpreted in an event hierarchy in which the B natural is heard as a mere way-station between more essential tones. It is the strength of the C-B♭ connection that causes this to happen.

Example 4.8: Stepwise connection relegates middle element to chromatic status

There is one possible exception to the successive semitone constraint. If the middle tone in a series of three is established as the most structural, the two flanking pitches could be included in a stable underlying scale. Example 4.9 presents a hypothetical example, in which C and G are immediately established as part of the tonic harmony, and an A♭ and F♯ surround G. Metric placement and relative length also work together to reinforce an event hierarchy in which G is more structural and A♭ and F♯ are subordinate to it.
Example 4.9: A hypothetical passage with successive diatonic semitones

The "deep scale" property

Theorists have long been interested in explaining the presence of one specific collection in much of the world’s music. This is set-type 7-35, often referred to as the diatonic collection, which provides the Medieval modes as well as the major and natural minor scales. Several properties of the collection have been suggested to be particularly advantageous for musical composition, thus making it the best of all possible options. However, while these properties may be perceivable and indeed actually advantageous, they do not necessarily represent constraints on scalar perception - it probably remains possible to successfully compose tonal music based on scales that lack these properties.

Perhaps the most popular observation about the diatonic collection is the "deep scale property," the recognition that every interval in the set appears a unique number of times and
each tone has a unique intervallic relation to all other tones. Imagining all possible combinations of tones within a C major scale, there is one tritone that can be made (B-F), two semitones (B-C and E-F), three major thirds, four minor thirds, five whole tones and six perfect fifths. If one begins at any one tone and considers all other tones, one sees a unique pattern of intervals -- the relationship from C to all other tones includes a whole tone (C-D), major third (C-E), perfect fourth (C-F), perfect fifth (C-G) et cetera, and this pattern taken from any other tone is going to be at least slightly different.

One argument on the advantage to using such a collection is that it is possible to recognize each tone's unique intervallic profile and thus know very explicitly "where we are" in the scale. It is also thought that the two least-common intervals in the collection, the tritone and semitone, are strong indicators of tonic by virtue of their rarity - they narrow down possible background collections and thus indicate a key center. There are some problems with this rare-interval theory of key-finding. Note the conceptual leap in this argument from identifying a particular diatonic collection to finding a tonic. Proponents of this model assume that we assume a diatonic basis and an Ionian modal orientation for a given stimulus. Western listeners may very well have this tendency (which is indeed demonstrable in the psychology lab), but a rare-intervals theory doesn't say anything about how tonal music based on other modal orientations or

---


102 Helen Brown and David Butler, "Diatonic Trichords as Minimal Tonal Cue-Cells," In Theory Only 5/6-7 (1981), 37-55.
other collections could possibly work – indeed, it suggests that these other musics would not be
tonal. Eytan Agmon has pointed out that the tritone is actually not necessary or sufficient for
tonic induction – tonal centers can be created without it and its appearance does not always
define a tonic. \(^{103}\) Thus, there are certainly other causal factors at work.

In general, I'd say that the lack of transpositional symmetry in the diatonic collection is
indeed useful for remaining tonally oriented. And as we'll see in chapter 5, the semitone (the
second-most rare interval) is particularly powerful in defining tonics. However, I doubt that this
advantage translates into a constraint - it seems equally possible to compose tonal music that is
based on collections that lack the deep scale property.

The diatonic collection and consonance

David Huron has also pointed out that the relative consonance of the diatonic collection
makes it very well-suited for composition. \(^{104}\) He employs two different measures of relative
consonance for his argument -- one is "aggregate dyadic consonance" which simply sums
together a consonance/dissonance score for all intervals that can be found within a collection.
(This is a highly generalized concept of consonance and dissonance by interval class that is
conceived as independent of a specific register, much different from our look at acoustic

\(^{103}\) Eytan Agmon, "Tonicity and the Tritone: Beyond the Rarity Issue," in Proceedings of the First International
Conference on Cognitive Musicology, ed. J. Louhivuori and J. Laaksamo (Jyväskylä: Jyväskylä yliopiston
monistuskeskus, 1993), 74-87.

\(^{104}\) David Huron, "Interval-Class Content in Equally-Tempered Pitch-Class Sets: Common Scales Exhibit Optimum
roughness in chapter 2.) By this measure set-type 7-35 scores very high - its maximal preponderance of perfect fifths makes it very consonant overall. (The only collections that score higher are subsets of the diatonic collection - the pentatonic scale (set-class 5-35) and the diatonic hexachord (6-32).)

In an interesting twist on the deep scale and rare interval theories, Huron also observes how well the interval vector of the collection corresponds to the relative consonance and dissonance of intervals in general. A collection scores higher if it contains common consonant intervals and rare dissonant ones. By this measure, the diatonic collection is again one of the best options, though the octochord 8-26 and the harmonic minor collection (7-32) score higher. Huron observes that the sets that score highly by both measures are those that are most commonly observed in world-wide practice.

Again, overall consonance is a property that suggests that composing with the traditional major and minor scales is indeed advantageous, as it allows a composer to create a variety of consonant sonorities. It remains unclear, however, at what point the overall level of dissonance might become a constraint, so that a collection actually becomes too dissonant to be perceivable as the scalar basis of a passage. There is probably is no such point that cannot be explained in other terms – a very dissonant collection would be likely to violate our successive-semitone constraint, for example.

Relative consonance provides an intriguing theory, however, on why subjects might prefer to associate major-mode tones with major triads and minor-mode tones with minor triads. Given a C major triad, subjects in the Krumhansl-Kessler probe-tone studies tended to give an A
or B a higher rating than an A♭ or a B♭. Given a C minor triad, the reverse was true.\textsuperscript{105} We can predict the preferred tone by looking at the implied simultaneity created by the interaction of probe and triad (example 4.10). This approach excludes the triad tones that engage in a voice-leading relation with the probe but includes all remaining notes that might be implicitly sustained. (For example, if a C major triad is followed by an A, we assume that there is a motion from G to A but the remaining tones implicitly continue, thus creating a (C, E, A) simultaneity. For a C major followed by a D we assume that there is motion from both C to D and E to D but the remaining G continues, thus creating an implicit (D, G) simultaneity.) In each case the simultaneity implied by the preferred diatonic tone is more consonant than the less preferred "chromatic" choice.

Implicit simultaneity might explain why subjects thought one tone "goes better" with a triad. However, simple acculturation provides an equally persuasive counterthesis - one can imagine that in a culture where most music was based in the "harmonic major" (e.g. a C major scale with an upper tetrachord of G-A♭-B♭-C) subjects might very well prefer a pairing of C-major and A♭.

\textsuperscript{105} Krumhansl and Kessler, "Tracing the Dynamic Changes…"
Example 4.10: Diatonic expectations and relative consonance

Non-traditional scales and scalar perception: The Ondine experiment

Both the Cohen and Krumhansl-Kessler experiments seem to engage a preference or expectation for the traditional major and minor scales of the Common Practice era. I was eager to test whether the sense of underlying scale was more difficult to observe when the sounding collections were somewhat unusual, so I designed a home experiment based on Ravel's
"Ondine," from *Gaspard de la nuit*. In this work each harmony involves a rapid oscillation between essential tones and decorations, so that each chord is practically a complete collection unto itself, and these collections change very frequently. As in the Cohen experiment I selected brief passages from this work and tested whether it was possible to aurally determine the scales implied therein. I was the sole subject for this experiment.

The first step was to obtain a MIDI recording of the piece and segment it. I wrote a crude piece of software which one could use to create this segmentation by ear, without viewing the actual pitch contents of the file. I chose some segments which seemed to conform to one collection and a few that apparently contained interesting transitions between collections. My choices are outlined in example 4.11. (In order to preserve a somewhat “naïve” perception of the work, it was necessary to choose these chunks quickly and somewhat arbitrarily, without a sustained hearing of the piece. These selections are not meant to represent the best or most musical segmentation of the work, and the resulting scalar responses are not necessarily meant to represent an intelligent analysis of the piece. This exercise is merely a test of whether scalar perception is possible in a somewhat unusual tonal context.)

---

Example 4.11: Segmentation of Ravel's "Ondine," from *Gaspard de la nuit*

- Presents a viable scalar collection.
- "Malformed" (Breaks the successive semitone constraint. Possibly chromatic or multiple scalar collections.)
A second computer program presented these 17 clips in random order and at random transpositions. Example 4.12 shows the user interface for the experiment. As the subject I was free to hit the "stimulus" button as many times as I wished. Next, I needed to select a tonic for the passage by pressing the 12 buttons in a column on the right-hand side of the interface. The buttons sound a tone as they are pressed so that one can find the desired pitch class by ear. After I chose a tonic the "scale" column was made available. This presents 12 chromatic tones that begin with the tonic as the lowermost pitch. One can select or deselect tones for the scale in an on/off action, and again the tones sound as the buttons are pressed. The "play" button below the column presented the selected pitches in ascending order, and I was free to directly compare the stimulus to the ascending scale if necessary. Finally, the "submit scale" button recorded the choice and moved on to the next stimulus.

Example 4.12: Ondine experiment graphical interface
Creating scales by ear that matched the content of the clips proved to be an easy task. Considering only the stimuli which presented what I am considering a single, well-formed scalar collection (with no violations of the successive semitone constraint), 55 out of 61 responses (or 90%) could be deemed completely “correct,” in that they accounted for all the pitch-classes present in the stimulus plus some possible "extra" notes that filled out the scale. Responses for each particular clip were very consistent, with little variation in tonics, and the chosen scales tended to be Dorian, Mixolydian, or a version of the so-called Heptatonia secunda collection (e.g. C, D, E, F, G, A♭, B♭, C.) Complete results are tabulated in example 4.13.

**Example 4.13: Ondine experiment results**

Guide to data

- Pitch contents of stimulus, as spelled in score. Collection begins on the most common specified tonic. Actual stimuli were transposed randomly. Stimulus is described as “malformed” if it violates the successive semitone constraint.
- Responses begin on the specified tonic, transposed back to original pitch level of score. Spelling reflects score and scalar norms — actual responses were enharmonically neutral.
- Comments. Traditional description of scalar response. “Wrong” responses omit pitches from the stimulus.

Stimulus 1 (m. 1) \(\{C^\#, E^\#, G^\#, A\}\)

\(\{C^\#, D^\#, E^\#, F^\#, G^\#, A^\#, B\}\times 6\) (Heptatonia seconda, C♯ major with i6, i7)

Stimulus 2 (m. 3) \(\{C^\#, D^\#, E^\#, G^\#, A\}\)

\(\{C^\#, D^\#, E^\#, F^\#, G^\#, A^\#, B\}\times 4\) (Heptatonia seconda, C♯ major with i6, i7)

\(\{C^\#, D^\#, E^\#, F^\#, G^\#, A^\#, B\}\times 3\) (C♯ Mixolydian — wrong, A♭ instead of A)

\(\{C^\#, D^\#, E, F^\#, G^\#, A^\#, B\}\times 1\) (C♯ Dorian — wrong, E instead of B♭)

Stimulus 3 (mm. 4-5) \(\{C^\#, D^\#, E^\#, G^\#, A\}\)

\(\{C^\#, D^\#, E^\#, F^\#, G^\#, A^\#, B\}\times 9\) (Heptatonia seconda, C♯ major with i6, i7)

\(\{C^\#, D^\#, E, F^\#, G^\#, A^\#, B\}\times 1\) (C♯ Dorian — wrong, E instead of B♭)
Stimulus 4 (m. 5) \{C, D, E, G, A, A, B\} (malformed)
\{C, D, E, F, G, A, B\} x 1 (C major – rejects the A)
\{E, F, G, A, B, C, D\} x 1 (E Dorian – Cx is not in stimulus, rejects C, A)
\{A, B, C, D, E, F, G\} x 1 (A Mixolydian – Cx is not in stimulus, rejects C, A)

Stimulus 5 (mm. 5-6) \{A, B, C, C, D, E, G\} (malformed)
\{A, B, C, D, E, F, G\} x 2 (Heptatonia seconda, A major with ↓, ↑)

Stimulus 6 (m. 6) \{A, B, C, D, E, F\}
\{A, B, C, D, E, F\} x 4 (Heptatonia seconda, A major with ↓, ↑)

Stimulus 7 (mm. 6-7) \{A, B, C, C, E, G\} (malformed)
\{A, B, C, D, E, F, G\} x 4 (Heptatonia seconda, A major with ↓, ↑)

Stimulus 8 (m. 7) \{E, G, A, B\}
\{E, F, G, A, B\} x 3 (E dorian)
\{A, B, C, D, E, F, G\} x 2 (A mixolydian)

Stimulus 9 (m. 8) \{A, B, C, D, E, G\}
\{A, B, C, D, E, F, G\} x 6 (A mixolydian)

Stimulus 10 (mm. 8-9) \{D, E, F, G, A, B\}
\{D, E, F, G, A, B\} x 1 (D dorian, rejects Cx)
\{A, B, C, D, E, F, G\} x 1 (A aeolian, rejects Cx)

Stimulus 11 (m. 9) \{C, D, E, F\}
\{C, D, E, F\} x 1 (C major – rejects A, B)
\{E, F, G, A, B\} x 1 (Heptatonia seconda, E major with ↓, ↑ - rejects F, B)

Stimulus 12 (mm. 9-10) \{D, E, F, G\}
\{D, E, F, G\} x 2 (D dorian – rejects A, B)
\{D, E, F, G, A, B\} x 1 (D mixolydian - rejects F, A, B)
\{G, A, B, C, D, E\} x 1 (G mixolydian - rejects A, B)

Stimulus 13 (m. 11) \{F, A, B\}
\{F, G, A, B, C\} x 4 (F dorian)

Stimulus 14 (mm. 11-12) \{F, A, B, C\}
\{F, G, A, B, C\} x 3 (F dorian)

Stimulus 15 (mm. 12-13) \{C, D, E, F\}
\{C, D, E, F\} x 2 (Heptatonia seconda, C major with ↓, ↑)
\{C, D, E, F, G\} x 1 (C mixolydian – wrong, A instead of A)
\{F, G, A, B, C\} x 1 (F dorian – wrong, E instead of E)

Stimulus 16 \{C, D, E, F\}
\{C, D, E, F\} x 2 (C mixolydian – wrong, A instead of A)
\{G, A, B, C\} x 1 (G dorian – wrong, A instead of A)
With passages that seem to present more than one scalar collection (which I've been calling "malformed") the preferred response was intimately tied to the event hierarchy. Shifts between collections were usually easy to hear, and it seemed most reasonable to match a scale to the main part of the passage.

Example 4.14 shows two such passages in detail. Stimulus #5 presents a fairly simple scenario, as the material from measure 5 yields to a new collection in measure 6. The sustained melodic B♯ presents the longest tone in the passage, giving the latter part of the segment a sense of rhythmic stability and finality, and the move from what is essentially a C♯ major seventh to an A♯ dominant ninth seems to favor the latter sonority as a move "forward." Thus, if one were to draw an event-hierarchical tree of the passage measure 5 would be subordinate to measure 6. Influenced by this end-weighted schema, I thought it made the most sense to select a scalar collection that matched the latter part of the passage, and I was able to produce the same collection both times.

Stimulus #11 presents a brief two-beat snippet with a different collection on each beat. Here the latter beat tended to sound more transitory - the A and B sounded like lower neighbors to A♯ and B♯ and the moving eighth notes in the melody implied a continuation. (Also, my hearing of the clip may have been polluted by exposure to the actual continuation in the piece, in which the B dominant sonority does indeed return to the original harmony on the downbeat of...
measure 10.) For one of the responses it seemed reasonable to select a scale (C♯ major) that matched the initial, more stable part of the passage, and it appears that on another hearing I attempted to create a scale that incorporated both parts (E♯ major with lowered 6 and 7, which includes A♯ as G♭ but still excludes B♭.)

Example 4.14: Influence of event hierarchy on scalar response

Overall, among the 17 responses to mixed-scale stimuli, 13 are arguably "correct" in that they adequately account for part of the stimulus. Three include tones that do not appear in the passage and exclude tones that are present, and are thus probably "wrong." The E♯ major response discussed above appears to include a mixture of tones from the passage. A proper large-scale experiment with multiple subjects could possibly ascertain whether people generally
follow a similar strategy of using event-hierarchical impressions to determine the underlying scale.

The scalar vs. chromatic distinction and musical experience

The theory of tonal hierarchy consigns chromatic tones to the lowest or most remote level of tonal organization, and for the most part this seems to be an adequate description of how they are heard. Generally speaking, a tone that does not belong to the underlying collection of a passage may sound "surprising," "ill-fitting," "distant" or "complex." (Of course, the idea of "chromaticism" also implies that these tones are a positive contribution to the overall effect, adding "color" to otherwise straightforward, unadorned music.) A note's status in the tonal hierarchy will tend to affect our interpretation of event hierarchy - if it does not belong to the underlying collection we are also likely to hear it as subordinate to other surrounding tones.

However, this general description leaves one large question unanswered. Do internalized expectations determine which tones are heard as chromatic, or does the music create this distinction through context? Is our sense of tonal hierarchy shaping our hearing, or is the sounding music shaping the hierarchy? Understanding the tonal hierarchy as the result of simple processes breaks this conundrum. The following discussion will consider common-practice passages that include chromatic tones, and examine how the sense of scale is established, how it feeds back into event hierarchy, and how it is potentially undermined by bottom-up events. (As in the final sections of chapter 3 I will make use of a variety of analytic techniques to examine the event-hierarchical context of specific moments.)
In doing so we must acknowledge three very familiar categories of chromatic event. The first is the one assumed by my description thus far, in which one's sense of tonal hierarchy remains more-or-less constant and stable throughout a passage. However, two other kinds of passages occur frequently in tonal music - instances of mode mixture, which modify the sense of underlying collection without moving the tonic, and modulations, which usually involve both a change in tonic and underlying collection. In these latter categories there is obviously something in the way new tones are presented that causes them to displace existing tones in our perceived underlying scale.

Our first excerpt will be the opening measures of Mozart's String Quartet in G Major, K. 387. The passage is presented in example 4.15, complete with an event-hierarchical interpretation that examines the chromatic passages in detail.
The opening measure of the piece conforms exclusively to the G major scale, and by the downbeat of measure two all seven tones have been presented. If one were to play the passage up to this point I think it would clearly communicate a G tonic and G major collection. It is within this context that our first chromatic tone sounds, a D♭ on the second half of the second downbeat. We know that D♭ conflicts with the existing scale because admitting it to the collection would violate the successive semitone constraint.

The tonal hierarchy model predicts that we will hear D♭ as subordinate to D and E, and that is certainly what happens -- it slides upwards without drawing much attention to itself. We
could say that it is quickly "resolved" by passing to E. Other contextual factors reinforce the impression that D and E are more structural than D♯ - both tones are on the beat, both are chord tones, and the E has an implicit quarter-note length. Considering an alternate version that simply ceases after the D♯ demonstrates how crucial this resolution is. A trailing D♯ eighth-note followed by silence (as in example 4.16) has a completely different character - it now sounds like an upper neighbor to D (or, in other words, an E♭) which at least temporarily displaces E♭.

(Following this truncated passage with some future E♭ would sound like a confirmation that it is the new ♭6, whereas a future E♭ would sound like a return to "the previous ♭6.")

**Example 4.16: Mozart K. 387 excerpt with abrupt cessation after D♯**

![Example 4.16: Mozart K. 387 excerpt with abrupt cessation after D♯](image)

Our expectations may have some influence on how we hear D♯ as it sounds --certainly the D♯ in example 4.16 is surprising. However, it is the immediate continuation to E that makes its subordinate status clear and reaffirms the intended underlying scale.
We can trace this tone's status in the tonal hierarchy through its specific relationships in the event hierarchy. As a chromatic tone D♯ could be considered a fourth-tier tone in the tonal hierarchy. G is the most central pitch class, and the other tonic triad tones (B and D) constitute the second tier of general importance. The other major-scale tones (A, C, E and F♯) occupy the third level. In this passage there is a specific event-hierarchical chain that leads from D♯ all the way back to G - D♯ is subordinate to E, E is ultimately subordinate to the D that arrives on the downbeat of measure four, and that D is subordinate to the final G in the bass.

As the E sounds on beat two, we hear our second chromatic tone in the bass, a G♯. This chromatic event is a bit more prominent than the D♯, but it also resolves clearly to a diatonic tone. The A minor sonority's position at the end of the phrase and its implicit length and metric position "underneath" the suspension figure (i.e. the sense that it really occurs on beat three) all work together to make G♯ subordinate to A. If we pause on this subphrase ending in measure two, there is really no impression that G♯ might belong to the underlying collection. G♯ still sounds like the tonic, and A♯ sounds like scale-degree 2. If G and A are still scalar tones there is simply no room to admit G♯. Since A is a third-tier tone, G♯'s subordinate relation to A also suggests that it is in the furthermost level of the tonal hierarchy.

One might ask, though, why it is that we don't uproot the G tonic and hear a modulation to A minor. After all, we have just heard a tonicizing V⁶ to i in that key and such a move would accommodate the G♯ as diatonic. Knowledge of the actual continuation of the passage is, of course, a strong influence - one might simply know that we will return to G in a mere two measures, and it is easy to auralize this relationship as we are paused in measure two.
However, there is something about this subphrase that seems extremely well-grounded in G major. It is remarkably difficult to imagine any continuation in A even with sustained effort. To further investigate, let us strip away all melody and rhythm and consider an abstract model of this passage, a I, V⁶/ii and ii in simple triads (example 4.17). If we work at it, we can hear these chords as ambiguous, and the deciding factor has to do with event hierarchy and the implied metric relationships that go along with it. If we hear the G major triad as a downbeat and the E⁶ and A minor as continuation, the A sounds like ii (example 4.17a). It remains difficult to imagine some continuing emphasis of A that would create a modulation. However, if we consciously imagine the first two chords to be a large-scale upbeat the balance shifts, and a modulation to A becomes plausible (example 4.17b). In order to achieve this we've changed the event hierarchy - G major is now subordinate to the other chords.

**Example 4.17: Modeling mm. 1-2 as three simple triads**

a. typical hearing, E⁶ and A minor as continuation

\[\text{I} \quad \text{V}^6/\text{ii} \quad \text{ii}\]

b. G and E⁶ as upbeats

\[\text{VII} \quad \text{V}^6 \quad \text{i}\]
In the actual passage the G-major material is so well-established that it is impossible to hear as a kind of auxiliary cadence that leads to A -- G major is privileged by initiating the phrase, by being played *forte*, and by extending via prolongation into the downbeat of measure two. Another event-hierarchical analysis (with some metric normalization to illustrate the relative "weight" of events) appears in example 4.18. The material would have to be radically reworked in order to tip the balance to A. Thus, relatively long-range event-hierarchy (i.e. the continuing influence of a G tonic from five beats ago) trumps the G♯ that sounds in the second measure. This influence of event hierarchy on implicit collection is similar to what we found in the Ondine experiment with the "malformed" stimuli -- the collection that corresponds to the primary part of the stimulus (in this case, the entire two-measure subphrase) seems to be the best representation of the passage as a whole.
Example 4.18: Measures 1-2 with some metric normalization

The next chromatic tone we hear is the C♯ in measure three, and this one has a few interesting properties. It isn't resolved to a third-tier tone like E or A -- it connects directly to a triad tone (D). Thus, there is no immediate contextual cue that this note is particularly remote. C natural (as part of the A minor triad) has been very prominent in measures two and three, but it is not hierarchically grounded in the same way the opening tonic G was. When the excerpt is
truncated after the arrival of D♭ the entire phrase tilts towards the downbeat of measure four, and I think that the arrival of the bass-note D in measure three ultimately dominates its metric span as well (example 4.19). If we pause at the end of this excerpt I think that C♯ could very well stand up as a new diatonic tone that has displaced C♯, and it is easy to imagine a concomitant shift to D tonic.
Example 4.19: Measures 3-4, truncated, with metric normalization
Mozart quickly "undoes" this new tone by passing from D to C♯ to C♭ in measure four, a frequent move in Common-Practice music that communicates that a tonicization of V is only temporary. Now C♯ is clearly subordinate to C♭, due to C♭'s status as part of the V♭.  

The A♯ that sounds on the third beat of measure four is also excluded from scalar status through what are, by now, familiar methods. If we pause at the end of beat three the note sounds like a substitute minor ♯3 with unresolved suspensions underneath (example 4.20). Such a tone would displace B♭, but the proper ♯3 arrives on the fourth beat, thus resolving the tone with a return of the real tonic triad. Metric placement, B♭'s position at the end of the phrase, and the implicit A-A♯-B passing motion all make it clear that the flanking pitches are more structural, and thus A♯ does not belong to the underlying scalar collection. 

Example 4.20: Measures 3-4, truncated at A♯
As we pause abruptly in midstream most of our chromatic tones have created disruptive and disorienting effects. Our sense of underlying collection seems to put up little resistance to admitting them into the scale and allowing them to displace other tones. (The one exception was G♯, which is so outmatched by the central G♭ that a truncated passage still sounds quite stable.) It is mostly what happens after a chromatic tone that seems to matter - if it is resolved quickly to a scalar tone (especially a third-tier one) with superior contextual position it seems to have no effect on the underlying collection. Also subtly important is the event-hierarchical status of the tone that would be displaced, as we saw with G♯ vs. G♭ - if the sounding event is more important than what preceded it may establish a new scalar tone, but if the conflicting tone remains contextually well-established the new tone has no effect.

Let us, then, consider a few passages in which new tones are successful in reshaping the underlying scalar stratum. The opening ten measures of Schubert's "Heidenröslein" present a typical modulation to the dominant, from G major to D major, which introduces C♯ as a new leading-tone to D (m. 6). (Thus, Schubert's C♯ can be compared to the brief C♯-D figure in K. 387, as both are ♯₄ in G.) Example 4.21 presents the entire passage with an event-hierarchical interpretation.
Example 4.21: Schubert's "Heidenröslein," mm. 1-12

Schubert's "Heidenröslein," mm. 1-12
The C♯ and the harmony it belongs to, V♯ of D major, are introduced in a rhythmically suggestive place. The first phrase has unfolded in a very straightforward block of four measures, and a second block has just begun. The A♯ is clearly parallel to the ii♯ of measure two, which branched forward in the event hierarchy - we expect the A♯ to similarly branch forward to something new. (This context is illustrated in example 4.22. Of course, the extreme familiarity of this tune also makes it difficult to imagine anything else happening at this juncture.) The C♯ does not seem to be outweighed by previous events - its position in a new hypermeasure compartmentalizes it away from the C♯'s that preceded it, and it points forward to some unknown quantity.

Example 4.22: The introduction of C♯, hypermetric context
What follows strongly reinforces C♯ as a diatonic tone - there are three more C♯'s and no conflicting C♭'s. Ultimately, Schubert's strategy for modifying the underlying collection seems straightforward - introduce a new tone in a rhythmically suggestive place, repeat it, and don't undercut it.

The Schubert song "Ständchen" offers a similarly straightforward instance of mode mixture. The piece is in D minor but has a propensity to gravitate towards the relative major, quickly cadencing on F in measures 16 and 22. In order to conclude the first stanza on tonic without forsaking the warmth of the major mode Schubert transposes his cadential material to D major in the final phrase (mm. 25-28, see example 4.23). F natural is replaced by F♯ in a deft cross-relation (mm. 23-24, F5 vs. F♯4). Right away this tonic with F♯ is superordinate to the F♯ that sounded over the dominant in the previous measure, giving the new tone a sense of stability and relative permanence. The major-mode material that follows closes out the four-phrase compound period -- it is thus conclusive and structurally "important." In order to remember a passage in the original mode that is equally significant we must think all the way back to the beginning of the song. The new B♭ also sounds within this secure D-major block (m. 25), connecting directly to A as a third-tier tone.
Example 4.23: Schubert's "Ständchen," end of first stanza (mm. 17-28)
After the cadence in the parallel major Schubert must make his way back to the minor mode. He achieves this with a brief postlude to the first stanza (mm. 29-36, example 4.24). This passage extends a long D pedal point, and is mostly in D major, but it begins with a move to the minor iv♭, a brief moment of doubt and unease (m. 29). The B♭ here is certainly locally scalar - it connects directly to a second-tier tone (A) and has no immediate competition from a B♭. Its position at the "fresh" start of a new phrase seems to make it less surprising - I think it registers more as a subtlety than an overt shift. Since it quickly yields to a tonic it also seems to have no lasting influence - the parallel moment with a major subdominant (m. 33) sounds like a similarly gentle shift back to B♭.
Example 4.24: "Ständchen", postlude to first stanza (mm. 27-36)
Overall it appears that our sense of an underlying scale is fairly pliable -- we are
generally willing to replace tones in the underlying scalar collection with new ones as they
sound. In the current discussion we've identified several factors that may determine whether a
tone is heard as scalar or chromatic. Proximity to "conflicting" tones (such as D and E vs. D♯)
can rule out scalar membership, due to the successive semitone constraint. Resolving to a third-
tier tone in the tonal hierarchy may also make chromatic status clearer, whereas tones that
resolve directly to members of the tonic triad may be more likely to achieve temporary diatonic
status. And metric position and other event-hierarchical factors seems to matter - new tones in
strong metric, harmonic, or phraseological positions are more likely to sound structural, whereas
those in weak positions sound temporary.

In general, our accumulated expectations of a particular underlying scale don't seem to
have much influence on our event-hierarchical interpretation. New “outside” tones may sound
surprising, but whether they are admitted into the collection seems to be influenced by the
contextual factors outlined above. Thus, ultimately, there are only a few driving forces that are
scalar, per se -- the structural constraints that scales are to be composed of steps and avoid the
steps-within-steps of successive semitones. These perceptual tendencies combine with the
structural ebb and flow of event hierarchy to define some tones as scalar and others as chromatic.

Because the scalar-chromatic distinction does not seem to actively shape interpretation, it
is somewhat dispensable in a theory of “how tonality works.” Our chapter 6 discussion of brief
dyad-and-monad figures will focus almost exclusively on intervallic rootedness and the
interaction of voice-leading and temporal ordering that I call the “finality effect.”
Chapter 5

Simple Diachronic Relations

The theory of tonal hierarchy tends, by its very nature, to be somewhat divorced from the actual flow of music in time. It asserts that as we experience a piece of music we utilize an internal organizational schema that remains stable and consistent, and the primary goal of theorists such as Krumhansl and Lerdahl is to capture that structure and describe it.

However, tonality is also inextricably bound to time. The seemingly static hierarchy that shapes our perception is strongly influenced by properties that emerge when one event meets the next. The current chapter will discuss a few of these properties that seem essential to the induction of a tonal center. In particular, we will focus on the phenomenon of voice-leading and what I call the “finality effect.”

The hierarchical structure of a single tone

We tend to think of event hierarchy as something that can only happen when two or more sounds occur in sequence. However, our tendency to organize time goes much deeper, all the way down to the single event. In examining the structure of a single tone we can gain an appreciation for why multi-event hierarchization is so ubiquitous.
The typical musical sound has a beginning (the attack), a continuation, and a somewhat indistinct ending (or decay.) In order to experience the event as unified and continuous we must knit each perceived moment into some kind of encoding that is presented to consciousness. One might imagine that such a code could be translated as "new thing (X), X continuing, X continuing and growing quieter" and so on.

This organization is hierarchical in that there is one specific moment in which the sound seems to essentially "happen," and the following moments are contingent upon it. We could draw this relationship with Lerdahl and Jackendoff's right-branching tree notation, as I've done in example 5.1. The subordinate branches in my diagram represent an unquantifiable series of following moments which are not actually discrete - the sound is, after all, continuous.

Example 5.1: Tones create right-branching hierarchy

See files ex5.1rightbranchtones.mp3 and ex5.1rightbranchtones.mp4 for audio and video demonstrations.

---

This structure makes good ecological sense. We know that the sounds we hear in our environment are the result of some action, and it is that action we want to attend to. It may be a sudden, percussive event with reverberation (e.g. a rock hitting another rock) or a decisive action with follow-through (like a footfall in noisy underbrush.) In these cases the attack of the sound represents "the important part" of the event.

These percussive sounds begin with an extremely quick rise in amplitude to an initial peak that we call the attack. The initial ramp-up to attack is often so brief as to be unperceivable. However, there are other sounds in our environment which have softer, more gradual attacks, like a spoken syllable with a soft consonant or the surging and receding of wind. These sounds require a different hierarchical structure. Their beginning is not their peak, and thus there is a sense that the initial increase in intensity leads forward to a high point which is the central moment of the event. Our encoding for this kind of event could be translated as "new thing (X), leading to louder X, leading to essential peak of X” and so on. Such an encoding must lag slightly behind real time, as each following moment supersedes the previous on the way to some unknown climax. We cannot recognize the peak of intensity until it is past. Such a lead-up can be illustrated with Lerdahl and Jackendoff’s left-branching notation, as in example 5.2.
Example 5.2: Soft attacks create left-branching structure

So far we’ve imagined sounds that are continuous and more-or-less featureless. What happens if there are other details that stick out, besides the attack? What if, instead of a simple decline in intensity, the sound continues with a change in pitch, a sudden shift in timbre, or a secondary attack? These situations would seem to call for a more complex representational structure that can account for multiple salient details at different times without sacrificing the unity of the sound – we need a recursive hierarchy.

If our hierarchy is able to encompass multiple layers of organization we can account for any sub-event that may attract our attention and still have a clear sense of how it fits in with the overall whole. Each local detail is knit together with the same kind of incremental encoding we’ve been discussing, but there is a larger structure to which everything belongs. Example 5.3 presents a musical tone with a slight pitch disturbance that occurs after the initial attack, as well as a tone with a secondary surge in amplitude. As the new details begin to emerge each moment ceases to branch back towards its predecessor -- this is no longer mere continuation of the sound.
but “something new.” However, the new sub-event is always connected back to the initial attack – it is a continuation of the complex sound that has already begun.

**Example 5.3: Tones with sub-details**

![Diagram](image.png)

See also ex5.3toneswithsubdetails.mp3 and ex5.3toneswithsubdetails.mp4 for audio and video.

Once we appreciate the structure of individual sounds the near ubiquitous hierarchical ordering of successive events becomes easy to understand. The same processes that organize individual tones can organize a series of tones, a phrase, a series of phrases, and so on. Each part belongs to some greater whole, and the concepts we attach to these structures (note, segment, phrase) need not have strict definitions.

This recursive organization must have some upper bound, of course, which would be the limits of attention and memory. The fine-grainedness of the hierarchy is similarly limited - we might grasp large-scale gestures without perceiving the details that make them up. (Indeed, repeated hearings often reveal small-scale nuances that escape us at first.) By fluidly organizing
the things we do perceive into a hierarchically organized structure our cognitive system makes
the best of what it gets.

**Two events**

The temporal microstructure of a single pitch event seems unlikely to have much
influence on tonality, of course. It is usually only when different pitch classes meet in
succession that a tonal center becomes clear. We can temporarily avoid the complexity of the
entire recursively hierarchical context of music by considering two discrete pitch events as our
basic diachronic unit.

Given two events in close proximity, our perceptual system will continue to apply the
same principles that organized a single tone – it will seek out what seems to be the most crucial
moment and understand the other event relative to it. It will assign one of two basic schemas,
either a right-branching hierarchy (meaning “beginning and continuation”) or a left-branching
one (meaning “precursor and goal.”) (example 5.4).
Example 5.4: Right- and left-branching structures

This decision will have a strong effect on our sense of tonal hierarchy. Given two events and no other information, the event that is judged to be superordinate in the event hierarchy will also contain the tones that are more central to the tonal hierarchy. To keep things simple, we’ll assume for the time being that the stimulus contains a tone that is going to be chosen as tonic – that tone is likely to be in the part of the stimulus that “wins” this juxtaposition of elements.

With a two-element stimulus that is otherwise free of context, our tonal and metric interpretations tend to align. Both categories of interpretation involve the same sort of judgment, after all, a determination that one thing is subordinate to another. We could say that each domain assigns an event hierarchy. Thus, our metric interpretation is often a clue as to our tonal interpretation. Example 5.5 presents two figures that are very similar but tend to have opposite hierarchical interpretations. 5.5a sounds like it is grounded in the first event – the tonic seems to be C, and the following D sounds like an unstable upper neighbor. The second example, however, seems to resolve to its second tone, F, which sounds like the tonic. Our metric sense of these fragments will tend to follow suit – the first one will sound like a downbeat followed by an upbeat, but the second is more likely be heard as an upbeat leading to a downbeat. (I’ll discuss why these two manifestations of set-type [027] behave so differently in chapter 6.) Of course,
we can consciously turn around our metric interpretation of 5.5b so that it sounds like a so-called feminine cadence (i.e. a motion to tonic where the goal arrives on a relatively weak beat), but I think we have a tendency towards the upbeat-oriented hearing.

**Example 5.5: Metric and tonal interpretations tend to align**

In general, we seem to have balancing tendencies that favor the first or the second event in our basic diachronic unit. I’m going to call the factors that favor the first event the “beginners’ advantage” and the factor that favors the second the “finality effect.”
Beginners’ advantage

The first event would seem to have a higher likelihood of being selected as hierarchically superordinate because it has actually set the perceptual process in motion and created our first impressions of the overall sound. If what follows remains consistent with that first impression, we will tend to assign the "beginning - continuation" schema to the overall event. Given two sounds with the same pitch content and equal intensity, a right branch would seem to be the default interpretation (example 5.6). Like the figure in example 5.5b, however, these two isolated, identical sounds remain highly ambiguous – if one is motivated it is trivial to hear the same figure as an upbeat followed by a downbeat.

Example 5.6: Right-branching hierarchy as default interpretation

<table>
<thead>
<tr>
<th>stimulus</th>
<th>interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>event hierarchy</td>
</tr>
<tr>
<td></td>
<td>meter</td>
</tr>
</tbody>
</table>
Arpeggiation

Melodic intervals that sound like arpeggiations also tend to be right branching. In these cases there is an implicit continuation of the first tone into the second event — we could say that the first event leaves a “trace,” a presence that lingers even though the tone has ceased to sound.\textsuperscript{108} If the first event presented only one tone, and the second event involves an implicit dyad, why isn’t the latter judged to be more important? The answer seems to be a form of the beginners’ advantage — the first tone has initiated an implicit harmony, and the second one is heard as a mere continuation of it (example 5.7). Interestingly, this will tend to happen even if the second tone adds information about the harmony that we did not anticipate with the first tone. Upon hearing the initial G of example 5.7 one would have no reason to expect a lower E to sound next — it is perhaps mildly surprising. Yet it still seems completely unproblematic that the E is a continuation of the harmony that began with the first tone. Once again, however, these intervals are metrically ambiguous and easy to intentionally reheat as an upbeat-downbeat figure.

Example 5.7: Arpeggiation tends to be right-branching

If the interval being arpeggiated is rooted it will have a strong influence on the implicit tonal center. Example 5.8 presents figures that outline major thirds, perfect fifths, and their inversions. Most of these are right-branching and downbeat-oriented, as we’d expect. However, if the intervallic root lies in the second event there may be an incentive to hear the figure as left-branching. This is especially true with $\hat{5}$-$\hat{1}$ figures, which engage Western cultural conventions and thus sound like a bass motion from dominant to tonic rather than the continuation of a single harmony.

(Also, for this listener the ascending minor sixth does not seem to behave like its inversional equivalent. The second tone sounds quite unstable, an upper neighbor to an anticipated perfect fifth. This is probably due to this interval’s somewhat weak rootedness and associations with musical contexts such as the prelude to *Tristan und Isolde*.)
Example 5.8: Arpeggiations of rooted intervals

likely right-branching interpretations

\[ \hat{5} \hat{1} \]

\[ \text{as left-branching} \]

implied C: V I

upbeat orientation

a somewhat complicated hearing of the ascending minor 6th

stimulus

implied hierarchy
The power of voice leading and the finality effect

Up to this point I’ve been vague about which intervals will tend to sound like arpeggations. Any interval will do so, as long as the two tones do not engage in a voice-leading relationship. Voice-leading relationships, on the other hand, rule out any sense of implied simultaneity – they are the opposite of arpeggiation.

We learned in chapter 4 that any melodic interval within Miller and Heise’s trill threshold will tend to create an ineluctable sense of connection and motion from the first pitch location to the second. Thus, major and minor seconds tend to create these relationships. They are relatively dynamic and disruptive, because the melodic motion that they create indicates that something in the sound source has changed.

This sense of connection can also occur in pitch-class space as well as pitch space, as sevenths and ninths create the impression of stepwise motion. We’ve discussed how the assimilation of many perceived overtones into one unified pitch percept can be understood as an axial pattern on a spiral-shaped pitch continuum. While pitches a seventh or ninth apart traverse a much greater frequency distance than those a second apart, they involve a similar rotation of this axial pitch-finding scheme (example 5.9.) It makes sense that such a shift would be associated with melodic connectedness, as “actual” connectedness in pitch space involves the same relationship.
Example 5.9: Pitch-space voice-leading as a rotation of axes

Voice-leading in itself might seem like a fairly neutral influence on hierarchy and tonality. Essentially what it is is motion, and depending on context that motion can be towards something structural (thus creating a left branch, as in example 5.10a) or away from it (as in 5.10b).
Example 5.10: Voice-leading as structurally neutral motion

a. motion towards structural tone
b. motion away from structural tone

However, when presented in a two-element stimulus that is followed by silence, stepwise motion has a decisive impact on our perceived tonal hierarchy. I call this phenomenon the “finality effect.” As the first note yields to the second it does not leave a “trace,” an implicit presence after it has ceased to sound. Instead, its trace is obliterated and replaced by one in a new pitch location. This second trace, however, is allowed to exist as the stimulus yields to silence – one could argue that the second tone thereby receives an implied agogic accent of indefinite length that makes it seem like the much more stable and “important” part of the two-event unit (example 5.11).
Example 5.11: Lingering trace creates agogic accent, finality effect

Example 5.12 presents every possible two-note stepwise configuration. All of these figures would be left-branching, and indicate a second tone that is more tonally central than the first. I think that most of them tend to suggest a tonic, except for the descending semitone. The descending semitone is not an idiomatic motion to tonic in Western music, and for that reason it seems more likely to be heard as a motion from $\hat{6}$ to $\hat{5}$ in the minor mode (or, perhaps, $\hat{4}$ to $\hat{3}$ in major.) These other scenarios still create a structural contrast between tones and lead us to an unheard tonic.
Example 5.12: All stepwise relationships and their hierarchical effect

The strong influence of the final state is somewhat dependent on our two tones being followed by silence. If our two-note unit is followed by more material (as one would expect in a typical, recursively hierarchical context) the trace of the second tone might not continue to exist, and the influence of the final state would be negated. This is what happens if we simply string together a repeated stepwise pairing – the trace of the second event is replaced in the stepwise return back to the first. Thus, with no finality effect we are left with a beginners advantage – the first tone will emerge as superordinate as it is prolonged from unit to unit (example 5.13).
Example 5.13: Repeated stepwise motion is neutral

However, if the trace of the second event is not displaced by subsequent events its influence can be decisive. This phenomenon is nicely demonstrated in an experiment designed by Jamshed Bharucha.\textsuperscript{109} Bharucha created short melodic figures from the tones of the B major and C major triads. Notes from each triad were joined in semitonal pairs, and each stimulus presented either a random permutation of the three ascending pairs (e.g. F#-G, B-C, D#-E) or the three descending pairs (e.g. E-D#, C-B, G-F#). These figures were played twice, followed by either a B major simultaneity or a C major one, and subjects were asked which chord “fit better” with the figure.

Bharucha’s figures seem to communicate a B or C triad by combining a small-scale finality effect and a larger-scale arpeggiation. The latter tone of each pairing is allowed to continue into the next unit, joining with the latter tone of the next pair in an implicit simultaneity (example 5.14). These figures are remarkably resilient – one can hear them in different metrical orientations (beginning on either an upbeat or a downbeat), add dynamic accents to the initial tones of each pair, or even stretch the rhythmic values so that the first tones receive an agogic

accent and one will still have the same hearing. (These alterations did not appear in Bharucha’s experiment – there the stimuli were presented with a strictly regular rhythm and amplitude.)

Example 5.14: Bharucha’s figures and the accumulation of traces

Bharucha’s subjects who reported some musical training were able to select the predicted triad 66% of the time. Untrained listeners, however, did no better than chance (50%). It was thought that the artificiality of the stimuli, which were computer-generated “Shepard tones” with short gaps between each note, was partly responsible for driving down accuracy – certainly a casual hearing of the figures in example 5.14 suggests that they do communicate a clear and unambiguous harmony.

The finality effect is a crucial structural force that cannot be appreciated by contemplating the materials of tonal music in the abstract. The relationship of, say, C to B would seem to be perfectly symmetrical – each tone is a half step from the other, and this interrelationship suggests little about their potential to create a tonal center. However, if we exploit the asymmetry of time either tone can be made superior to the other, and this relationship will in turn suggest a hierarchical orientation that has implications for all twelve pitch classes.
Theorists are, of course, generally invested in the fact that tones tend to make significant stepwise connections, and it is of particular interest that we can often anticipate these connections before they occur. Non-chord tones within harmonies tend to resolve by step to a chord tone, and tones outside of the tonic triad often seem to impart a palpable expectation that they will connect to a tone in the tonic harmony. However, the perceived attraction of a tone to a future position is an effect that can only occur after a tonal center is clearly established. What’s fascinating about the Bharucha experiment discussed above is that the same stepwise figures can create tonality, by retroactively suggesting that one pitch is subordinate to another.

While melodic expectation is indeed interesting, it seems possible that we may sometimes put the cart before the horse by assuming that stepwise motions are a response to projected tonal expectations rather than a direct cause of tonal perceptions. This is frequently the case with local harmonies that are elaborated with non-chord tones – musicians are generally taught that in the majority of instances non-chord tones “must” resolve by step, with the implication that this is the result of an expectation that is created as soon as the dissonance is sounded. The reality of the situation is often that failure to resolve a non-chord tone threatens to create an unintended tonal effect, altering the intended harmony or otherwise rendering it unclear. Resolving the tone by step indicates its subordinate status and clarifies which pitches “belong” to the underlying sonority – this is often a retroactive process rather than a projective one.

We will have an opportunity to observe the structure-making power of voice-leading motions in the following chapter, which examines the tonal properties in a collection of brief harmonic and melodic fragments. In general, the beginners’ advantage, finality effect, and the disruptive role of voice-leading will all have a crucial influence on how each stimulus is heard.
Chapter 6

The Dyad Plus Monad Experiment

In previous chapters we’ve examined several primitive perceptual processes that contribute essential elements to our experience of tonal music. With the experiment described in this section we bring these properties together and observe how they interact. The execution of this somewhat unorthodox home experiment actually predates most of the research and theory that has been presented up to this point – it was a crucial source of discovery that helped to establish the elements required for an adequate account of tonality.

The experimental design originated with a fairly naïve conception of how harmonies form and what our experience of them is like. I wanted to establish principles for the formation of coherent underlying sonorities in an arpeggiating or otherwise “broken” texture. Example 6.1 presents such a passage from the literature, the first 14 measures of the Presto from Bach’s Partita for Solo Violin in B minor. The harmonic analysis indicated below the staff is William Rothstein’s, and he in turn cites composed-out accompaniments by Schumann and Brahms as precedents.111

---

I think that most listeners who are familiar with western classical music would agree that the Presto delivers the same sort of harmonic richness that a more “full” passage would offer, without resorting to any simultaneity of tones. Victor Zuckerkandl has written a beautiful description of this phenomenon, asserting that “the chord is not in the tones but somehow above or behind or about them, a radiation, an aura, one further step removed from materiality.”\textsuperscript{112}

I originally designed the dyad plus monad experiment in an attempt to derive rules for this sort of harmonic formation. I conceived of the dyad as the main harmonic material in the experiment and the subsequent note as a “probe tone.” I expected that the initial vertical interval would create some sense of underlying sonority that would either include or exclude the following tone, and thus the task of evaluating the stimuli would entail a simple binary judgment of whether the tone “belongs to” the harmony or not. Example 6.2 illustrates this plan with a pair of figures.

Example 6.2: Original conception of the experiment

\[ \begin{align*}
\text{E “belongs”, therefore} & \quad \text{is a harmony.} \\
\text{F# does not belong, therefore} & \quad \text{is not a harmony.}
\end{align*} \]

However, I quickly realized that these categories of “belongingness” and “exclusion” were an inadequate description of what was happening. Example 6.3 presents a few more possibilities that quickly became apparent. As I expected, some notes did “belong” with the initial dyad and create an arpeggiative continuation from it (example 6.3a), and some tones acted as non-chord tones that seemed to demand resolution back to a more stable position (i.e. they were “excluded,” example 6.3b). Others, however, appeared to be more stable than the tones that preceded – they were the resolution, and it was one or both of the previous tones that were somehow excluded (example 6.3c). While some figures implied a single harmony, others seemed to be a miniature progression of two harmonies (example 6.3d), or even imply an additional, future point of arrival (example 6.3e). It was clear that these figures were highly suggestive and versatile fragments of the full musical texture, and it would take more than a simple binary judgment to adequately describe them.
Example 6.3: Interpretations of the dyad + monad stimuli

a) arpeggiation

b) non-chord tone, left-branching

c) non-chord tone, resolution

d) two harmonies

e) implied third event

I thus built up a computerized interface and set out to consider a randomized pool of these stimuli, recording whatever properties might be salient, without any overarching theoretical agenda. In a way I had stumbled into a process of Husserlian “eidetic variation.”¹¹³ As a founding father of the philosophical discipline of Phenomenology, Husserl endeavored to redefine the foundations of our knowledge of the world in terms of human experience. He recommended that we rebuild the sciences through a process of strict empiricism, starting from a point of extreme skepticism. This skeptical stance, known as the epoché, requires one to initially set aside what one knows about a subject at hand. Then, one selects an area of inquiry and

considers every possibility within that domain, gradually observing the variant and invariant properties within it and thus deriving “essences.” This is the process of eidetic variation.

The problem with the epoché is one of beginning, of defining the area of study. In music it seems difficult to investigate any general notion (like the domain of “tonality”) without evoking a priori definitions such as “chord” or “scale” or “harmony” – this would seem to be the exact sort of received knowledge we are supposed to resist. However, by stumbling into the dyad-plus-monad experiment through a basic misconstrual of what I might observe, I felt I had unwittingly achieved this goal and begun the process. The pool of 275 stimuli certainly seems arbitrary enough, with many unusual figures that one would never offer as good examples of “tonal” or “harmonic” phenomena.

It must be admitted that as a laboratory experiment, the results of this study remain provisional and vulnerable to criticism. While I am fairly confident that I lacked strong preconceived notions of how tonality operated in these figures, the possibility of pre-existing biases cannot be entirely discounted. It remains possible that my observations are idiosyncratic or strongly reflect the influence of formal theoretical training.

Thus, the discussion that follows should be considered as part of a pilot study for a more robust and scientific set of trials. I’ll attempt some analyses of the data in order to demonstrate the influence of intervallic rootedness, melodic anchoring, and other factors, and I’ll discuss a possible predictive model of tonal perception. I’ll discuss variants of the experiment that would be much more practical to execute with a large body of subjects. Ultimately the value of the current chapter rests not on the data I’ve collected but on the intuitive persuasiveness of the resulting theory.
Designing the experiment

The pool of stimuli consists of 275 combinations of dyads and single tones. The dyads range in size from a vertical semitone up to a major seventh. Each dyad is combined with an additional tone which may lie anywhere from a major seventh below the lower note to a major seventh above the upper note. Any added tones that create octave duplications or unisons with the dyad are excluded. Example 6.4 presents a schematic representation of the full set of stimuli.

Example 6.4: Generating the set of 275 stimuli

As a result we have every possible combination of three intervals (excluding octaves and unisons) that fits within a two-octave range and avoids gaps of more than a major seventh. The restriction on gaps was meant to mimic conventional standards of upper-voice harmonic writing,
and it was thought that the three tones might thus consistently retain the potential to form a blended sonic object. The exclusion of pitch-class-duplicating intervals was made on the grounds that such stimuli were not expected to produce interesting harmonic effects – in the original “belongingness” paradigm, at least, one would expect all redundant tones to “belong” or to create an continuation of the harmony implied by the original dyad. In making these restrictions we lose the ability to observe some potentially interesting effects (e.g. that of doubling and register on intervallic rootedness), but gain from a slight simplification in the pool of stimuli and a narrowing of the scope of the investigation.

I designed a computerized interface to present stimuli and record responses. The stimuli were selected at random from the full pool of 275 until each figure had been heard three times. They were transposed at random but restricted so that they fit within a fixed registral range of C4 to C6. The dyads and monads were played using a default Windows MIDI piano sound for one second each (or quarter note = 60) with no sustain. The interface allowed me to rehear each figure as many times as desired before recording my observations. The trial was done incrementally in a series of sessions of varying length, over a period of several weeks.
The user interface includes a spatial arrangement of checkboxes which mimic the intervallic distance of the stimulus tones in pitch-space. These checkboxes can be clicked to indicate voice-leading connections between tones (or “streams”) and the implied continuation of a tone into a subsequent event (or a “trace.”) Example 6.5 shows this part of the display, in which a trace has been indicated, represented by the dotted arc extending from the topmost box. In this image an additional implied tone has been selected (by a method I’ll discuss below), and it is represented by a checkbox surrounded by parentheses. A voice-leading connection from the lower dyad tone and this new implied tone is represented by the straight, solid line between checkboxes.

Such an elaborate visual representation seemed necessary in order to record complex observations about the stimuli with confidence that the intended details would be attributed to the correct tones. However, it does have the disadvantage of potentially suggesting properties and relationships through visual cues, rather than through hearing alone. With myself as subject, I did not feel that this was a problem, because it was usually easy to identify the intervallic content of the figures by ear. Thus, the aural presentation seemed more explicit than the on-screen interface. A subject might, however, be tempted to rely on the visual cues to guess as to what response was appropriate. Such unintended consequences might be mitigated by a more abstract
layout that can still represent higher and lower pitches but otherwise obscures intervallic distance.

**Example 6.6:**

Chromatic pitch-selector buttons.

A vertical column of 24 buttons numbered 0-23 from bottom to top offers another means of entering information (example 6.6). These buttons act as a chromatic keyboard of sorts, allowing the user to select specific pitches as implied tones and to indicate the implied tonic. The buttons always correspond to the pitches from C4 to C6, and they sound a tone as they are pressed. A black dot also appears next to the buttons to indicate the last key pressed.

Thus, the buttons acted as a fixed pitch reference. The stimuli were transposed at random, however, so for a subject like myself who lacks absolute pitch the process of selecting a tone required a “hunt and peck” technique in which undesired notes are sounded until one eventually arrives at the desired one. Since using the interface in this way involves sounding more pitches, the process can suggest relationships that are not necessarily already auralized by the subject — one might accidentally select a tonic or implied tone and subsequently decide that it is compelling. An interface that could take input that is sung might be less suggestive, but ultimately allowing the subject to generate any sound whatsoever still allows for some degree of experimentation and accidental discovery. I do think that a process of discovery is appropriate for this experiment -- as long as the subject arrives at interpretations he or she finds plausible the data is, in a sense, authentic. One expects that most figures will have a
limited number of likely interpretations, and that the subject will select from this set of possibilities.

Once implied tones are indicated they are essentially added to the stimulus – they appear in the checkbox representation and can optionally be heard in the playback. Subjects can also create a third event that follows the dyad and monad, representing an implied continuation or resolution. The interface also allows one to assign a root or tonic to any of the three events. One can even assign more than one root to the figure, with a non-tonic root on an earlier event and the ultimate tonic on a later one. However, in practice I found that this rarely seemed appropriate – I entered multiple roots for less than 1% of stimuli.

A third type of input allows a user to select from different instances of Lerdahl and Jackendoff’s hierarchical tree notation to indicate the structural priority and “flow” of the different elements in the stimulus. These are presented in examples 6.7. A collection of two-branched trees accounts for items that do not imply a third event -- one can indicate if the dyad and monad figure implies a progression away from something stable (or a “right-branch,” example 6.7a), a progression towards stability (a left-branch, example 6.7b), an arpeggiation with beginning and continuation (right-branching tree with dot, example 6.7c), or an arpeggiation that seems to lead towards its second event (left-branch with dot, example 6.7d).
Example 6.7: Two-event tree interface

As I proceeded with the experiment it initially seemed as though the combination of these two-element trees with the data on implied third events would be sufficient to imply a multi-level, three-element hierarchy. In many cases this remains true. The third element typically fulfills one of two different functions – either it furnishes a return to the initial sonority and creates a prolongational structure (example 6.8a), or it provides a new point of resolution that was not in the first two events, making the dyad-and-monad figure some sort of non-tonic harmonic material (example 6.8b and c). It is easy to combine the chosen two-element tree with a higher level structure, either a right-branching arpeggiation (for prolongations) or a left-branching progression (for non-tonic resolutions).\textsuperscript{114}

\textsuperscript{114} One might notice that I’ve conflated the categories of prolongation (which Lerdahl and Jackendoff indicate with a white dot) and arpeggiation (indicated with a black dot.) I do not think that we need this distinction in the current chapter.
Example 6.8: Extending the two-event trees to account for three events

a. 
\[
\begin{array}{ccc}
\text{right branch} & + & \text{prolongational return} \\
\text{arpeggiation} & + & \text{interpreted as non-tonic (with resolution)} \\
\text{left branch} & + & \text{interpreted as non-tonic (with resolution)} \\
\end{array}
\]

b. 
\[
\begin{array}{ccc}
\text{right branch} & + & \text{higher-level arpeggiation} \\
\text{arpeggiation} & + & \text{interpreted as non-tonic (with resolution)} \\
\text{left branch} & + & \text{higher-level left branch} \\
\end{array}
\]

c. 
\[
\begin{array}{ccc}
\text{right branch} & + & \text{higher-level arpeggiation} \\
\text{arpeggiation} & + & \text{interpreted as non-tonic (with resolution)} \\
\text{left branch} & + & \text{higher-level left branch} \\
\end{array}
\]

(There are a few cases, however, in which simply adding a third branch to the original indicated tree would be incorrect according to the Lerdahl and Jackendoff theory. Example 6.9 presents two such situations. The right-branching \{G, B\} + A figure in example 6.9a combines with a resolution to a future implied tonic. It could be argued that the intervening monad A should branch forward, following the larger hierarchical flow from initial dyad to the point of
resolution. Also, in 6.9b the intervallic proximity of the F# monad to its implied G resolution creates a grouping condition that would probably cause the F# to branch forward. These exceptions have little consequence for our investigation – we can regard it as significant that the monad remains at the lowest level of structure, inferior to the initial dyad regardless of whether it branches forward or backward.

Example 6.9: Three-event trees that disrupt the two-event structure.

As the experiment progressed I did design more buttons that indicated some of these multi-level, three-event trees, and the data is a mixture of two-event trees for proper two-event interpretations, two-event trees combined with implied third events, and three-event trees. A better interface would allow for the spontaneous creation of any possible multi-level hierarchical
structure. As it stands the data is reasonably complete, but it necessitated an undesirable layer of post-hoc analysis to interpret third events and thus codify the actual structures indicated.

**Presenting the data**

As I discuss specific interpretations of figures I’ll use some consistent notational conventions which are illustrated in example 6.10. The “stimulus number” will appear to the upper left of the notated figure – this numbering scheme follows the generative sequence illustrated in example 6.4, above, and can be used as a token of reference. Each figure will be notated so that the chosen tonic is C, even though in the actual trials the stimuli were transposed at random. Register will be manipulated so that the figure fits comfortably on the staff. Reported voice-leading connections will be indicated with solid lines, implied continuations with dotted arcs, and implied tones with parentheses. Finally, the number of times this particular interpretation was entered into the interface will be indicated to the lower left of the excerpt – an indication of (3/3) means that the interpretation was consistent across 3 hearings of the stimulus, whereas (1/3) would mean that I only chose this interpretation once and there were two other hearings which differed in some way.
Example 6.10: Presenting a specific hearing of a stimulus.

FINDINGS

Stepwise motion vs. arpeggiation

In general, we can group our stimuli into two categories – those that are heard as arpeggiation and those that imply some kind of progression or motion. This opposition is built in to the very definition of the terms – an arpeggiation is the impression of an underlying continuity or stability, that the sounding tones are all “part of the same thing.” Progression, on the other hand, involves a transition from one thing to another. The main factor that seems to determine whether a stimulus will be heard as an arpeggiation or progression is the perception of stepwise motion. Steps and arpeggiation are mutually exclusive. 41% of our stimuli were judged to contain voice-leading connections between the dyad and the monad, and 30% of our
stimuli were judged to be arpeggations. I indicated both an arpeggiation and a step only 4 times out of 826 hearings, for an overlap of a mere half percent.

There are a few ways that the perception of stepwise connection can be elicited. The typical voice-leading relationship involves a literal, horizontal major or minor second. We could call these “direct” connections, or “pitch-space” connections. Notes that are a seventh or ninth (or perhaps even a fourteenth or sixteenth) apart occasionally sound like they are connected as well – we may have the impression that, despite the shift in register, these intervals still essentially involve the movement of a voice up or down by a single increment. We might call this “pitch-class voice leading” and imagine that the connection is happening in “pitch-class space.” The computerized interface allows one to indicate pitch-class voice leading in two different ways – one can connect notes across registers, or one can create an implied octave duplication of one of the tones and indicate a connection to that. In the data for all pitch-class voice-leading connections I preferred to mark cross-registral connections over the implied octave duplications at a rate of about 2 to 1. In virtually all cases I only indicated stepwise connections for members of interval class one or two.\textsuperscript{115}

Example 6.11 presents a sampling of reports involving horizontal pitch intervals 10, 11, 13 and 14. Also included (as 6.11i) are our only figures which include pitch interval 22. While responses to the 6.11i stimuli indicated some cross-registral connections, the sample is so small (and the figures are so unusual) that it is probably unwise to conclude much from these.

\textsuperscript{115} There was exactly one instance in which I entered a voice leading connection for a member of interval class 3, or, more specifically, a major sixth. This may have been the result of confusion, as the sixth overlapped with a minor seventh in the figure, \{B4, C5\} + D4.
Example 6.11: Figures that evoked pitch-class voice-leading connections

a) Pitch-interval +14 (Ascending major 9th)

Connection to octave-displaced implied tone.

“Direct” pitch-class connection.

b) Pitch interval +13 (Ascending minor 9th)

c) Pitch-interval +11 (Ascending major 7th)

d) Pitch-interval +10 (Ascending minor 7th)
e) Pitch-interval -10 (Descending minor 7th)

f) Pitch-interval -11 (Descending major 7th)

g) Pitch-interval -13 (Descending minor 9th)

h) Pitch-interval -14 (Descending major 9th)

i) Pitch-intervals +22 and -22 (Minor 14ths)
Example 6.12 tallies the frequency of marked connections in every stimulus that contains a horizontal interval of interval class one or two, and we can see that the members of this category vary somewhat in their typical connecting strength. The black bars emanating from the left indicate the frequency with which I marked these specific intervals with some kind of voice-leading relationship (including relationships to an octave-displaced implied note.) The lighter bars on the right indicate the percentage of figures that were marked as arpeggiations (which, I’ve argued, are the “opposite” of stepwise motion.) All figures that are not arpeggiations are still progressions of some sort, so the white space in the center of the graph represents figures that were interpreted with Lerdahl and Jackendoff’s right and left-branching trees (sans arpeggiation dot) but did not receive any voice-leading indications.
Example 6.12: Frequency of marked voice-leading connections

Considering the chart we see that while all members of interval class one and two have the power to create voice-leading connections, some intervals are stronger or more salient than others. The semitone appears to be the strongest possible connection, which is consistent with our discussion of the trill threshold in chapter 4. The literal pitch-space intervals of one or two semitones are stronger than all their octave displacements, and among the wider intervals the descending ones seem more likely to imply stepwise motion than their ascending equivalents.
One can also see the superior voice-leading strength of interval class one over two throughout all of the intervals – the black bars showing explicitly marked connections are perhaps a bit inconclusive (though the overall averages still show a trend, with 34% for intervals 11 and 13 and 27% for 12 and 14), but there is a clearer difference in the frequency of arpeggiations, as interval class one creates very few (4% overall) but IC 2 is a little more conducive to being heard as part of a single unified sonority (with 14%).

As I noted above, every stimulus that is not judged to be an arpeggiation is some kind of “progression” – either a motion between implied harmonies or a linear motion that resolves a perceived non-chord tone. If 30% of the stimuli were arpeggiations, that leaves the remaining 70% as progressions (example 6.13). Most of these progressions do contain horizontal instances of interval classes one and two. Yet, voice-leading connections were only noted about 41% of the time.
Example 6.13: Summary of voice-leading in arpeggiations and progressions

The disparity between what is contained in the stimuli and what is marked is consistent with the aim of the experiment, which was to indicate perceptually salient characteristics. In rehearing the figures, some connections do seem to be underreported – once one is conscious that
a voice-leading interval is present it is usually easy to intentionally hear a stepwise motion from one tone to another, and the fact that the figures are marked as progressions also suggests the underlying causal effect of steps. In the course of the trials I could have systematically tried to detect all of these intervals and mark them thoroughly – I consciously resisted such an approach, choosing to mark motions that seemed particularly overt. Thus, the lower scores for certain intervals reflects a lower level of salience, but not necessarily a lack of voice-leading effect. I would expect a similar pattern to emerge from a trial with a pool of subjects who have no theoretical motivation to mark all ICs 1 and 2 as steps but are merely instructed to detect a sense of motion or connection.

Since I’ve emphasized a causal connection between progression and steps, one might be curious as to nature of the progressions without steps, which constitute 14% of the overall responses. These figures can be categorized into two groups – those with an IC 1 or 2 in the initial, vertical dyad, and some triadic set-types that were not interpreted as arpeggiations. Samples of these hearings are reproduced in example 6.14.

The dissonances in the initial vertical dyad seem to have implied some instability or conflict that needed to be resolved, as in a suspension-like figure, or marked the initial sonority as a non-tonic harmony that needed to progress to something else. I’ve provided two instances of these figures in examples 6.14 a-b. The G-F dyad of 6.14a implied a specific tone of resolution in the second event. (Therefore, one could argue that the progression does involve a voice-leading connection after all.) The G-A dyad in 6.14b marks the initial dyad as distinct from the solo Í that follows but does not seem to imply linear motion.
There were also many instances in which triadic set-types (especially the more unusual ones, the diminished and augmented triads) were interpreted in two distinct parts (examples 6-14c-d). This disjunction seems to be created when the first dyadic element has a strong harmonic implication (i.e. of a major or minor triad) and the second element is not consistent with it. Again, there does seem to be a stepwise relation between an implicit tone created by the initial event (a G, scale-degree 5 in both 6.14c and d) and the second event. In the discussion that follows I’ll refer to the relationship between the triadic implications of the first event and the second event as “harmonic compatibility.”

**Example 6.14: Some progressions that lack voice-leading intervals (ICs 1 and 2)**

a. initial dyad leads to implied resolution
b. dyad is subordinate, no resolution chosen

c. set class 3-12 (augmented triad) interpreted as major harmony with NCT
d. set class 3-10 (diminished triad) interpreted as minor harmony with NCT
One additional phenomenon that drives down the marked voice-leading scores for individual intervals is intervallic collisions, when two members of IC 1 or 2 connect to the following monad. When the two potential voice-leading connections approach the tone from opposite directions, this is not necessarily a problem (example 6.15a), but when the intervals approach from the same direction, one tends to mask the other (example 6.15b). Of all the stimuli with a unidirectional voice-leading interval conflict, 24% have both of the connections marked, 42% have only one of the two intervals marked, and 34% are totally unmarked. For bidirectional figures, however, 61% have both intervals marked and the remaining 39% have one indicated voice-leading connection.

**Example 6.15: Voice-leading interval collisions**

a. bidirectional conflict  

![Example 6.15a](image1)  

b. unidirectional conflicts  

![Example 6.15b](image2, image3, image4)
Voice-leading and the “finality effect”

In our chapter 5 I discussed the tendency of the second of two events to emerge as hierarchically superordinate when connected by strong voice-leading intervals (and followed by silence, as in our test figures.) I dubbed this phenomenon the “finality effect.” We can see the overall influence of the finality effect by looking at the correlation between the intervallic content of our stimuli and the indicated hierarchical structure.

Within each figure there are three intervals that we can observe (example 6.16). There is one initial vertical interval that constitutes the first event and then two horizontal intervals that connect the tones of the dyad to the subsequent monad. These horizontal intervals are created by the second event and are essentially part of it. If the presence of a particular interval in the first position is matched with a right-branching hierarchical structure (example 6.16b), we can say that the interval “wins” in the event-hierarchy (or, rather, emerges as hierarchically superordinate). If a horizontal interval corresponds to a left-branching hierarchy (which points to the second event, example 6.16c), it too can be described as winning. The winning percentage of an interval (or the tendency of the hierarchical interpretation to align with its position in the stimulus) is noteworthy because the hierarchically superordinate element tends to have an influence on the overall tonal orientation of the figure.
Example 6.16: Intervallic winners and losers

It should be noted that, in one sense, arpeggiation can be considered “ties,” in that no interval has demonstrated a decisive influence on the hierarchical structure. (While they do generally receive a right-branching tree, this is in my opinion a default “beginning-continuation” schema that is applicable to the vast majority of instances. We would need other contextual factors that are not available in this experimental design to create an arpeggio that has a “precursor-arrival” structure.) However, arpeggiation are also “wins” in the sense that all intervals have an opportunity to influence the tonal orientation. In the discussion that follows we’ll consider arpeggiation to be a third category that is distinct from wins and losses.

We can observe the finality effect by considering the winning percentages of the various interval classes, segregating them by interval position (example 6.17). The rooted intervals (interval classes 4 and 5) have a significant effect on event hierarchy – they have very strong winning percentages when they appear in the initial dyad and (for reasons I will discuss below) slightly weaker effects when included in the horizontals. Our voice-leading intervals, ICs 1 and 2, are the most asymmetrical on the chart -- as a vertical in the initial dyad they have a negative effect on the winning percentage, but as a voice-leading interval to the second event they have
very strong results. Interval class 1 in particular has both the weakest and strongest scores of all. This strong showing in the second column is the “finality effect” in action.

Example 6.17: Winning percentage for the various interval classes

<table>
<thead>
<tr>
<th>Initial Dyad</th>
<th>Horizontal Interval</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>11%</td>
<td>18%</td>
</tr>
<tr>
<td>5</td>
<td>52%</td>
<td>44%</td>
</tr>
<tr>
<td>4</td>
<td>51%</td>
<td>26%</td>
</tr>
<tr>
<td>3</td>
<td>36%</td>
<td>26%</td>
</tr>
<tr>
<td>2</td>
<td>13%</td>
<td>46%</td>
</tr>
<tr>
<td>1</td>
<td>7%</td>
<td>53%</td>
</tr>
</tbody>
</table>

Example 6.18 presents a selection of hearings which seem to be influenced entirely by voice-leading and the finality effect, determining both the event-hierarchical tree and the choice of overall tonic. Acoustically rooted intervals (which are our other major tonal influence) are, for the most part, not present. Even the essential Western cadential motion from 2 and 7 to 1 (example 6.18a) could be argued to be a wholly linear and rhythmic phenomenon, as it involves an unrooted minor third or major sixth resolving by step. The figures in example 6.18c do contain a perfect fourth and fifth, respectively, but these intervals seem to “lose out” to the finality effect.
Example 6.18: Selected figures influenced by the “finality effect”

a. typical cadential formulae

\[
\begin{align*}
\text{#51} & \quad (3/3) \\
\text{#196} & \quad (2/3) \\
\text{#20} & \quad (1/3)
\end{align*}
\]

b. chromatic convergence

\[
\begin{align*}
\text{#107} & \quad (1/3) \\
\text{#145} & \quad (1/3)
\end{align*}
\]

c. finality effect trumps rooted interval

Our results also show a definite influence of intervallic rootedness on the event hierarchy and the choice of tonic. As I argued in chapter 3, certain intervals found within the harmonic series tend to point to one tone as hierarchically central. We expect such an interval to behave consistently whether it is “right side-up” or “inverted” – a perfect fifth will point to its lower note, but when it is reconfigured into a perfect fourth it should point to its upper note.

We can track the influence of intervallic rootedness by again considering every instance of each interval, tabulating the relationship of the chosen tonic relative to the interval’s position.
in the stimulus. As an example of how we will track this, let us consider a perfect fifth that appears as an initial dyad. Let’s assume that (as will often be the case) the chosen tonic matches the bottom note of this fifth. Regardless of the actual transposition of the stimulus, we’ll normalize the fifth as $[0, 7]$ and call the tonic 0. However, if the chosen tonic does not align with the bottom note, but instead is a whole step above it, we would judge it relative to the fifth and call it 2. We’ll use mod 12 pitch-class integers, so a tonic a half-step below the bottom note will be tabulated as an 11, and so on.

Since rooted intervals and their inversions behave differently, we will track mod 12 pitch-class intervals rather than interval classes. We would consider an initial perfect fourth, for example, as $[0, 5]$, and we’ll expect a significantly different result in our tabulation of tonics. Here the most common result is likely to be the top tone, or 5.

Speaking more formally, we might call the lower note of the initial dyad $x$, the upper note $y$, and the following monad $z$. We’ll count the first vertical interval as $y-x$ and the tonic as $\text{mod12}(\text{tonic}-x)$. This gives us the results we’ve discussed thus far.

In evaluating our horizontal intervals one unusual twist is in order. Normally one might expect to take the “directed interval” from the dyadic tones to the following monad. So, for instance, a fifth down from the lower tone $x$ would normally be negative 7, which would then be converted to its mod 12 equivalent of 5. (Formally, we’d be taking $\text{mod12}(z-x)$.) However, this falling fifth tends to function in the same way as the vertical fifth – the lower note points to a tonic. In order to achieve a parallel tabulation of tonics, we’ll have to take the inverse directed interval – to count it “backwards” from the monad to the dyad tones. The formal definition would be $\text{mod12}(x-z)$. The tonic will also be calculated relative to the new note, $\text{mod12}(\text{tonic}-z)$. 
All of these calculations are demonstrated in example 6.19. Each figure will create results for the three different intervals contained therein. This particular stimulus has an initial perfect fourth (or interval 5), and the tonic lies a step below the bottom note for a result of 10. It’s also an instance of inverse directed interval 2 (in the whole step down from $x$ to $z$) with a relative tonic of 0, and an inverse directed interval of 7 (in the perfect fifth down from $y$ to $z$), also with a relative tonic of 0. As we survey the aggregate of all these tonic-to-interval relationships we will be able to see which intervals have a strong influence over the resultant tonic and which ones seem to end up in a random jumble of configurations. Example 6.20 presents the tonics that are combined with every interval, with the results separated by position.

**Example 6.19**

a. sample figure with three interval / tonic combinations
Example 6.20: Roots chosen with the various intervals.

b. method of calculation

**Initial dyad**

- Vertical interval 1

**Horizontal with monad**

- Inverse directed interval 1 (half step down and octave equivalents)

- Vertical interval 2

- Inverse directed interval 2 (whole step down and octave equivalents)
vertical interval 3
inverse directed interval 3
(minor third down and octave equivalents)

vertical interval 4
inverse directed interval 4
(major third down and octave equivalents)

vertical interval 5
inverse directed interval 5
(perfect fourth down and octave equivalents)

directed interval 6
(tritone and octave equivalents)
vertical interval 7
inverse directed interval 7
(perfect fifth down and octave equivalents)

vertical interval 8
inverse directed interval 8
(major third up and octave equivalents)

vertical interval 9
inverse directed interval 9
(minor third up and octave equivalents)

vertical interval 10
inverse directed interval 10
(whole step up and octave equivalents)
In general, we see a significant influence of the intervals 4, 5, 7, and 8 (i.e. the perfect fifth, major third, and their inversions.) The perfect fifth is by far the strongest tonic-defining interval, inducing a key center that aligns with its root 72% of the time overall. Its inversion, the perfect fourth, points to a tonic on its upper note 55% of the time. The major third also has a significant but weaker effect, defining the tonic 39% of the time (49% when in the first vertical) and 33% of the time when inverted. (The overall average for any tonic position is a mere 8%.)

Other intervals show a weaker influence on the tonic and can probably be considered unrooted. As was discussed above, the interval classes 1 and 2 show a marked asymmetry that is due to the finality effect – as horizontal intervals they are key-defining, but in the first position their effect is close to the mean and barely favors one tone over the other. Interestingly, an initial vertical of interval 10 exhibits the opposite effect that we might expect from an interval that appears in the overtone series – its top tone emerges as root 24% of the time, but the bottom tone performs at the level of chance with a mere 8%. In my opinion this is persuasive evidence that the minor seventh is, in fact, not acoustically rooted. The tritone is certainly unrooted, with its tendency to suggest a tonic that is not one of its tones. (A root of 1, or the tone above the
bottom-most tone, was the most common result, representing the typical 7-1 resolution in Western classical music.)

The only interval class left to discuss is [3], or the minor third and its inversion. According to our theory of acoustically-based rootedness, this interval is not rooted. It does have a tendency to favor its lower tone as tonic over the upper one – the lower tone of interval 3 was selected 29% of the time (as opposed to 11% for the upper tone) and the major sixth also favored its top tone 28% to 11%. However, there is a strongly competing interpretation of the interval as the upper tones of a major triad – this interpretation yields a tonic of 8 for the minor third (23% overall) and a tonic of 5 for the major sixth (also 23%). When the minor third appears as the initial vertical, this alternate configuration actually beats the would-be rooted one. No other interval has such a strongly competitive alternate key center. It seems likely that these two competing interpretations are the result of projecting common triadic configurations onto the third and not some deeper acoustically-based tonic-making force. I’ll discuss the idea of “triadic projection” more below.

**Rooted intervals and “winning percentage”**

We can also consider the influence of the intervals by considering their winning percentage. I’ve already asserted that the event that dominates the event hierarchy is also likely to define the overall tonal hierarchy. However, in the case of rooted intervals the tonal hierarchy and event hierarchy are in a sort of circular relationship – tonal hierarchy can actually precede event hierarchy and shape the structural flow of events.
We can unpack this circular relationship by first reviewing the “finality effect” and the cases where the event hierarchy does indeed drive the tonal hierarchy. Example 6.21a presents a two-note figure that seems to be influenced entirely by voice-leading and the sense that the second element supplants the first. Here it seems clear that there is nothing about C that would make it superior to B other than the influence of voice-leading and the temporal configuration of tones – reversing the order gives the opposite effect.

However, a rooted interval points to a tonic regardless of which tone appears first (example 6.21b). As soon as the second tone sounds it is clear that the two are in this specific relationship, that one is a central tone and the other is subordinate. The second tone is characterized immediately, at its moment of attack, and the first tone is characterized in retrospect, at its moment of termination.

**Example 6.21: Finality effect vs. rooted interval effect**

a. finality effect is dependent on temporal order

```
<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>C</td>
</tr>
<tr>
<td>C</td>
<td>C</td>
</tr>
</tbody>
</table>
```

```
<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>C</td>
</tr>
<tr>
<td>C</td>
<td>B</td>
</tr>
</tbody>
</table>
```

b. rooted interval defines tonic regardless of order

```
<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>C</td>
</tr>
<tr>
<td>C</td>
<td>C</td>
</tr>
</tbody>
</table>
```

```
<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>C</td>
</tr>
<tr>
<td>C</td>
<td>G</td>
</tr>
</tbody>
</table>
```

```
<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>C</td>
</tr>
<tr>
<td>C</td>
<td>B</td>
</tr>
</tbody>
</table>
```

```
<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>C</td>
</tr>
<tr>
<td>C</td>
<td>E</td>
</tr>
</tbody>
</table>
```
Since this tonic-subordinate role is established instantaneously, it has the potential to define our sense of structural flow in the figure at hand. Example 6.22a presents a figure in which the perfect fifth appears first, as a vertical. As the second tone sounds, it is likely to be heard as tonally subordinate to the initial tonal center and thus a motion away from a more stable point. It will thus receive a right-branching hierarchical tree, and the sense of hierarchical flow and the perceived tonal hierarchy will be consistent.

Example 6.22: Intervallic rootedness precedes event hierarchy

a. initial dyad is rooted interval, defines tonic

Example 6.22b presents a figure in which the first event presents tonally neutral material and the second event creates a falling fifth. Here the second tone is likely to be instantaneously characterized as tonic and be interpreted as a motion toward a goal. It will receive a left-branching hierarchical tree.
Example 6.22 (cont.)

b. second event makes rooted interval, defines tonic

Thus, a rooted interval is likely to have a relatively high winning percentage as well as a strong influence on tonic. We’ve already summarized winning percentage by interval class in example 6.17, where we could see that interval classes 4 and 5 did relatively well. IC 5 was the strongest factor on event hierarchy across the board, “winning” 46% of the time, and when IC 4 was in the initial dyad it was as strong as IC 5 – both interval classes won about 50% of the time in the first position. (Remember that arpeggiations are considered ties, and thus there are three possible outcomes. Anything above 33% shows a significant effect.)

While these intervals are powerful as initial dyads, their effect is largely diminished as horizontals leading to the second event. Example 6.23 again summarizes the winning percentages of interval classes [4] and [5], separating the “right-side up” and inverted forms. We can see that only the perfect fifth wins frequently as both a vertical and a horizontal.
Example 6.23: Winning percentages for rooted intervals

This seems due to two factors – inverted intervals cannot “win” as horizontals, and interval 4 is frequently overwhelmed by the interval in the first position. The reason the two inverted intervals (the perfect fourth and minor sixth) perform poorly in the second position is simple, given the discussion above – if they do manage to influence the overall tonal orientation, they will point back to the initial event as tonally superordinate rather than emphasizing the new event. There are figures in which interval 5 helps to clarify and extend an arpeggiation (example 6.24a), and it can “win” in the second position by establishing a 5 (example 6.24b), but it is also frequently trumped by other tonal factors (example 6.24c).
Example 6.24: Various effects of interval 5

a. perfect fourth clarifies, extends arpeggiation

The poor performance of interval 4 is somewhat surprising, given that a new tone creating a descending major third should point to the second event as tonally important.

However, the major third appears to have difficulty competing with roots that are established in
the initial dyad – it cannot beat a vertical fifth (example 6.25a), an initial major third (6.25b), and it even has trouble displacing an orientation established by a minor third (6.25c).

Example 6.25: Interval 4 in second position fails to establish root.

a. Interval 4 (C - A♭) is trumped by initial 7 (C - G).

b. Interval 4 (C - A♭) is trumped by initial 4 (C - E).

c. Interval 4 (G - D♭) is trumped by initial orientation (implicit C - G fifth?)

Intervals + roots + position + hierarchical status

However, there is one measure by which horizontal thirds show some tonic-making power, namely root choice cross-referenced with interval position and event-hierarchical status.

We have already looked at the roots that tend to be chosen with each interval in example 6.20, and though some intervals showed very strong tendencies toward certain outcomes, these results were surrounded by a large number of other roots with very low percentages, as if there were a certain level of randomness or “noise” in the data. This noise is due to competition between the three intervals in the stimulus – as we’ve seen in the examples above, virtually any interval in any position can be trumped by another. From the point of view of the losing interval these results seem chaotic and random.
Considering only “winners” will filter out some of these undesirable effects. As we see in example 6.26, when interval 4 does succeed in creating a progression towards the second event it consistently points to its lower note for a result of root 0. In general, the three columns represent an increase in entropy from left to right – the first column represents the influence of a single dyad, two intervals compete to define the second events represented in the second column, and all three intervals are free to influence the result of arpeggiations, which are represented in the right-most column.

Example 6.26: Intervals associated with various roots, separated by position and event-hierarchical status
vertical interval 3, winners only (winning % is .38)

inverse directed int 3 (minor third down) winners only (winning % is .27)

int 3 either position arpeggations only (arpeggiation % is .45)

vertical interval 4, winners only (winning % is .54)

inverse directed int 4 (major third down) winners only (winning % is .26)

int 4 either position arpeggations only (arpeggiation % is .36)

vertical interval 5, winners only (winning % is .49)

inverse directed int 5 (perfect fourth down) winners only (winning % is .30)

int 5 either position arpeggations only (arpeggiation % is .35)

vertical interval 6, winners only (winning % is .11)

inverse directed int 6 (tritone) winners only (winning % is .18)

int 6 either position arpeggations only (arpeggiation % is .32)
vertical interval 7, winners only
(winning % is .54)

inverse directed int 7
(perfect fifth down)
winners only
(winning % is .57)

int 7 either position
arpeggiations only
(arpeggiation % is .31)

vertical interval 8, winners only
(winning % is .49)

inverse directed int 8
(major third down)
winners only
(winning % is .26)

int 8 either position
arpeggiations only
(arpeggiation % is .45)

vertical interval 9, winners only
(winning % is .35)

inverse directed int 9
(minor third up)
winners only
(winning % is .25)

int 9 either position
arpeggiations only
(arpeggiation % is .45)

vertical interval 10, winners only
(winning % is .13)

inverse directed int 10
(whole step up)
winners only
(winning % is .45)

int 10 either position
arpeggiations only
(arpeggiation % is .16)
With these sorted results we see clear-cut trends for the rooted interval classes [4] and [5]. In “root position” (as 4 and 7) the intervals have a pronounced effect in all three columns (save, perhaps, for the major third when it appears in arpeggations – there the root only has a slight edge over other possibilities). The inverted forms (5 and 8) are influential as initial dyads, but they cannot sway the results in the second column. (When these inverted intervals do win in the second event, they actually tend to favor the “wrong” note.)

We also see a strong indication of the “finality effect” with voice-leading intervals, as 1, 2, 10 and 11 all have high winning percentages in the second column and the majority of instances point to the second event as tonic (root 0).

Other tonal projections

So far our most strongly influential intervals are either acoustically rooted or acting as voice-leading connections, and they point fairly consistently to one particular tone as a potential tonic. However, it must be noted that there are other common interval-root associations in the
data that cannot be directly attributed to these causes. For example, the minor third is frequently
associated with its lower tone as tonic, and yet there is no theory that can account for this in
terms of rootedness. It is also just as frequently heard as the upper part of a major triad, even
when no tonic tone is explicitly present. How can we explain these cases?

I would argue that many of these instances are the results of triadic projection, in which
the third evokes a more elaborate tonal context due to culturally ingrained memory and
association. Such an effect may not be directly attributable to the basic acoustic property of
intervallic rootedness, but, as I’ll argue below, it remains dependent on the rooted fifth as an a
priori source of structuring and internal coherence.

To understand how triadic projection works, it might help to think of the process of
memory, recognition, and association in general. Some cognitive theorists think that our general
concepts are stored in memory as networks that connect various properties into a somewhat
consistent configuration.\footnote{\(116\) e.g. John R. Anderson, "A Spreading Activation Theory of Memory," \textit{Journal of Verbal Learning and Verbal Behavior} 22 (1983), 261-95.} If we think of the general concept of an apple, we think of a limited
collection of shapes, colors, tastes, and textures which are all interconnected. We know other
things about them as well, that they have stems, that they are healthy, that you can buy them in
bushels, that they are seasonal to the summer and fall – all of these facts are constituted in a web
of neurons. The recognition of any one of these aspects is thought to stimulate the other nodes in
the network, making the wider context potentially available to consciousness. (It has been
demonstrated in timed recognition tasks, for example, that once a particular word is processed related words become “primed” and will tend to be recognized faster.\textsuperscript{117}

It is perhaps uncontroversial to suggest that the memory of a melodic fragment (or, in this case, a single interval) will tend to involve an encoding of the harmonic and metric context in which it is heard.\textsuperscript{118} One may be able to abstract the figure away from its original context, but it is also likely that recall from memory will also evoke these original properties. Our extremely simple and ambiguous stimuli appear to have the power to evoke harmonies and progressions that have been heard before, even in the absence of the strong acoustic cue of rootedness. In the data these associations appear to be less strong and less consistent than those evoked by the rooted intervals, and can thus be considered a secondary force driving tonal perception, subordinate to rootedness and the anchoring effects of voice-leading.

In addition, it can be argued that these associations are actually constructed out of the rooted intervals. This is apparent in instances in which the minor third (or its inversion, the major sixth) is heard as the upper tones of a major triad, despite the absence of a sounded root (example 6.27). Exposure to repeated hearings of actual major triads with all three tones seems necessary to develop this musical memory and engage this interpretation. Once it is engaged, the tones are characterized by their relationship to an implicit $\hat{I}$. Thus, even though the intervals are not sounded there is a real, auralized perfect fifth and major third relationship which makes the percept stable and tonally focused.


Example 6.27: Minor third evokes missing root.

The three figures in example 6.27 present instances in which the members of interval class [3] were interpreted as 3 and 5 relative to an absent 1. The first two figures actually include a rooted interval class 4 in the second part of the stimulus (E-A♭ in 6.27a, E♭-G in 6.27b), but the triadic projection is apparently a stronger influence on the overall interpretation. Example 6.27c presents a minor-seventh fragment that is perhaps unlikely to be heard as a tonic sonority. In
these hearings the horizontally unfolding minor third seems to have evoked the tonic, even though the overall sonority requires some resolution before it is completely stable.

**Non-tonic interpretations**

Up to this point I’ve discussed our stimuli as though they always evoke a tonic harmony. We might expect a voice-leading interval, rooted interval, or other tonal projection to point to a Í which is either explicitly or implicitly part of the overall sonority presented by the figure. However, our experimental design allows for the indication of future tonics that lie outside of the sounding harmony – such choices were usually accompanied by the entry of implied third events of resolution. Example 6.28 presents a sampling of non-tonic hearings that were entered in the course of the experiment.

**Example 6.28: Some figures with non-tonic resolutions**

a. major triad
Because I was interested in resisting the influence of received tonal theory, my experimental design did not include any way to indicate the traditional harmonic functions of tonic, subdominant, and dominant. The only data that indicate a non-tonic hearing are the selected future tonic tones and the third implied events of resolution. Further complicating the picture is the fact that not all third events indicate a non-tonic function – they can also indicate an
implied prolongational structure with a return to the tonic (after a brief departure in the second event.)

I thus had to engage in a post-hoc review of the data to determine whether each hearing appeared to represent a grounded, tonic interpretation or a non-tonic interpretation. I defined these concepts in fairly broad terms. A tonic hearing includes the tonic tone, either explicitly or implicitly, and the overall sonority appears to be rooted on that 1. A non-tonic harmony either excludes the tonic altogether (as in a dominant sonority) or appears to be rooted on another tone, as is the case with subdominants, submediants, and the like. In these latter cases another root seems to be “undermining” the tonic. We are thus conflating together all of the traditional functional categories into two catch-all definitions of tonic and non-tonic. In examining the data I also felt a need to create a third subcategory, the “dissonant tonic” in which the tonic tone was present and grounded but other tones appeared to demand resolution – a suspension figure, for example, might create a dissonant tonic.

In my post-hoc review I thought that 30% of responses indicated a non-tonic interpretation, and 6.4% were dissonant tonics. Non-tonic interpretations seem normal in contexts like ours where a brief figure is followed by silence – that space allows a subject to easily imagine a future tone of resolution, and thus even the major triads (which we might expect to be solidly grounded on a 1) were occasionally heard as dominants.

Some figures, however, were always heard as non-tonic. We can consider at least three possible underlying causes that might place a sonority rather decisively in this category. The first, most traditional explanation would probably be a relatively high level of dissonance. However, I do not think we have definitive evidence of such a trend. Example 6.29 summarizes
the extent to which the various intervals are associated with non-tonic interpretations. (All of the intervals that participate in arpeggiation are counted, as well as those that “win” in the two-position event hierarchy. “Losing” intervals presumably have less influence on the choice of tonic and are excluded.)

Example 6.29: Intervals and non-tonic interpretations

In looking at the initial, vertical dyad, there is generally a higher rate of non-tonic interpretations for intervals traditionally considered dissonant (the seconds, sevenths and tritone) than with the traditionally consonant ones. However, intensity of dissonance does not seem to correlate with the frequency of non-tonic function - the tritone is by far the most frequently non-tonic, whereas the seconds, which are generally more intensely dissonant, have a lower rate.
Minor thirds and their inversions, on the other hand, are almost as frequently non-tonic as
seconds. One might explain these inconsistencies by considering the secondary effect of
“masking,” in which one tone of a simultaneous major or minor second emerges as more salient
than the other. Masking might be causing figures with initial seconds to be heard more as
dissonant tonics (as one tone emerges as “structural” and the other is heard as a mere irritant),
thus suppressing non-tonic percentages that might otherwise be robust.

In considering the horizontal intervals created between the dyad and monad, there is even
less of a clear correlation. Again, the tritone emerges as the strongest nontonic force.
Secondmost, however, is the minor third and its inversion. In this context our other traditionally
dissonant intervals (the seconds and sevenths) are again being processed differently, here serving
as voice-leading intervals which frequently point to tonics.

A second explanation of non-tonicity would be the influence of tonal convention and
remembered tonal context. After all, figures which could be viewed as subsets of the dominant
seventh chord (including the diminished triads) did have a high tendency to be judged as
dominants. The tritone in particular appears to have functioned as a very strong marker of
dominant function. One could simply argue that the frequency of dominant-seventh-to-tonic
resolutions in Western classical music creates an association between the tritone and its
conventional resolutions.

However, another underlying cause of this strong association between certain intervals
and non-tonic function might be a lack of intervallic rootedness. As I’ve argued, a rooted
sonority tends to point to a putative tonic. Lacking such a strong cue, however, it becomes easier
to imagine some “outside” tonic to which the currently sounding tones will resolve. In our first
column of example 6.29 there is actually a clear division between the rooted intervals and all others, unlike the somewhat muddled distinction between consonant and dissonant intervals. In the second column, however, which considers the effects of horizontals, the results are less conclusive, which could again be attributed to the competing effect of voice-leading.

**Predictive Model**

In looking at such a large pool of stimuli we’ve gained very specific information about how certain figures can be heard. We’ve identified a few forces that have an influence on the data (i.e. intervallic rootedness, voice-leading, and other acculturated interval-root associations), and these properties seem to interact in a complex way. How can we combine this information into a more generally applicable model, one that we could possibly apply to a wider variety of stimuli? I think the answer lies in using the statistical probabilities derived from experimental data channeled through a theory of structure.

As I discussed in my opening chapter, the investigation of tonality though statistical analysis is hardly new, and remains very influential throughout the field of music psychology. In general, the effect of key is thought to be caused by the relative distribution of tones in a passage, with the more central tones of tonic and dominant appearing most frequently and the chromatic tones appearing the least. Applying statistical methods to musical stimuli is perhaps attractive because it feels objective – a suppositionless tool is applied to a body of works, it identifies a trend, and the trend proves to be somewhat predictive, therefore the link between tone distribution and tonal hierarchy appears to be true in a way that the somewhat vague and perhaps
untestable assertions of traditional harmonic theory are not. However, many music theorists find this approach to be unsatisfying, because it observes essentially unordered collections of tones.

In our direct experience of music the ordering of tones and the specific intervallic relations between them seem to have consequences. It seems to matter if a sequence of events is arbitrarily reversed, for example, or if one specific tone is substituted for another. And certainly, if our minds are keeping a tally of the most frequent tones and interpreting the stimulus accordingly, we cannot directly observe it doing such a thing.\textsuperscript{119} It simply doesn’t feel like this is how tonality actually works.

Rather than relying on statistical analysis it is also possible to construct what I’m going to call a “purely algorithmic” model of tonality. In doing so one might observe a pool of stimuli, intuit the forces that seem to be at work, and construct an algorithmic model that quantifies and combines certain structural properties in a way that produces consistent, desirable results. My model of intervallic rootedness from chapter 3 is such an algorithm, combining an arbitrary logarithmic registral weighting with an intervallic template of values. Here one is perhaps asserting that the offered numeric values are substituting for some real, as-of-yet-undiscovered neural structure that could be similarly observed and quantified. However, the somewhat artificial nature of such a model can be frustrating, and it is not always clear how the asserted scaling and algorithmic interaction could possibly be implemented in the brain. There is also a problem in drawing sensible conclusions from the numerical results – in my chapter 3 discussion, for example, I needed to define yet another arbitrary value, a threshold for what

\textsuperscript{119} Of course, this is not necessarily a problem for theorists of mind who hold that some perceptual processes are inaccessible to consciousness, e.g. Jerry A. Fodor, \textit{The Modularity of Mind}. (Cambridge, MA: MIT Press, 1983).
constitutes a clear and decisive result as opposed to an ambiguous one. Thus, as with a purely statistical model, a purely algorithmic model can seem inauthentic.

The solution seems to be to combine the real-life measurements of statistics with intuitions about the structure of musical experience. We can combine percentages derived from data into an algorithm with multiple interacting components that have been derived from general observation – in presenting such a model it is clear that each percentage stands for a cognitive process that behaves a certain way (i.e. makes a certain judgment a certain amount of the time) and balances with other forces with a relative strength or weakness which can again be expressed in percentages. The predictions of such a model can then be compared to the actual data and evaluated for goodness of fit.

Example 6.30 presents a flowchart for a predictive model of the dyad-and-monad experiment. Given a figure, it attempts to find the most likely perceived root and hierarchical structure. Since many of our figures in the experiment elicited more than one interpretation, it considers multiple possibilities and evaluates their relative likelihood.
Example 6.30: The predictive model

**PHASE ONE: ARPEGGIATION OR PROGRESSION?**

```
stimulus
  ↓
  yes
  ↓
  does it have voice-leading intervals?
    ↓
    yes
    ↓
    are the three tones “harmonically compatible”?% Arp % Prog
    no
    ↓
    no
```

**PHASE TWO: BRANCHING + ROOT PREDICTIONS**

```
arpeggiation
  ↓
root predictions from list of interval-root associations
  ↓
left
  ↓
left- or right-branching?
right
  ↓
predictions combined, top 3 are normalized to 100%
```

The first determination we must make is the probability that the figure will be heard as an arpeggiation or some kind of progression. The most crucial determinant for this seems to be the presence of voice-leading intervals, so this is the first property observed. If voice-leading intervals are present, we evaluate the likelihood of a progression or arpeggiation based on values derived from the data. The voice-leading intervals are grouped into categories with similar
percentages, as shown in example 6.31. The four groupings have a theoretical logic as well, with semitone connections being the strongest progression-making force, whole-tone connections the next strongest, then, for the most part, all octave-displaced semitones as next-strongest and all octave-displaced whole tones as the weakest connections. (The descending major ninth, interval -14, crosses categories and is grouped with the octave-displaced semitones. This could be altered for theoretical consistency but, possibly, with a loss of accuracy.) Thus, most of the stimuli with voice-leading intervals will be evaluated both as potential arpeggiation and potential progressions, except for those with semitones, which only create progressions.

Example 6.31: Arpeggiation and progression percentages for stimuli with voice-leading intervals

<table>
<thead>
<tr>
<th>Horizontal Intervals</th>
<th>Predictions</th>
</tr>
</thead>
<tbody>
<tr>
<td>{-10, 10, 14}</td>
<td>26.2% ARP</td>
</tr>
<tr>
<td>{-14, -13, -11, 11, 13}</td>
<td>5.8% ARP</td>
</tr>
<tr>
<td>{-2, 2}</td>
<td>2.4% ARP</td>
</tr>
<tr>
<td>{-1, 1}</td>
<td>0% ARP</td>
</tr>
</tbody>
</table>

A lack of voice-leading intervals does not mean that the stimulus will necessarily be heard as an arpeggiation, however – the other side of our flowchart also admits some ambiguity between arpeggiation and progression. As was discussed above, there are some stimuli amongst those with no voice-leading intervals which occasionally seem to lack “harmonic compatibility” between the initial dyad and the second event. One broad category is stimuli with an initial, vertical interval class [1] or [2] – these intervals have a certain tendency to sound like unstable elements that crave resolution. Also, figures that outline the less common triadic types
(diminished and augmented) have a tendency to imply an entire triad in their initial tertian dyad, which conflicts with the second element. Thus, these two groups seem to arpeggiate much less frequently than the figures that form major and minor triads. Since their percentage values are so close, our predictive model lumps the first two classes of figure together into one category and considers major and minor triads to be their own category (example 6.32).

**Example 6.32: Arpeggiation and progression in non-VL figures**

<table>
<thead>
<tr>
<th>Category</th>
<th>ARP (%)</th>
<th>PROG (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>major + minor triads</td>
<td>90.7</td>
<td>9.3</td>
</tr>
<tr>
<td>first vertical is [1] or [2]</td>
<td>55.2</td>
<td>44.8</td>
</tr>
<tr>
<td>dim + augmented triads</td>
<td>53.3</td>
<td>46.7</td>
</tr>
</tbody>
</table>

Thus, in the first stage of the predictive model a figure is assigned a probability of being an arpeggiation or a progression, selected from six pairs of values. We then apply slightly differing algorithms to flesh out each possibility and predict the most likely roots. The procedure for evaluating arpeggiations is fairly simple. We start with a list of the 24 most common interval-root associations in arpeggiations (example 6.33). (Ranking the interval-root associations allows the most significant data to rise to the top, and an arbitrary cutoff of 24 excludes associations that seem incidental.) The list is ordered by overall commonness across all stimuli – thus members of interval class [1] and [2] appear fairly low on the list, since they appear less frequently in arpeggiations. However, the second column of values indicates how strongly the interval is associated with a particular root whenever it does appear.
**Example 6.33: Ranked interval-root associations for arpeggations**

<table>
<thead>
<tr>
<th>interval</th>
<th>root</th>
<th>overall % of instances</th>
<th>% when interval is present</th>
<th>comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>5</td>
<td>.255</td>
<td>.80</td>
<td>rooted interval</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>.203</td>
<td>.72</td>
<td>rooted interval</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>.124</td>
<td>.29</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>.12</td>
<td>.28</td>
<td>3 and 5 of major triad</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>.116</td>
<td>.36</td>
<td>rooted interval</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>.112</td>
<td>.29</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>.112</td>
<td>.36</td>
<td>rooted interval</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>.112</td>
<td>.38</td>
<td>dominant resolution</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td>.096</td>
<td>.30</td>
<td>3 and 5 of minor triad</td>
</tr>
<tr>
<td>4</td>
<td>9</td>
<td>.084</td>
<td>.27</td>
<td>3 and 5 of minor triad</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>.08</td>
<td>.21</td>
<td>3 and 5 of major triad</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>.076</td>
<td>.25</td>
<td>dominant resolution</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>.072</td>
<td>.19</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>5</td>
<td>.072</td>
<td>.45</td>
<td>dominant resolution</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>.068</td>
<td>.23</td>
<td>dominant resolution</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>.056</td>
<td>.20</td>
<td>dominant resolution</td>
</tr>
<tr>
<td>9</td>
<td>10</td>
<td>.056</td>
<td>.13</td>
<td>dominant resolution</td>
</tr>
<tr>
<td>9</td>
<td>7</td>
<td>.056</td>
<td>.13</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>.052</td>
<td>.13</td>
<td>dominant resolution</td>
</tr>
<tr>
<td>11</td>
<td>11</td>
<td>.052</td>
<td>.30</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>.052</td>
<td>.30</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>.044</td>
<td>.28</td>
<td>dominant resolution</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>.044</td>
<td>.15</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>.04</td>
<td>.26</td>
<td></td>
</tr>
</tbody>
</table>

This approach conveniently captures several different processes at work – it shows the rooted intervals 4, 5, 7 and 8 pointing to a tonic tone, other intervals engaging a triadic projection that also suggests a tonic, and some non-tonic interpretations that imply a future resolution. (The strong association of interval 6 with the root of 1, for example, reflects a traditional resolution of $\hat{7}$ ($+4$) to $\hat{1}$.)
The model assumes that the three given intervals in a figure will each potentially suggest one or more roots. It creates an inventory of the root associations for each interval and their strengths (i.e. the frequency that the interval is actually associated with the particular root.) Example 6.34 shows the tally of root associations for the intervals within a major-triad figure.

**Example 6.34: Root predictions for a major-triad figure**

The figure

combined probabilities for a C root

\[
\begin{align*}
x (\text{int } 3) &= .21 & \text{not } x &= .79 \\
y (\text{int } 4) &= .36 & \text{not } y &= .64 \\
z (\text{int } 7) &= .72 & \text{not } z &= .28 \\
\end{align*}
\]

When two intervals potentially engage the same tonic, they act together in a

multiplicative fashion. We can imagine that each interval may or may not elicit the associated
root by itself, or that both might work in concert. Given A and B pointing to the same target C, the overall probability of C is the sum of the probabilities that A will “succeed” and B will “fail,” A \times (1-B), that B will succeed and A will fail, (1-A) \times B, and that both will succeed, A \times B. As a shortcut we can subtract the probability that both will fail, (1-A) \times (1-B), from 100\%.

Example 6.35 illustrates a more complex, three-way combination of probabilities, for which the shortcut is again the possibility that all will fail, (1-A) \times (1-B) \times (1-C) subtracted from 100\%.

Progressions follow a similar procedure which I’ll detail below. Then, for each figure, the root predictions for both arpeggiation and progressions are multiplied by their overall ARP and PROG percentages and combined into a single list. All but the top 3 results are discarded, since the actual data only contains 3 interpretations per stimulus. These three results are normalized to add up to 100\%. Example 6.35 shows the top 3 results for our major triad figure, which, unsurprisingly, are all arpeggiation.

**Example 6.35: Top 3 results for Stimulus No. 70**

<table>
<thead>
<tr>
<th>top 3 results</th>
<th>normalized to add to 100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Root C .86 × ARP (.907) = .78</td>
<td>.53</td>
</tr>
<tr>
<td>Root F .48 × ARP (.907) = .43</td>
<td>.29</td>
</tr>
<tr>
<td>Root E .29 × ARP (.907) = .26</td>
<td>.18</td>
</tr>
</tbody>
</table>

---

120 This method of combining probabilities is similar to the approach discussed at [http://www.mathpages.com/home/kmath267.htm](http://www.mathpages.com/home/kmath267.htm), accessed 1/13/2014.
In order to measure how well our results fit our data, we try to round our final percentages into whole-number predictions that add up to 3. Dividing the terms in example 6.36 by .33 we get 1.6, .88, and .55. Since these numbers round to 2, 1, and 1, producing a surplus prediction, we gradually increase our divisor until the undesired prediction is eliminated – our final result is 2 predicted arpeggiations with root C, and 1 with a dominant resolution to F. Comparing this to our actual results, 3 arpeggiations with a root of C, we conclude that we were 66% correct.

The predictions for progressions follow a similar procedure to what I’ve just detailed. We work off of two interval-and-root lists, one for initial verticals and one for horizontals (example 6.36). Both tally up the results for intervals that are participating in progressions, regardless of whether they win in the local hierarchy or lose. As above, the interval lists capture multiple tonal forces at work. We see the influence of intervallic rootedness, triadic projection, and non-tonic resolutions, and, for the first time, the remarkably strong force of voice-leading and the “finality effect” in the horizontals (i.e. the values for inverse directed intervals 1, 2, 10 and 11 with a root on 0, their target note.) The root implications of all intervals are then combined in a similar fashion to arpeggiations.
Example 6.36: Interval-root lists for progressions

a. initial dyads (verticals)

<table>
<thead>
<tr>
<th>interval</th>
<th>root</th>
<th>overall % of instances</th>
<th>% when interval is present</th>
<th>comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>0</td>
<td>.073</td>
<td>.75</td>
<td>rooted interval</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>.066</td>
<td>.58</td>
<td>rooted interval</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>.056</td>
<td>.53</td>
<td>rooted interval</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>.047</td>
<td>.5</td>
<td>rooted interval</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>.033</td>
<td>.26</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>.028</td>
<td>.35</td>
<td>₃ and ₅ of major triad</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>.024</td>
<td>.25</td>
<td>₃ and ₅ of major triad</td>
</tr>
<tr>
<td>10</td>
<td>8</td>
<td>.021</td>
<td>.16</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>10</td>
<td>.019</td>
<td>.2</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>.019</td>
<td>.2</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>.017</td>
<td>.19</td>
<td>dominant resolution</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>.017</td>
<td>.19</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>.017</td>
<td>.15</td>
<td>dominant resolution</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>.016</td>
<td>.22</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>.016</td>
<td>.15</td>
<td>dominant resolution</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
<td>.016</td>
<td>.12</td>
<td>dominant resolution</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>.016</td>
<td>.12</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>10</td>
<td>.014</td>
<td>.16</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>4</td>
<td>.014</td>
<td>.16</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>.012</td>
<td>.15</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>.012</td>
<td>.15</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>.012</td>
<td>.13</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>.012</td>
<td>.13</td>
<td>dominant resolution</td>
</tr>
<tr>
<td>11</td>
<td>5</td>
<td>.012</td>
<td>.14</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>.01</td>
<td>.15</td>
<td>dominant resolution</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>.01</td>
<td>.15</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>.01</td>
<td>.11</td>
<td>dominant resolution</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>.01</td>
<td>.1</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>.01</td>
<td>.11</td>
<td></td>
</tr>
</tbody>
</table>
b. horizontal intervals (from dyad to monad)

<table>
<thead>
<tr>
<th>interval</th>
<th>root</th>
<th>overall % of instances</th>
<th>% when interval is present</th>
<th>comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>0</td>
<td>.122</td>
<td>.7</td>
<td>rooted interval</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>.108</td>
<td>.42</td>
<td>melodic anchoring</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>.099</td>
<td>.42</td>
<td>melodic anchoring</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>.087</td>
<td>.39</td>
<td>melodic anchoring</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>.068</td>
<td>.27</td>
<td>melodic anchoring</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>.059</td>
<td>.24</td>
<td>(reverse) melodic anchoring</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>.056</td>
<td>.24</td>
<td>(reverse) melodic anchoring</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>.05</td>
<td>.35</td>
<td>rooted interval</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>.045</td>
<td>.37</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>.045</td>
<td>.31</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>.045</td>
<td>.31</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>.043</td>
<td>.3</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>.043</td>
<td>.25</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>.042</td>
<td>.33</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>.04</td>
<td>.16</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>.038</td>
<td>.17</td>
<td>(reverse) melodic anchoring</td>
</tr>
<tr>
<td>1</td>
<td>9</td>
<td>.037</td>
<td>.15</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>.035</td>
<td>.23</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>.033</td>
<td>.22</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>4</td>
<td>.033</td>
<td>.13</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>11</td>
<td>.031</td>
<td>.12</td>
<td>(reverse) melodic anchoring</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>.03</td>
<td>.12</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>.028</td>
<td>.18</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>.026</td>
<td>.18</td>
<td>dominant resolution</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>.026</td>
<td>.21</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>5</td>
<td>.026</td>
<td>.12</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>9</td>
<td>.026</td>
<td>.1</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>.024</td>
<td>.14</td>
<td>dominant resolution</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>.024</td>
<td>.14</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>6</td>
<td>.024</td>
<td>.11</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>.023</td>
<td>.18</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>.023</td>
<td>.13</td>
<td>dominant 7th fragment</td>
</tr>
<tr>
<td>6</td>
<td>11</td>
<td>.021</td>
<td>.12</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>.019</td>
<td>.08</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>.019</td>
<td>.08</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>9</td>
<td>.019</td>
<td>.13</td>
<td>3 and 5 of minor triad</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>.019</td>
<td>.13</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>.019</td>
<td>.15</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>.017</td>
<td>.1</td>
<td></td>
</tr>
</tbody>
</table>
Example 6.37 shows the interval-root tallies and combined root predictions for figure No. 91, a somewhat ambiguous stack of perfect fourths.

**Example 6.37: Root predictions for Figure No. 91**

Once a collection of roots is predicted, is it time to determine whether the progression would be locally right-branching or left-branching. As I argued earlier, the intervallic content of the figures often seems to define the overall tonal orientation first, which then determines the relative structural importance of the two events. Thus, for each potential root we consider the
resulting scale-degree identity for the tones in the figure. The branch with the most “central” tone is predicted to win in the event hierarchy. We consider only dominants and tonics, with priority given to the tonic. Example 6.38 shows the branching predictions for all five possible roots from 6.37. Note that the suggestion of a G♭ root does not create any tonally central pitch-classes and is discarded.

Example 6.38: Branching predictions for the stacked-fourth figure

As was noted above these progression predictions are multiplied by their overall PROG score and combined with the arpeggiation root predictions. The top three results are selected and the percentages are normalized. Example 6.39 shows our top three predictions for the stacked-fourth figure.
Example 6.39: Top 3 results for the stacked-fourth figure.

<table>
<thead>
<tr>
<th>top 3 results</th>
<th>normalized to add to 100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Root B♭, right-branching .68 \times PROG (.738) = .50</td>
<td>.41</td>
</tr>
<tr>
<td>Root C, left-branching .57 \times PROG (.738) = .42</td>
<td>.35</td>
</tr>
<tr>
<td>Root F, right-branching .39 \times ARP (.738) = .29</td>
<td>.24</td>
</tr>
</tbody>
</table>

As before these three results will be divided by .33 and rounded in order to match our percentages to the actual results. Here the predictions round neatly to forecast one instance each, and, happily, they are a 100% match with the actual data. Example 6.40 presents the three interpretations I entered for figure 91, transposed so that the root is always C.

Example 6.40: Actual results

![Musical examples](image)

At present each individual stage of the algorithm performs with what seems like acceptable accuracy. The initial arpeggiation vs. progression predictions are 85% accurate. Our progression prediction algorithm can match actual hierarchical branchings with 76% accuracy.
and can match both roots and branchings 60% of the time. Our arpeggiations algorithm can predict the roots of arpeggiations with 65% accuracy. However, when all of these parts of the model are combined the result is a somewhat dismal 42% match with actual data. However, the measure of accuracy I am adopting here is probably a bit unrealistic – rounding our percentages to whole-number predictions and matching them to a small handful of data points is probably akin to evaluating a similarly small number of coin flips – we know what the probabilities would be in such a situation but the actual results will involve quite a bit of variance. An experiment with more trials for each figure and a more sophisticated measure of correlation is probably in order before we can entirely discount this approach.

One aspect of the experiment which I think the model has successfully captured is the ambiguity of these stimuli, the fact that the same figure can elicit more than one interpretation. About two thirds of our stimuli evoked more than one root or hierarchical branching, and the algorithm gives a plausible account of how this plurality might come about. In our stacked fourth figure, for example, the model captures the tendency of the initial fourth to suggest a root, the suggestive but ambiguous motion of a melodic descending fourth (associated both with a root of C and F) and the emphasis created by the implicit pitch-class step up into the second event. All of these factors balanced to predict the three interpretations found in the experiment.

At present the algorithm is obviously tailored to the very specific and somewhat artificial design of this experiment. One of the major modifications that will be necessary to make it applicable to more diverse stimuli will be a registral weighting of tones and intervals, so that a perfect fifth with a registraIly emphasized bass will count more than a perfect fifth in the upper structure of the chord. We would also need to explicitly disentangle the vertical and the horizontal and develop a theory of how multiple successive events interact in a recursive event
hierarchy. In general, the addition of more tones and more events must have the effect of reducing ambiguity in the model, since the typical well-formed succession of fully-realized sonorities is not tonally ambiguous.

Executing the experiment with multiple subjects

Many modifications are also in order if a dyad-plus-monad experiment is to be executed with an actual pool of subjects. Perhaps the first priority would be to sharply reduce the number of stimuli, since an experiment with hundreds of trials is simply not practical. A strategic sampling of the figures with a variety of hypothesized properties (and, perhaps, a perfectly balanced distribution of intervals) could be selected. Subsequent trials could test a partially overlapping set of the figures, so that one could eventually amass a wide pool of data while insuring that subjects are responding with consistency.

The open-ended investigation of properties must also be curtailed, focusing on one aspect of the stimuli at a time. I can imagine experiments focused solely on the saliency of voice-leading connections, the implicit continuation of tones, the distinction between arpeggiation and progression, the assignment of a tree branching, or the selection of a tonic. The user interface must also be simplified to admit only one kind of feedback, without multiple ways to express the same result.

Finally, one would have to decide what kind of subject we are interested in. Some of these properties (like voice-leading connectedness) may be a fairly direct result of low-level
auditory processes, and thus it might make sense to test the ability of the general public to perceive them in a consistent way. However, I suspect that many of the tonal tendencies I’ve measured, while acoustic in origin, are also cultivated by intensive musical experience and training. Thus one would necessarily have to focus on a narrower population that exhibits a baseline sensitivity to tonality. One approach that seems promising is to screen participants with a simple tonic-finding test that uses short passages of real music. René van Egmond and Mila Boswijk found that administering such a test was useful in identifying groups of similar subjects, more so than the typical questionnaire about years of musical experience and training.¹²¹

In general I remain confident that the dyad-plus-monad design is a very fertile testing ground for the functioning of tonality, neither too abstract and impoverished nor too complex. Examining these figures in a rigorous fashion has inspired a vision of tonality that is not merely attributable to an internalized sense of statistical regularities, but rather the result of specific, interacting primitive auditory processes.

By disassembling the tonal hierarchy into its constituent parts one gains a better understanding of what it is and how it works. We can isolate each individual process and see how it interacts with others. Each tendency becomes more nuanced and limited – not an absolute principle of structure but a potential or a probability.

What we’ve established in the current study is, of course, quite simple. Intervallic rootedness creates the vertical hierarchization of tonality and accounts for the primacy of the tonic triad. Short-term memory for pitch locations combined with a constraint on successive semitones generates a sense of underlying scale. The rhythmic combination of events combines with the force of voice-leading to create motions to and from structural reference points. These three primitive processes may be sufficient to define tonality at its most basic level. Adding the projective force of meter and a general sense of recursion could possibly expand the model to the point where it could account for the conventional observations we make about harmony and key in Common Practice-era music.

One might protest that such an account is actually too simple, in that it fails to embrace any of the things we find interesting about music. The current account fails, even, to account for interkey relationships or the tension and release we find within musical phrases, topics that Krumhansl and Lerdahl have explored within their own work. Some of these failures, are, of course, due to the limitations of scope within any self-contained study. However, I think there is
one distinct advantage to keeping our cognitive theories as simple as possible – it creates a dividing line between what is innate and cognitively embedded and what is cultural or conceptual – or, to put it more bluntly, what is “natural” and what is man-made.

If music is, in fact, a reappropriation of simple, everyday properties of hearing and communication, it follows that we can gain an appreciation for the works we consume on a daily basis as the result of centuries of cultural engineering, the careful assembly of basic elements into more and more sophisticated aural structures that are then transmitted from generation to generation. Viewing music as a complex artifice may better account for our experience as learners, teachers, and analysts, that the tonal distinctions we eventually learn to make are the results of some effort, and perhaps not automatic or universal. Consequently, a single piece of music tends to demand repeated hearings and an endless process of discovery. Perhaps it is this extravagant complexity that we value in music, properties that extend well beyond the boundaries of what we are “hard-wired” to perceive.
Bibliography


