Tension-driven Automatic Music Generation

Thesis

How to cite:

For guidance on citations see FAQs.

© 2021 Germán Ruiz Marcos

https://creativecommons.org/licenses/by-nc-nd/4.0/

Version: Version of Record

Link(s) to article on publisher's website:
http://dx.doi.org/10.21954/ou.ro.000142cf

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online’s data policy on reuse of materials please consult the policies page.
Tension-driven automatic music generation

Germán Ruiz-Marcos

Supervised by
Dr Robin Laney and Dr Alistair Willis

School of Computing and Communications
Faculty of Science, Technology, Engineering and Mathematics
Open University

August, 2021

A dissertation submitted for the degree of Doctor of Philosophy.
Abstract

The Ancient Greeks are one of the first civilisations we know of to have created algorithms to compose music. Since then, algorithmic techniques have vastly improved with increasingly sophisticated computers. In the last two decades, much research in this area has focused on two goals: designing algorithms which generate music as close as possible to that of human composers and implementing those algorithms to automatically generate music in interactive scenarios, such as video games.

To meet these goals, automatically generated music should:

■ focus on higher-level concepts, such as musical tension,
■ have long-term structure, and
■ be able to adapt to changes in real time.

Combining these three requirements is, however, a challenging task. This dissertation investigates three steps to overcome this challenge. First, we argue that Lerdahl’s model of musical tension is suited to the automatic generation of tonal music that has long-term structure and that matches a given tension profile. By means of an illustrative example, we review Lerdhal’s model and implement a novel computational system to automate it. Second, we show that an effective generation strategy is to combine statistical methods with both rule-based methods and generative grammars to create a music generation system. Third, we implement the system and evaluate it through a collection of computational tests and empirical studies.

Our evaluation shows that:

(1) the system works effectively in real time, as long as the input tension profiles do not contain too many steep transitions,

(2) the hierarchical structure perceived by listeners matches the patterns intended by the system in the generated music, and

(3) tension-changing input profiles are accurately matched by the generated music.
Acknowledgements

It goes without saying that it would have not been possible to carry out the present research without the support and advice given by many people, to whom I am very thankful, especially the following:

Robin Laney and Alistair Willis, my supervisors, who gave me the freedom to research creatively, whose expert advice and constant support helped me grow into a rigorous researcher and who did everything possible so that I always found myself in a warm and welcoming environment.

Fred Lerdahl and Carol Krumhansl, whose work is the main inspiration of this research, who kindly provided lots of data that guided the progress of my research project and who were open to long discussions about theoretical and methodological aspects of their work.

Ben Winters and Colin Johnson, my examiners, who took the time and effort to carefully read my work, who made the viva a once-in-a-lifetime amazing experience and whose comments have greatly benefited this dissertation. I would also like to thank Soraya Kouadri for dealing with the examination process and for acting as the viva chair.

Shailey Minocha, my third party monitor, whose door was always open and who cared about my research project from the beginning to the end.

Marian Petre and Daniel Gooch, who untiringly organised and hosted the PG-Forum (Post-graduate research students Forum), at the School of Computing and Communications and the Knowledge Media Institute, which allowed me to improve my research skills. I would also like to extend my gratitude to all the guest speakers and colleagues that made the forum possible by sharing their knowledge and experience.

The members of the NLP (Natural Language Processing) group, who sat through all the stages of this research and whose feedback helped me shape my research project.

Allan Jones and Simon Holland, who kindly accepted to act as my examiners in a mock viva and whose incredibly valuable comments helped me be better prepared for my examination.

Robert Samuels and Allan Jones, whose feedback on the preliminary stages of my research helped me define my research approach.
Alvaro Faria and Paul Garthwaite, who helped me find my way out of some rabbit holes I went down concerning the statistical analysis.

Duncan Banks, who helped me better understand the intricacies of physiology-centred studies.

The many friends who probably got tired of hearing me speak about my research project and thankfully pulled me out of the office, from time to time, to get some fresh air.

And last, but not least, my family, who were always by my side along the bumpy road a PhD journey may get to be.
The work presented in this dissertation has led to the following publications and research activities:

**Conference papers (peer-reviewed)**


**Doctoral consortium papers**


Poster presentations


Germán Ruiz-Marcos (2019). My computer writes music on its own... does yours? In: 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW). Available at: https://doi.org/10.21954/ou.rd.13026956.v1

Demos


Repositories

Germán Ruiz-Marcos, Robin Laney, and Alistair Willis (2020). autognomus: AUtomatic Tension-Oriented GeNerator Of MUSic [Software]. Available at: https://doi.org/10.21954/ou.rd.15028599.v1

Germán Ruiz-Marcos, Robin Laney, and Alistair Willis (2020). AuToTen: an automatic calculator of tension in tonal music [Software]. Available at: https://doi.org/10.21954/ou.rd.13026578.v1
# Contents

1 Introduction

1.1 Problem domain: tension-driven automatic music generation

1.2 Motivations: state-of-the-art challenges

1.3 Research approach: goals, questions and objectives

1.4 Dissertation outline

2 Modelling Musical Tension: scope, definition and models

2.1 Framing the scope of the study of musical tension

2.1.1 Why is it so challenging to define the concept of musical tension?

2.1.2 Towards a definition of the concept of musical tension

2.1.3 Defining the concept of musical tension

2.1.4 A taxonomy of the approaches to modelling musical tension

2.2 Theoretical accounts of musical tension

2.2.1 Meyer's theory of musical emotions and meaning

2.2.2 Narmour's implication-realisation model

2.2.3 Swain's dimensions of harmonic rhythm

2.2.4 Lerdahl and Jackendoff's Generative Theory of Tonal Music (GTTM)

2.2.5 A critical look at theoretical accounts of musical tension

2.3 Empirical studies of musical tension

2.3.1 Studies that record human judgements

2.3.2 Studies that record physiological responses
2.3.3 Studies that record neural activity .......................... 30
2.3.4 A critical look at empirical studies of musical tension .... 31
2.4 Generation models of musical tension ........................ 33
2.4.1 Lerdahl’s Model of Tonal Tension (MTT) ................. 34
2.4.2 Farbood’s Parametric-Temporal Model of Musical Tension ........................ 34
2.4.3 Herremans and Chew’s Tension Ribbons .................... 35
2.4.4 A critical look at generation models of musical tension .... 37
2.5 Which models of tension are best suited to our goals? ...... 38
2.6 Summary and Conclusions ..................................... 40

3 Lerdahl’s Model of Tonal Tension: background, overview and application 43
3.1 Background .................................................. 44
3.2 A Generative Theory of Tonal Music (GTMM) ................ 46
3.2.1 The grouping structure .................................. 49
3.2.2 The metrical structure ................................... 51
3.2.3 The time-span reduction ................................ 55
3.2.4 The prolongational reduction .............................. 60
3.3 Tonal Pitch Space (TPS) ...................................... 67
3.3.1 The pitch-class level ..................................... 67
3.3.2 The chordal level ...................................... 69
3.3.3 The regional level ...................................... 72
3.4 Model of Tonal Tension (MTT) ............................... 75
3.4.1 The harmonic sequential model ........................... 75
3.4.2 The harmonic hierarchical model ......................... 77
3.4.3 The melodic model ...................................... 80
3.5 Empirical evaluation ........................................... 83
3.5.1 Participants and materials ................................ 83
3.5.2 Procedure ............................................... 83
3.5.3 Results .................................................. 84
3.5.4 Conclusions ............................................. 87
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.6</td>
<td>Summary and Conclusions</td>
<td>88</td>
</tr>
<tr>
<td>4</td>
<td>Automatically Calculating Tonal Tension: <em>challenges, implementation and evaluation</em></td>
<td>89</td>
</tr>
<tr>
<td>4.1</td>
<td>Why is it so challenging to fully automate Lerdahl’s Model of Tonal Tension?</td>
<td>90</td>
</tr>
<tr>
<td>4.1.1</td>
<td>Calculating $i$ in the Tonal Pitch Space</td>
<td>91</td>
</tr>
<tr>
<td>4.1.2</td>
<td>Calculating $j$ in the Tonal Pitch Space</td>
<td>92</td>
</tr>
<tr>
<td>4.1.3</td>
<td>Representing the prolongational reduction</td>
<td>93</td>
</tr>
<tr>
<td>4.1.4</td>
<td>Calculating distances within the Tonal Pitch Space across regions</td>
<td>93</td>
</tr>
<tr>
<td>4.2</td>
<td><em>AutoTPS</em>: automating Lerdahl’s Tonal Pitch Space</td>
<td>94</td>
</tr>
<tr>
<td>4.2.1</td>
<td>Automatically calculating $i$ in the Tonal Pitch Space</td>
<td>95</td>
</tr>
<tr>
<td>4.2.2</td>
<td>Automatically calculating $j$ in the Tonal Pitch Space</td>
<td>97</td>
</tr>
<tr>
<td>4.2.3</td>
<td>Automatically calculating $k$ in the Tonal Pitch Space</td>
<td>99</td>
</tr>
<tr>
<td>4.3</td>
<td><em>AuToTen</em>: automating Lerdahl’s Model of Tonal Tension</td>
<td>101</td>
</tr>
<tr>
<td>4.3.1</td>
<td>A representation of the prolongational reduction</td>
<td>102</td>
</tr>
<tr>
<td>4.3.2</td>
<td>Automatic calculation of tonal tension</td>
<td>104</td>
</tr>
<tr>
<td>4.3.3</td>
<td>Automatic feature extraction</td>
<td>105</td>
</tr>
<tr>
<td>4.4</td>
<td>Computational evaluation</td>
<td>106</td>
</tr>
<tr>
<td>4.4.1</td>
<td>Evaluating <em>AutoTPS</em>’s accuracy</td>
<td>106</td>
</tr>
<tr>
<td>4.4.2</td>
<td>Evaluating <em>AuToTen</em>’s accuracy</td>
<td>107</td>
</tr>
<tr>
<td>4.5</td>
<td>Summary and Conclusions</td>
<td>108</td>
</tr>
<tr>
<td>5</td>
<td>Automatic Music Generation: <em>definition, scope and methods</em></td>
<td>113</td>
</tr>
<tr>
<td>5.1</td>
<td>Framing the scope of the field of Automatic Music Generation</td>
<td>114</td>
</tr>
<tr>
<td>5.1.1</td>
<td>Defining the concept of Automatic Music Generation</td>
<td>114</td>
</tr>
<tr>
<td>5.1.2</td>
<td>A taxonomy of the main methods in the field</td>
<td>116</td>
</tr>
<tr>
<td>5.2</td>
<td>Which computational methods are best suited to our goals?</td>
<td>117</td>
</tr>
<tr>
<td>5.2.1</td>
<td>Non-AI generation methods</td>
<td>117</td>
</tr>
<tr>
<td>5.2.2</td>
<td>Soft computing methods</td>
<td>119</td>
</tr>
<tr>
<td>5.2.3</td>
<td>Symbolic AI methods</td>
<td>123</td>
</tr>
<tr>
<td>5.3</td>
<td>Hybridisation</td>
<td>127</td>
</tr>
</tbody>
</table>
5.4 Summary and Conclusions .................................................. 129

6 The Music Generator: architecture, design and implementation ..................... 131

6.1 autognomus: the music generator ........................................ 132
   6.1.1 Compositional Strategy ............................................. 132
   6.1.2 TAnDeM: autognomus’s architecture .............................. 133
   6.1.3 How does autognomus work? ...................................... 135

6.2 Themer: the generator of themes ......................................... 136
   6.2.1 Step 1: defining the input parameters .............................. 137
   6.2.2 Step 2: generating the theme’s structure ............................ 138
   6.2.3 Step 3: generating the chord labels in the theme’s first semi-phrase 144
   6.2.4 Step 4: generating the chord labels in the theme’s second semi-phrase 151
   6.2.5 Step 5: generating a melody to fit the theme’s first semi-phrase 156
   6.2.6 Step 6: generating a melody to fit the theme’s second semi-phrase 168
   6.2.7 Step 7: generating the theme’s representations .................... 171

6.3 Arranger: the generator of arrangements .................................. 171
   6.3.1 Step 1: generating a theme’s arrangement .......................... 172
   6.3.2 Step 2: generating the arrangement’s representation .............. 174

6.4 Developer: the generator of developments .................................. 174
   6.4.1 Step 1: generating a theme’s development .......................... 175
   6.4.2 Step 2: generating the developments’ representations .......... 183

6.5 Morpher: the morphing system ........................................... 183
   6.5.1 Step 1: reading the input data ...................................... 185
   6.5.2 Step 2: generating new harmonic material that matches input de-
             grees of tension .................................................. 185
   6.5.3 Step 3: generating new melodic material that matches input de-
             grees of tension .................................................. 191
   6.5.4 Step 4: generating a new arrangement that matches the input tension 199
   6.5.5 Step 5: playing the generated music in real time ................ 202

6.6 Summary and conclusions .................................................. 203
7 Evaluation: methodology, computational tests and empirical studies

7.1 Evaluating the Music Generator .................................................. 206
  7.1.1 To what extent is the Music Generator applicable in real time? ... 206
  7.1.2 To what extent does the generated music have long-term structure? 210
  7.1.3 To what extent does the generated music match input tension profiles? .................................................. 210

7.2 Computational test ................................................................. 211
  7.2.1 Materials ................................................................. 211
  7.2.2 Procedure ............................................................. 211
  7.2.3 Results ................................................................. 212

7.3 Empirical study I ................................................................. 214
  7.3.1 Participants ............................................................. 214
  7.3.2 Materials ............................................................... 215
  7.3.3 Procedure ............................................................. 216
  7.3.4 Results ................................................................. 217

7.4 Empirical study II ................................................................. 227
  7.4.1 Participants ............................................................. 227
  7.4.2 Materials ............................................................... 227
  7.4.3 Procedure ............................................................. 230
  7.4.4 Results ................................................................. 231

7.5 Discussion ................................................................. 239
  7.5.1 The extent to which the Music Generator is applicable in real time 239
  7.5.2 The extent to which the generated music has long-term structure 240
  7.5.3 The extent to which the generated music matches input tension profiles .................................................. 241

7.6 Summary and Conclusions .................................................. 242

8 Conclusions: dissertation review, research insights and contributions, limitations and future directions

8.1 Dissertation review .......................................................... 246
8.2 Research insights ........................................ 249
8.3 Research contributions ..................................... 249
8.4 Limitations ................................................... 251
  8.4.1 The design and implementation of the Music Generator .... 251
  8.4.2 The evaluation methodology ............................ 253
  8.4.3 The quality of the automatically generated music ............ 255
  8.4.4 The scope of the present research ....................... 256
8.5 Future work .................................................. 257
8.6 Closing remarks ............................................. 258

Appendix A  Experimental materials and additional data 259
  A.1 Experimental music materials ................................ 259
  A.2 Experimental interface materials ............................ 266
  A.3 Graphical representations ................................... 270

Glossary of musical terms 281

References 295
# List of Figures

1.1 An outline of the scope of this dissertation. ............................ 6

2.1 A classification of the approaches to modelling musical tension. .... 19

2.2 A flowchart of the bottom-up component of Narmour’s implication-realisation model, for a given three-note melodic line, x-y-z. The flowchart focuses on the intervals between the notes in the given melodic line and the direction (i.e. ascending or descending) between consecutive notes. .............. 23

3.1 An intuitive reduction of the first phrase of the theme in Mozart’s Ah vous dirai-je, Maman, K. 265/300e, with regards to the level of importance of its notes. ............................................................ 48

3.2 GTTM’s grouping structure of the first phrase of the theme in Mozart’s Ah vous dirai-je, Maman, K. 265/300e. ............................................................ 50

3.3 GTTM’s metrical structure of the first phrase of the theme in Mozart’s Ah vous dirai-je, Maman, K. 265/300e. ............................................................ 54

3.4 An example of possible hierarchical tree structures from the reduction of the first phrase of the theme in Mozart’s Ah vous dirai-je, Maman, K. 265/300e. . 56

3.5 GTTM’s time-span reduction (top) and time-span segmentation (bottom) of the first phrase of the theme in Mozart’s Ah vous dirai-je, Maman, K. 265/300e. . 59

3.6 Types of branching connections in GTTM’s prolongational reduction. . . . . . . 62

3.7 GTTM’s basic form ............................................................ 65

3.8 GTTM’s normative structure ............................................................ 65
<table>
<thead>
<tr>
<th>Figure Number</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.9</td>
<td>GTTM’s basic form + normative structure</td>
<td>65</td>
</tr>
<tr>
<td>3.10</td>
<td>GTTM’s prolongational reduction and chord labels of the first phrase of the</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>theme in Mozart’s Ah vous dirai-je, Maman, K. 265/300e.</td>
<td></td>
</tr>
<tr>
<td>3.11</td>
<td>TPS’s basic space oriented toward I/C.</td>
<td>68</td>
</tr>
<tr>
<td>3.12</td>
<td>TPS’s basic space (compact version) oriented toward I/C.</td>
<td>68</td>
</tr>
<tr>
<td>3.13</td>
<td>A transition from the basic space oriented toward I/C to that oriented toward</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>V/C</td>
<td></td>
</tr>
<tr>
<td>3.14</td>
<td>Major-key diatonic circle-of-fifths (chord label format).</td>
<td>70</td>
</tr>
<tr>
<td>3.15</td>
<td>C major’s diatonic circle-of-fifths (roots format).</td>
<td>70</td>
</tr>
<tr>
<td>3.16</td>
<td>TPS’s basic space (compact version) oriented toward V/C.</td>
<td>71</td>
</tr>
<tr>
<td>3.17</td>
<td>Graphical representation of the chordal space.</td>
<td>72</td>
</tr>
<tr>
<td>3.18</td>
<td>Major-key chromatic circle-of-fifths (key label format).</td>
<td>73</td>
</tr>
<tr>
<td>3.19</td>
<td>Major-key chromatic circle-of-fifths (roots format).</td>
<td>73</td>
</tr>
<tr>
<td>3.20</td>
<td>TPS’s basic space (compact version) oriented toward I/G.</td>
<td>73</td>
</tr>
<tr>
<td>3.21</td>
<td>Graphical representation of the regional space.</td>
<td>74</td>
</tr>
<tr>
<td>3.22</td>
<td>GTTM’s prolongational reduction and chord labels of the first phrase of the</td>
<td>77</td>
</tr>
<tr>
<td></td>
<td>theme in Mozart’s Ah vous dirai-je, Maman, K. 265/300e (numbered version).</td>
<td></td>
</tr>
<tr>
<td>3.23</td>
<td>TPS’s anchoring space (compact version) in the key of C major.</td>
<td>80</td>
</tr>
<tr>
<td>4.1</td>
<td>Ascending major-key chromatic circle-of-fifths.</td>
<td>96</td>
</tr>
<tr>
<td>4.2</td>
<td>Descending major-key chromatic circle-of-fifths.</td>
<td>96</td>
</tr>
<tr>
<td>4.3</td>
<td>An example of the calculation of $j(I/C \rightarrow I/D^\flat)$ using AutoTPS.</td>
<td>100</td>
</tr>
<tr>
<td>4.4</td>
<td>An example of the calculation of $j(I/C \rightarrow I/C^\sharp)$ using AutoTPS.</td>
<td>100</td>
</tr>
<tr>
<td>4.5</td>
<td>An example of the calculation of $j(I/C \rightarrow II/C)$ using AutoTPS.</td>
<td>100</td>
</tr>
<tr>
<td>4.6</td>
<td>TPS’s basic space (compact version) oriented toward the tonic chord of C^\sharp</td>
<td>101</td>
</tr>
<tr>
<td></td>
<td>major or D^\flat major.</td>
<td></td>
</tr>
<tr>
<td>4.7</td>
<td>TPS’s basic space (compact version) oriented toward I/C (repeated figure).</td>
<td>101</td>
</tr>
<tr>
<td>4.8</td>
<td>Prolongational matrix, (a), associated with the prolongational reduction of the</td>
<td>103</td>
</tr>
<tr>
<td></td>
<td>theme in Mozart’s Ah vous dirai-je, Maman, K. 265/300e, (b).</td>
<td></td>
</tr>
</tbody>
</table>
4.9 *Grail* theme from Wagner's *Parsifal*, and its chord labels, transcribed from Lerdahl and Krumhansl (2007, p.343). .................................................. 107


5.1 A taxonomy of the main methods in the field of Automatic Music Generation. 118

5.2 A grammar example, transcribed from Holtzman (1980, p.55), that generates the ABA form typical in the classical sonata. ................................. 127

6.1 A diagram of *autognomus*'s *TAnDeM* architecture. .............................. 134

6.2 An example of the application of *Themer*'s Step 1. .............................. 137

6.3 GTTM's *basic form + normative structure* including its prolongational beginning (PB) and its cadential structure (CS). ................................. 138

6.4 A prolongational structure including all of *Themer*'s possible chords, which are labelled in *Themer* as: (1) First Chord, (2) Reprise Chord, (3) Left Subother Chord, (4) Other Chord, (5) Right Subother Chord, (6) Left Subnormative Chord, (7) Normative Chord, (8) Right Subnormative Chord, (9) Sub\textsubscript{3} Subdominant Chord, (10) Sub\textsubscript{2} Subdominant Chord, (11) Sub\textsubscript{1} Subdominant Chord, (12) Subdominant Chord, (13) Cadential\textsuperscript{6\textdegree} Chord, (14) Dominant Chord and (15) Last Chord. ................................. 141

6.5 An example of the application of *Themer*'s Step 2. .............................. 144

6.6 An example of the application of *Themer*'s Step 3 including: (a) PB's structure from Figure 6.5, (b) the generation of First Chord, (c) the generation of Normative Chord, (d) the generation of Reprise Chord and (e) the generation of Left Subnormative Chord. ................................. 150

6.7 An example of the application of *Themer*'s Step 4 including: (a) CS's structure from Figure 6.5, (b) the generation of Last Chord, (c) the generation of Dominant Chord, (d) the generation of Subdominant Chord and (e) the generation of Sub\textsubscript{1} Subdominant Chord. ................................. 155

6.8 An example of the generation of the first reference pitch in the *melodic skeleton* against the First Chord, I/C, in Figure 6.6. ................................. 157
6.9 An example of the generation of the second reference pitch in the melodic skeleton that was started in Figure 6.8, against the First Chord in Figure 6.6. 158

6.10 An example of the weightings, (a) \( w_l \) and (b) \( w_r \), associated with the notes in C major’s diatonic scale when given the pair of reference pitches in Figure 6.9a. 159

6.11 An example of the generation of one new pitch, \( b_4 \), to fit the melodic skeleton in Figure 6.9a. 159

6.12 An example of the generation of another new pitch, \( a_4 \), to fit Figure 6.11. 160

6.13 Original palette of note values implemented in Themer. 162

6.14 An example of the assignment of the note value to the first pitch in Figure 6.12. 163

6.15 An example of the assignment of the remaining note values in Figure 6.14. 164

6.16 An example of the types of transformations implemented in Themer, including: (a) the first phrase of the theme in Mozart’s Ah vous dirai-je, Maman, K. 265/300e, as the reference theme, (b) the transposition of the reference theme down a fourth, (c) the inversion of the reference theme, (d) the retrogression of the reference theme and (e) the inverted retrogression of the reference theme. 166

6.17 An example of the application of Themer’s Step 5 including: (a) PB’s structure from Figure 6.6, (b) the generation of a new melodic line against First Chord, (c) the transposition of First Chord’s melodic line against Reprise Chord, (d) the inversion of First Chord’s melodic line against Normative Chord and (e) the generation of a new melodic line against Left Subnormative Chord. 168

6.18 An example of the application of Themer’s Step 6 including: (a) CS structure from Figure 6.7, (b) the repetition of First Chord’s melodic line against Last Chord, (c) the inversion of First Chord’s melodic line against Dominant Chord, (d) the transposition of First Chord’s melodic line against Subdominant Chord and (e) the retrogression of Subdominant Chord’s melodic line against \( Sub_1 \) Subdominant Chord. 170

6.19 An example of the application of Arranger’s Step 1 to the theme generated in Figures 6.17 and 6.18. 174
6.20 An example of the incorporation of new Subordinate Chords (dotted lines) into the hierarchical structure in Figure 6.3... 175

6.21 The two possible new hierarchical trees that Developer calculates to generate the development of a theme whose hierarchical structure is that of Figure 6.3. These trees incorporate one new Subordinate Chord (dotted line) per branch, either left or right, except for the first and last chords in the tree, whose branching directions are fixed. ... 177

6.22 An example of the application of Developer’s Step 1 including: (a) PB’s chords from Figure 6.17, (b) the generation of new right-branching Subordinate Chords (bold) according to Rule 6.5 (c) the generation of melodic lines against the newly generated Subordinate Chords as retrogressions of those in (b), (d) CS’s chords from Figure 6.18, (e) the generation of new right-branching Subordinate Chords (bold) according to Rule 6.5 and (f) the generation of melodic lines against the newly generated Subordinate Chords as retrogressions of those in (c). ... 180

6.23 An example of the application of Morpher’s Step 2 including: (a) the theme previously generated in Figure 6.19 and (b) the newly generated chords to match the tension profile in Figure 6.24. ... 189

6.24 Input tension profile to be matched by the theme in Figure 6.23a. ... 190

6.25 Diatonic index-based templates of the possible new melodic lines, per measure, pre-defined in Morpher. ... 194

6.26 Example of the application of Figure 6.25a to I/C. Melodic lines a to p refer to the templates 1 to 16 in Figure 6.25a, respectively. ... 195

6.27 An example of the application of Morpher’s Step 3 including: (a) the new harmonic sequence in Figure 6.23b, (b) the measures where new melodic material needs to be generated so that the input tension profile in Figure 6.24 is matched and (c) the newly generated melodic material. ... 197
6.28 An example of the application of *Morpher*’s Step 4 given Figure 6.24 as the input tension profile to be matched, including: (a) the theme in Figure 6.19 as input theme, (b) the measures that need to be filled with new material to match the input tension profile and (c) the newly generated material according to the harmonic sequence previously generated in Figure 6.23b and the new melodic lines previously generated in Figure 6.27c. 

7.1 Judgements of tension provided by the participants in Lerdahl and Krumhansl (2007) when they listened to Figure 7.2 (solid line) and the data points read by *autognomus* at the beginning of each measure if the solid line was fed as input into *autognomus* (dashed line). 

7.2 Reduction of Chopin’s Prelude No. 9 in E major, Opus 28, used in the empirical study in Lerdahl and Krumhansl (2007). 

7.3 Duration of a measure with regards to tempo, in BPM, for time-signatures with binary division, 2/4 and 6/8 (solid line); with ternary division, 3/4 and 9/8 (dashed line); and with quaternary division, 4/4 and 12/8 (dotted line). 

7.4 Degrees of correlation, $r$, between the tension profiles of IGA’s one hundred pieces and the degrees of tension read by *autognomus*. 

7.5 Virtual slider used in *autognomus*’s empirical studies. 

7.6 Graphical representation of the predictions of musical tension of each multiple regression in Table 7.6 against the corresponding participants’ judgements. 

7.7 Input tension profiles *Morpher* is fed with. 

7.8 Graphical representation of the cliffhanger profiles regressions’ predictions against participants’ judgements. 

7.9 Graphical representation of the decreasing profiles regressions’ predictions against participants’ judgements. 

7.10 Graphical representation of the surprise profiles regressions’ predictions against participants’ judgements. 

7.11 Graphical representation of the unresolved profiles regressions’ predictions against participants’ judgements.
A.1 Score of the four themes generated with Themer that were used in the first empirical study. ................................................................. 261
A.2 Score of the four pieces generated with Morpher, based on the cliffhanger profile, that were used in the second empirical study. ................. 262
A.3 Score of the four pieces generated with Morpher, based on the decreasing profile, that were used in the second empirical study. ................. 263
A.4 Score of the four pieces generated with Morpher, based on the surprise profile, that were used in the second empirical study. ................. 264
A.5 Score of the four pieces generated with Morpher, based on the unresolved profile, that were used in the second empirical study. ................. 265
A.6 Information about the empirical study participants were first presented with. 267
A.7 Consent form participants were presented with and must agree with to proceed. 268
A.8 Questionnaire participants were presented with to collect data about them. . 269
A.9 Interface participants were presented with to record their judgements of tension of the twenty pieces of music used in the empirical study. ................. 269
A.10 Average judgements of tension, of all 85 participants, corresponding to the four themes used in the first study. ................................. 271
A.11 Average judgements of tension, grouped by training, corresponding to the four themes used in the first study. ................................. 272
A.12 Average judgements of tension, of all 85 participants, corresponding to the cliffhanger pieces used in the second study. ....................... 273
A.13 Average judgements of tension, grouped by training, corresponding to the cliffhanger pieces used in the second study. ....................... 274
A.14 Average judgements of tension, of all 85 participants, corresponding to the decreasing pieces used in the second study. ....................... 275
A.15 Average judgements of tension, grouped by training, corresponding to the decreasing pieces used in the second study. ....................... 276
A.16 Average judgements of tension, of all 85 participants, corresponding to the surprise pieces used in the second study. ....................... 277
A.17 Average judgements of tension, grouped by training, corresponding to the surprise pieces used in the second study. ................................. 278
A.18 Average judgements of tension, of all 85 participants, corresponding to the unresolved pieces used in the second study. ......................... 279
A.19 Average judgements of tension, grouped by training, corresponding to the unresolved pieces used in the second study. ......................... 280
List of Tables

1.1 Main tasks, and their motivations, that Pearce et al. (2002) identify with the approaches for the automation of the compositional process, as well as some tension-driven examples. ................................................................. 2

2.1 Connections between our definition of musical tension and the main interpretations in the literature. ................................................................. 17

2.2 An example of the application of Narmour’s model according to Figure 2.2. . 23

3.1 Stepwise calculation of the degrees of sequential tension in the first phrase of the theme in Mozart’s *Ah vous dirai-je, Maman*, K. 265/300e. ............... 78

3.2 Stepwise calculation of the degrees of hierarchical tension in the first phrase of the theme in Mozart’s *Ah vous dirai-je, Maman*, K. 265/300e. .............. 80

3.3 Stepwise calculation of the degrees of attraction in the first phrase of the theme in Mozart’s *Ah vous dirai-je, Maman*, K. 265/300e. ..................... 82

4.1 Test cases used to evaluate the accuracy of *AutoTPS*. .............................. 110

4.2 Test cases used to evaluate the accuracy of *AuToTen*, which include the degrees of surface tension, hierarchical tension and attraction in the *Grail* theme from Wagner’s *Parsifal* transcribed from Lerdahl and Krumhansl (2007, p.343). . 111
4.3 Test cases used to evaluate the accuracy of AuToTen, which include the
degrees of surface tension, hierarchical tension and attraction in Bach’s chorale
Christus, der ist mein Leben transcribed from Lerdahl and Krumhansl (2007,
pp.344-345). ................................................................. 111

6.1 PB’s harmonic template given an N-measure theme. .......................... 142
6.2 CS’s harmonic template given an N-measure theme. ......................... 143
6.3 Allowed ranges of pitches pre-defined in Themer according to the input character.160
6.4 Example of the development that Developer will generate when Figure 6.3 is
the structure of the input given theme, when Subordinate Chords are left-
branching subordinates (i.e. those in Figure 6.21a) and when the transforma-
tion applied is the inversion. ............................................ 179

7.1 A summary of the average correlation coefficients, $r_{av}$, and the percentage of
significant regressions, $\%$sig, between the tension profiles of IGAs one hun-
dred pieces and the degrees of tension read by autognomus in real time. .... 212
7.2 A summary of the participants that took part in the empirical study grouped
by gender and years of musical training. ..................................... 214
7.3 Criteria used to determine the type of Intra-class Correlation Coefficient (ICC)
one needs according to Koo and Li (2016); Perinetti (2018). .................. 219
7.4 Degrees of inter-subject agreement for all eighty-five participants across the
four themes used in the study. ............................................ 220
7.5 Results of the independent regressions of the participants’ judgements of ten-
sion, of the four generated themes, against $cont_{melody}$, $cont_{bass}$, and $cont_{soprano}$.224
7.6 Results of the multiple regressions, that include the corresponding salient fea-
tures, of the four generated themes. ..................................... 225
7.7 A summary of the participants that took part in the empirical study grouped
by gender and years of musical training (repeated table). ...................... 227
7.8 Degrees of inter-subject agreement for all eighty-five participants across the
sixteen pieces used in the study. .......................................... 232
7.9 Results of the linear regressions of the average participants' judgements against the input tension profiles.
List of Acronyms

**GTMM**  Generative Theory of Tonal Music . . . . . . . . . . . . . . . . . . . ix
**TPS**  Tonal Pitch Space . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . x
**MTT**  Model of Tonal Tension . . . . . . . . . . . . . . . . . . . . . . . . . . . x
**IGA**  Interactive GTTM Analyser . . . . . . . . . . . . . . . . . . . . . . . . . 93
**ICC**  Intra-class Correlation Coefficient . . . . . . . . . . . . . . . . . . . . . xxiv
Chapter 1

Introduction

In 1957, Hiller and Isaacson (1957) published the score of the *Illiac Suite*, a four-movement suite for string quartet, which is considered to be the first piece of music automatically generated by a computer. Since then, many other authors and composers have investigated the automatic generation of music using computers. In the last two decades, a particular interest has grown with regards to the automatic generation of music driven by tension. According to Papadopoulos and Wiggins (1999), most of the music automatically generated by computers in the late twentieth century tended to lack specific intention and so could sound meaningless. To allow human composition to be simulated more closely, they proposed that music generation systems should evaluate and refer to a specific feature, such as musical tension.

Musical tension is considered to be an essential tool for evoking emotions (Herremans et al., 2017). Tension-driven music generation systems exist, but they usually focus on a particular task, such as imitating a specific style or generating the soundtrack of a single video game, while focusing on tension. Although the implementation of these systems has contributed to a better understanding of the problem of automatically generating music with regards to tension, there still exists a gap when it comes to finding different approaches which might generalise the solution to the problem at hand. Finding these approaches could translate into the application of tension-driven generation systems beyond academia, in our day-to-day lives. Some of these applications might in-
clude assisting composers with their compositional process, generating music for video creators, providing inspiration for novice musicians or generating music which adapts to our personal needs and tastes.

In this dissertation, we will investigate why it is challenging to automatically generate music with regard to tension. We will then explore how to overcome these challenges.

1.1 | Problem domain

Tension-driven automatic music generation

According to Pearce et al. (2002), approaches for the automation of the compositional process can be identified with at least four different tasks. Table 1.1 includes these tasks, and their motivations, as proposed by Pearce et al., as well as a collection of systems that illustrate the application of each task in the context of the automatic generation of music with regards to tension.¹

Table 1.1: Main tasks, and their motivations, that Pearce et al. (2002) identify with the approaches for the automation of the compositional process, as well as some tension-driven examples.

<table>
<thead>
<tr>
<th>Approach for the automation of the compositional process</th>
<th>Task</th>
<th>Motivation</th>
<th>Tension-driven example</th>
</tr>
</thead>
<tbody>
<tr>
<td>algorithmic composition</td>
<td>expansion of the compositional repertoire</td>
<td>Connectionist Approach to Driving Chord Progressions Using Tension (Melo and Wiggins, 2003)</td>
<td></td>
</tr>
<tr>
<td>design of compositional tools</td>
<td>development of tools for composers</td>
<td>Hyperscore (Farbood, 2011)</td>
<td></td>
</tr>
<tr>
<td>computational modelling of musical styles</td>
<td>proposal and evaluation of theories of musical styles</td>
<td>Mezzo (Brown, 2012)</td>
<td></td>
</tr>
<tr>
<td>computational modelling of music cognition</td>
<td>proposal and evaluation of cognitive theories of musical composition</td>
<td>MorpheuS (Herremans and Chew, 2017)</td>
<td></td>
</tr>
</tbody>
</table>

The tasks in Table 1.1 may be combined with and/or may concern other disciplines beyond academia. In Pearce et al. (2002), this approach is labelled as “interdisciplinary

¹The MorpheuS system in Table 1.1 is, in turn, based on the computational modelling of music cognition presented in Herremans and Chew (2016).
research”. Concerning this type of research, there has been, in the last decade, a special interest in the community towards the automatic generation of music that adapts to a given narrative (Herremans et al., 2017), such as those of films and video games. According to Collins (2009, p.5):

> video games are perhaps an ideal media form for procedural music; after all, many elements of gameplay (...) are unpredictable and occur in a non-linear fashion.²

Because of games’ non-linearity, among other reasons, the use of automatically generated music for games has rapidly increased and is currently seen as one of the most viable approaches in the future of game music (Plut and Pasquier, 2020). What is more, automatically generated music for games has been shown to increase the players’ experienced tension during game-play (Plut and Pasquier, 2019). This should not be a surprise as sound designers tend to decide on the emphasis and importance of game sounds based on points of tension and release within the game (Collins, 2008, p.92).

The benefits introduced in the previous paragraph have increased interest in interdisciplinary research on game-related systems that specifically focus on tension. These systems usually generate music as either adaptations or variations of pre-existing music, as for example in Sonancia (Lopes et al., 2016), or as whole new compositions, as for example in Escape Point (Prechtl, 2016).

Despite the potential benefits of automating the compositional process, we are still not using computer generated music in our day-to-day lives. To make it happen, according to Herremans et al. (2017), the newly composed music should match higher-level concepts. They suggest that tension-driven generated music could fill this gap by supporting a narrative, which would provide music generation systems with the agency they currently lack. But, how can tension be used to fill this gap?

²It is important to note that, to Collins, “procedural music” specifically refers to rule-based automatic music generated in real time.
1.2 | Motivations

State-of-the-art challenges

In Section 1.1, tension has been identified as one of the main areas of improvement to automating the compositional process. But what does the concept of tension actually entail in the context of music? Existing tension-driven music generation systems, such as those in Table 1.1, implement tension in many different ways. For instance, in Morpheus, the degree of tension of a group of notes, in a given piece of music, is associated with its dispersion in the tonal space, to its distance to the piece's key and to its movement in the “spiral array” (Chew's (2014) model of tonality). In Mezzo, tension depends on the chords' intervals and root motion, inspired by Cope's (2005) theories. In Escape Point, tension inversely correlates with Piston's (1959) usual chord progressions; the less probable a progression, the tenser.

Although being different, the above interpretations of tension produced pieces of music that successfully satisfied the needs of each research project. However, to generalise and better understand how to use tension to automate the compositional process, the concept of tension needs to be thoroughly dissected. Thus, the first challenge that motivates this dissertation is to frame the concept of musical tension based on a critical review of its different interpretations in the literature.

In the last two decades, many scholars have addressed this challenge, in one way or another, to investigate the automatic generation of music. However, by only focusing on this challenge, one takes the risk of producing music which only has theoretical interest. This often happens when the generated music turns out to be repetitive, contains abrupt transitions or lacks a well-defined structure. Thus, automatically generated music should also have a coherent musical discourse. The literature refers to this intra-piece coherence in different ways, such as large-scale form (Aspromallis and Gold, 2016), progression and development (Carnovalini and Rodà, 2020), large-scale structure (Hall and Pearce, 2021) or long-term structure (Herremans et al., 2017). The latter, long-term structure,\(^3\)

---

\(^3\)Such is the impact and importance of long-term structure that it is one the key features in state-of-the-art commercial music generators, such as AIVA (Aiva-Technologies, 2016) or Google's Magenta (Google-
Automatically generating music with long-term structure entails finding or describing patterns, themes or motifs within the musical context. Systems that focus on the automatic generation of music that has long-term structure are often designed for a specific application, such as generating danceable music (Anderson et al., 2013), modelling musical expression and performance (Carnovalini and Rodà, 2019), supporting interactive experiences (Aspromallis, 2020) or matching tension (Herremans and Chew, 2017), among others. However, mainly because of the complexity of framing musical structure, the specific application is often treated as a secondary feature and gets subjugated by a stronger focus on long-term structure.

In this dissertation, the specific application concerns the generation of music that matches a given tension profile. Thus, the second challenge that motivates this dissertation is to find a balanced focus between long-term structure and tension when investigating new ways to automatically generate music.

Interest in automating of the compositional process has increased following the recent growth of narrative-based experiences, which often include improvised content or depend on the actions of participants (Aspromallis and Gold, 2016). The literature refers to these narratives as being non-linear, meaning that the timing of events and actions are unpredictable (Collins, 2009). For the music that supports these experiences to be as immersive as possible, it should be able to adapt to the narratives in real time. However, the possibility of generating music in real time depends on the models on which the different generation systems are based. Thus, the third challenge is to combine the model of musical tension and the computational method for automatically generating music in a way that is applicable in real time.

---

4These experiences include live improvisations and performances (Eigenfeldt, 2014), live coding (Brown and Sorensen, 2009), interactive concerts (Liem et al., 2015), virtual environments (Robertson et al., 1998), interactions with the natural environment (Nikolaidis et al., 2012), interactive films (Casella and Paiva, 2001) and video games (recall Sonancia (Lopes et al., 2016) and Escape Point (Prechtl, 2016)).
1.3 | Research approach

Goals, questions and objectives

In Section 1.2, three main challenges have been identified concerning the automatic generation of music with regards to tension. Putting the focus on overcoming these challenges means pursuing three specific goals. These three goals focus on automatically generating music which:

(i) matches input tension profiles,

(ii) has long-term structure, and

(iii) is generated in real time.

Now that our goals have been defined, it is possible to fully frame the scope of this dissertation. As shown in Figure 1.1, this dissertation focuses on the computational techniques within the field of algorithmic composition where tension, long-term structure and real time meet.

Figure 1.1: An outline of the scope of this dissertation.
The pursuit of our goals, while considering the challenges proposed in Section 1.2, can be phrased in the form of the following Research Question:

**Research Question**

*How can tonal music be generated in real time so that it has long-term structure and matches a given tension profile?*

In order to address our Research Question, the research presented in this dissertation adopts the following objectives:

- to frame the scope of and to define the concept of musical tension, and to develop a taxonomy of the main approaches to modelling musical tension,
- to frame the scope of and to define the concept of automatic music generation, and to develop a taxonomy of the main methods to automatically generate music, and
- to design and to implement a system capable of generating tonal music that has long-term structure and matches input tension profiles in real time, and to evaluate the capabilities and limitations of the system.

### 1.4 | Dissertation outline

Chapter 1 has framed the scope of the dissertation and has defined our goals and Research Question.

Chapter 2 identifies the model of tension best suited to our goals. To do so, Chapter 2 takes the following steps: framing and defining the concept of musical tension, reviewing the main approaches to modelling the concept, identifying the requirements that a model of musical tension should meet so that it suits our goals and determining the models of musical tension that best suit the identified requirements.

Chapter 3 reviews the model of tension identified as the one best suited to our goals. To do so, Chapter 3 takes the following steps: contextualising the model, reviewing its components and its evaluation, and providing an illustrative example of the calculation of tension according to the model.
Chapter 4 automates the model of tension identified as the one best suited to our goals. To do so, Chapter 4 takes the following steps: identifying the challenges concerning the automation of the model, implementing the model computationally and evaluating the accuracy of the implementation.

Chapter 5 determines the computational methods best suited to our goals. To do so, Chapter 5 takes the following steps: defining and framing the scope of the concept of Automatic Music Generation, reviewing the main methods in the field and identifying which methods are best suited to our goals, and determining how to apply the identified computational methods so that our goals can be met.

Chapter 6 designs and implements a system capable of generating tonal music that has long-term structure and matches input tension profiles in real time. To do so, Chapter 6 takes the following steps: determining the compositional strategy best suited to our goals, designing the architecture of a music generation system to answer our Research Question, and designing and implementing the components of the music generation system.

Chapter 7 evaluates the performance and applicability of the music generation system. To do so, Chapter 7 takes the following steps: defining the methodology for evaluating the performance and applicability of the music generation system with regards to our goals, designing the appropriate tests according to the defined evaluation methodology, presenting the procedures and the results of the tests, and analysing and discussing the impact of the results with regards to our Research Question.

Chapter 8 draws conclusions from the conducted research. To do so, Chapter 8 takes the following steps: summing up the conducted research, discussing the main conclusions acquired from our research, pointing out the main research contributions, discussing the limitations of the conducted research and proposing directions for future work.

At the end of this dissertation there is an appendix that includes the experimental materials and some additional data used in Chapter 7, a glossary of musical terms (the reader should note that this dissertation uses American music-theory terminology) and a list of references.
Chapter 2

Modelling Musical Tension

Scope, definition and models

In this chapter, we identify the model of tension best suited to our goals. To do so, we will:

(1) frame and define the concept of musical tension,

(2) review the main approaches to modelling the concept of musical tension,

(3) identify the requirements that a model of musical tension should meet so that it suits our goals, and

(4) determine the models of musical tension that best suit the identified requirements.

The first step is addressed in Section 2.1, the second step is addressed in Sections 2.2, 2.3 and 2.4, and the third and fourth steps are addressed in Section 2.5.
2.1 | Framing the scope of the study of musical tension

Music, in the Western tonal tradition, is often described by means of patterns of tension and relaxation (Krumhansl, 1996). This description of music was adopted in the musical analysis treatises of the early years of the twentieth century, influenced by Gestalt psychology (Bent and Drabkin, 1987).¹

Such is the importance of tension in music that many scholars consider it a key underpinning element of music psychology that is essential to the listening experience. For instance, according to Granot and Eitan (2011, p.219):

[tension is] one of the most fundamental concepts in the analytical, aesthetic and psychological discourse on music.

Having said that, what do we exactly mean by tension in the context of music? The truth is that, despite the fact that tension may play a key role in the psychological understanding of music, there is still no agreed definition. Why is that?

2.1.1 | Why is it so challenging to define the concept of musical tension?

Musical tension is an ill-defined concept. To support this claim, let us consider the following scenario: a Westerner is listening to a sequence of chords that ends on a perfect cadence. What degree of tension may the listener perceive at the end of the cadence?

On one hand, one could follow the definition of tension provided by Bigand and Parn-cutt (1999, p.242):

¹According to Rothfarb (1991), one of the key contributions to defining a modern music psychology was the work of Ernst Kurth, which shows a strong affinity with Gestalt musical principles. Kurth (1917) conceives three levels of activity in the creation of music: a will or energy, carried by melodic, harmonic and rhythmic elements, which drives music forward and defines coherent units within the musical discourse; a “play of tensions”, drawn from the unconscious mind, describing the music’s arcs of growth; and the conscious understanding of the “play of tensions” once it takes form with regards to the musical sound.

Other music analysts have described music in a similar way, supporting these claims. For example, Kurth’s first level of activity matches the idea of “musical force”, understood by Zuckerkandl (1956) as a dynamic component, known as force, equilibrium, tension or direction, which pushes music forward. Kurth’s second and third levels of activity can be linked to the work of Meyer (1956, p.43), to whom “[the] understanding and enjoyment [of music] depend upon the perception of and response to attributes such as tension and repose, instability and stability, and ambiguity and clarity”.

10
strong musical tension at the end of a fragment evokes the feeling that there must be a continuation of the sequence, [whereas] low musical tension evokes the feeling that the sequence could naturally stop at this point.

The above definition relates to the cognitive approaches that emphasise the role tonal function may play in perceiving musical tension. These approaches mainly focus on the hierarchies within a sequence of chords and on the perceptual distances between the sequence’s chords.

According to Bigand et al. (1996, pp.126-128):

important events in a tonal hierarchy instill weak or null musical tension and less important ones create strong musical tension (...) [and] greater spatial distances between musical events correspond to greater degrees of perceived musical tension.

The last chord in the sequence presented in the above scenario may be an important one in the sequence's hierarchy, since it is a tonic chord where the sequence will repose. Likewise, in most psychological representations of the tonal space, the transition from a dominant chord to a tonic chord is associated with one of the smallest distances within the space.\(^2\) Therefore, following this line of thought, the listener may perceive a low degree of tension associated with the last chord in the cadence.

On the other hand, one could follow the definition of tension provided by Huron (2006, p.421), to whom tension describes:

those feelings that arise immediately prior to an expected event.

The biggest difference between Bigand and Parnscutt’s definition of tension and that of Huron’s is that the former considers how listeners may respond after the event of study has happened, whereas the latter considers the listeners’ response before the event of study has happened. Taking that into account, in the last chord of the sequence in the above scenario, which is the event of study, Huron’s interpretation of tension will

\(^2\)See Krumhansl (2001) for more detail.
focus on the impact the dominant chord may have upon the tonic chord, with regards to tension.

Listeners may expect the dominant chord to resolve to a tonic chord, mainly because of feeling the need of resolving the leading tone, included in the dominant chord, to the root of the tonic chord. These expectations would probably be greater in the case of a dominant seventh chord, where listeners may also feel that the dominant’s seventh tone needs to resolve to the tonic’s third, as well as because of the existence of a tritone relation between the leading tone and the seventh of the dominant chord. According to Huron (2006, pp.306-307), listeners may perceive a rise in tension because of this expectation of resolution. Therefore, following this line of thought, the listener may perceive a high degree of tension associated with the last chord in the cadence.3

Notice how, for the same scenario, two opposite conclusions have been drawn above. Based on the work by Bigand and Parncutt (1999), it was concluded that the end of a perfect cadence may be associated with low tension. However, based on the work by Huron (2006), the opposite conclusion was reached, associating the end of a perfect cadence with high tension. Why is this?

The main reason two opposite conclusions were drawn above is that Bigand and Parncutt (1999) and Huron (2006) use the term “tension” to describe two different processes. The former conceive tension as a post-outcome response whereas the latter conceives tension as a pre-outcome response.

Some authors, such as Granot and Eitan (2011), hypothesise that, in a perfect ca-

---

3 Some readers may feel this conclusion is somewhat counterintuitive. Most likely this will happen because of the different interpretations each of us has when differentiating between concepts such as tension, expectation, surprise or stress. However, this dilemma has been introduced here motivated by a discussion with an anonymous contributor (i.e. conference reviewer) in June 2019, whose argument was the following:

I tend to agree more with the Huronian view that tension is indeed increased when expectation is maximum (and the resolution is ideally delayed). According to this view, \( V \rightarrow I \) is the transition with the highest amount of tension and not the most relaxing. Note that tension rises before \( I \) is reached and the resolution is indeed relaxing, a fact that explains why relaxing is being reported eventually. I suspect that (...) participants tend to confuse ‘tension’ with ‘stress’ and report these terms interchangeably. So for example, a transition \( I \rightarrow V \) is the most surprising and therefore if one is asked for its tension amount will probably report a high tension when actually he/she will mean high surprise, (i.e., stress). It is indeed very difficult to empirically discriminate between tension and surprise induced stress.
dence, a Westerner will perceive a rise in tension when the dominant chord is being played, as expectations for resolution will rise, and will perceive that tension falls when the tonic chord is played, as expectations will be realised. The truth is that both Bigand and Parncutt’s and Huron’s approaches would describe the musical forces involved in a perfect cadence in a similar way, analogous to the rise and fall description provided by Granot and Eitan. Although, in these descriptions, the term tension would be used to express different phenomena.

In the case of Huron (2006), his work defines tension as a pre-outcome response. However, his work also considers post-outcome responses, to which he refers with the terms “prediction” and “reaction”. Prediction concerns feelings, either positive or negative, that are evoked in response to the relative success or failure in predicting some outcome. Reaction concerns fast responses, immediately evoked after an outcome, which commonly assume a worst-case assessment of the outcome. Therefore, in a perfect cadence, a Huronian approach will probably associate the dominant chord with a high expectation of resolution and the tonic chord with a low expectation of ensuing events. According to Huron, these expectations will translate into a high-degree tension response, before the tonic chord (i.e. pre-outcome), and a low-arousal prediction/reaction response, after the tonic chord (i.e. post-outcome). Notice how, although the second response does not explicitly use the term tension, the combination of both responses (tension and prediction/reaction) does match Granot and Eitan’s rise and fall hypothetical behaviour concerning a perfect cadence.

In the case of Bigand and Parnutt (1999), their work defines tension mostly in the context of harmonic vertical motion (i.e. the chord’s hierarchical importance and their distances in the tonal space). This view is shared by other scholars, such as Lerdahl (1988; 1996; 2004; 2007). Lerdahl uses the term “tension” as a post-outcome response that concerns the distances between chords in the tonal space. However, his work also considers the influence that horizontal motion (i.e. melodic arrangement) may have on the conveyance of tension. He studies this other contribution using the term “attraction”. The already discussed effects of the leading tone, as well as the seventh tone in the case of a dominant seventh chord, will result in an increase of the perceived attraction, before
Chapter 2. Modelling Musical Tension

Framing the scope of the study of musical tension

the tonic chord (i.e. pre-outcome), followed by a tension decrease, after the tonic chord (i.e. post-outcome), associated with a close transition in the tonal space. Notice again how, although the first response does not explicitly use the term tension, the combination of both responses (attraction and tension) does match Granot and Eitan's rise and fall hypothetical behaviour concerning a perfect cadence.

There are two key messages to be taken away from the above discussion. First, the term “tension” is sometimes used in the literature to refer to different processes. This issue could result in misunderstandings and so tension was described as an ill-defined concept at the beginning of this section. Second, despite this trade-off, most approaches, such as Bigand and Parnscutt's and Huron's, will probably result, in the end, in similar descriptions concerning perceived musical forces, although different terms might be used in such descriptions.

In order to overcome the lack of agreement concerning the definition of the concept of tension, a new definition that encompasses all descriptions and points of view is needed. Recall that this generalisation was phrased in Section 1.2 as the first challenge to be met.

2.1.2 | Towards a definition of the concept of musical tension

In the music-related literature, there is no agreement upon the use of the term tension. Some scholars use it to describe the behaviour of a very specific musical force while others use it to describe a broader range of behaviours. In order to generalise the definition of the concept of musical tension, we grouped together the main forces that may be associated with musical tension into a single definition as follows:

### Defining musical tension: first attempt

**Musical tension** refers to the reaction listeners experience as a consequence of the expectations generated by the musical discourse, either when these expectations emerge, cease, change, evolve, are fulfilled, are violated, are lacking, are unclear, concern stability or trigger some non music-related emotional responses.

Expectations that *emerge* concern feelings of prediction or anticipation. For instance,
a dominant chord may result in the emergence of the expectation that it will resolve into a tonic chord, and so listeners may perceive a rise in tension associated with this feeling of anticipation.

Expectations that cease concern the termination of the feelings immediately evoked after an expected outcome. For instance, the end of a musical phrase may result in the cessation of preceding expectations if listeners feel the music could or should naturally stop, and so they may perceive a fall in tension associated with the expectation of no continuation.

Expectations that change concern the modifications of immediately preceding expectations. For instance, adding a seventh to a preceding dominant chord (i.e. \( V \rightarrow V^7 \)) may result in the change of preceding expectations, as the feeling of resolving to the tonic will probably be greater, and so listeners may perceive a rise in tension associated with this greater feeling of anticipation.

Expectations that evolve concern long-term feelings of reinforcement caused by complex assessments which are influenced by the musical context. For instance, a pattern or theme that is presented as a main idea in the musical discourse may result in the evolution of expectations if the main idea gets transformed once listeners had internalised it, and so they may perceive a change in tension associated with this transformation.

Expectations that are fulfilled concern the realisation of preceding predictions. For instance, a tonic chord after a dominant chord may result in the fulfilment of preceding expectations, since it was the outcome listeners probably anticipated, and so they may perceive a fall in tension associated with the realisation of their expectations.

Expectations that are violated concern the denial of preceding predictions. For instance, a deceptive cadence may result in the violation of preceding expectations, as the cadence’s last chord may not be the outcome listeners probably anticipated, and so they may perceive a rise in tension associated with the denial of their expectations.

Expectations that are lacking concern the inability to generate any clear expectation. For instance, a musical sequence that abruptly stops may result in lack of expectations, as listeners may not be able to anticipate what might be coming next, and so they may perceive a rise in tension associated with the non-existence of expectations.
Expectations that are unclear concern equally probable continuations or a too irregular context. For instance, a musical language unfamiliar to listeners may result in unclear expectations, and so they may perceive a rise in tension associated to confusing expectations.

Expectations that concern stability relate to the role each note and chord may play in the musical context. For instance, a chord transition between two chords in distant keys may result in states of instability, and so listeners may perceive a rise in tension associated with the chord transition.

Finally, there may be expectations that trigger a non music-related emotional response. For instance, a piece of music that is part of the soundtrack of a movie and is played during an emotional scene may trigger some expectations when played again, as it might remind listeners of a particular emotional scene in the movie, and so they may perceive a change in tension associated with these emotion-related expectations.

Table 2.1 connects our definition of tension with the literature’s main interpretations of the concept of tension. The table relates expectations to either feelings or to components of models of tension/expectation.

To sum up, we have proposed a definition of tension that generalises the concept by encompassing the main interpretations that exist in the literature, as shown in Table 2.1. In Section 2.1.3, we complete our definition of tension.

2.1.3 | Defining the concept of musical tension

There is a key issue that needs to be mentioned concerning Section 2.1.2’s definition of tension. Notice that it implies that any musical expectation could be related to a tension response. This implication presents some advantages and some disadvantages. On one hand, the definition captures most of the descriptions of tension provided in the literature. Thus, it may be of use to the community. On the other hand, the definition may come into conflict with those of other scholars. Some authors meticulously distinguish between concepts such as tension, surprise, suspense or stress. However, what our definition is doing is putting tension on top of all these concepts and encompassing them
Table 2.1: Connections between our definition of musical tension and the main interpretations in the literature.

<table>
<thead>
<tr>
<th>Expectations that:</th>
<th>Relate to:</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>feelings or component of a model</td>
<td>reference</td>
</tr>
<tr>
<td><strong>emerge</strong></td>
<td>prediction and anticipation</td>
<td>Lehne and Koelsch (2015)</td>
</tr>
<tr>
<td></td>
<td>gravity, magnetism and inertia</td>
<td>Larson (1993; 2004)</td>
</tr>
<tr>
<td></td>
<td>tension response</td>
<td>Huron (2006)</td>
</tr>
<tr>
<td></td>
<td>expectation-tension</td>
<td>Margulis (2005)</td>
</tr>
<tr>
<td></td>
<td>attraction response</td>
<td>Lerdahl (2004)</td>
</tr>
<tr>
<td><strong>cease</strong></td>
<td>reaction response</td>
<td>Huron (2006)</td>
</tr>
<tr>
<td></td>
<td>expectancy of no continuation</td>
<td>Bigand and Parncutt (1999)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Melo and Wiggins (2003)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Larson (1993; 2004)</td>
</tr>
<tr>
<td><strong>change</strong></td>
<td>dissonance</td>
<td>Lehne and Koelsch (2015)</td>
</tr>
<tr>
<td></td>
<td>surface tension</td>
<td>Lerdahl (2004)</td>
</tr>
<tr>
<td><strong>evolve</strong></td>
<td>imagination and appraisal responses</td>
<td>Huron (2006)</td>
</tr>
<tr>
<td></td>
<td>structural expectation</td>
<td>Lerdahl and Jackendoff (1983)</td>
</tr>
<tr>
<td><strong>are fulfilled/violated</strong></td>
<td>implication-realisation</td>
<td>Narmour (1990)</td>
</tr>
<tr>
<td></td>
<td>implicative denial</td>
<td>Lerdahl (2004)</td>
</tr>
<tr>
<td></td>
<td>expectancy-denial</td>
<td>Margulis (2005)</td>
</tr>
<tr>
<td></td>
<td>prediction response</td>
<td>Huron (2006)</td>
</tr>
<tr>
<td></td>
<td>surprise</td>
<td>Margulis (2005)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Huron (2006)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lehne and Koelsch (2015)</td>
</tr>
<tr>
<td><strong>are lacking</strong></td>
<td>suspense</td>
<td>Meyer (1956)</td>
</tr>
<tr>
<td></td>
<td>uncertainty</td>
<td>Lehne and Koelsch (2015)</td>
</tr>
<tr>
<td><strong>are unclear</strong></td>
<td>conflict and uncertainty</td>
<td>Lehne and Koelsch (2015)</td>
</tr>
<tr>
<td></td>
<td>regularity of the musical context</td>
<td>Granot and Donchin (2002)</td>
</tr>
<tr>
<td><strong>concern stability</strong></td>
<td>tonal stability</td>
<td>Lerdahl (2004)</td>
</tr>
<tr>
<td></td>
<td>expectancy-tension</td>
<td>Margulis (2005)</td>
</tr>
<tr>
<td></td>
<td>dissonance and instability</td>
<td>Lehne and Koelsch (2015)</td>
</tr>
<tr>
<td><strong>non music-related response</strong></td>
<td>appraisal response</td>
<td>Huron (2006)</td>
</tr>
<tr>
<td></td>
<td>emotional events</td>
<td>Lehne and Koelsch (2015)</td>
</tr>
</tbody>
</table>

all. So, why is this definition of tension preferred in this dissertation over other existing definitions?

The definition of tension previously introduced in Section 2.1.2 seems to be the one that best suits our goals. Although from a theoretical or musicological point of view it might be possible to easily differentiate between concepts such as tension, surprise, suspense or stress, the truth is that, perceptually, it is not an easy task at all. Recall that this dilemma was already introduced in footnote 3 (see page 12), where an anonymous contributor claimed that, in certain scenarios, a listener “will probably report a high tension when actually he/she will mean high surprise, (i.e. stress)”. This assertion seems to
Chapter 2. Modelling Musical Tension

Framing the scope of the study of musical tension

Imply that listeners are somehow mistaken when they refer to the tension they perceive; but, should not a theoretical interpretation of tension match what listeners perceive and not the other way around? This is why Section 2.1.2’s definition of tension is preferred over others, as it will probably account for most of the different interpretations listeners may have of the concept of tension.

Concerning Section 2.1.2’s definition of tension, there is a final note to be made. In the literature, approaches for the study of musical tension are typically classified as based on either cognitive or psychoacoustic features (Bigand et al., 1996; Granot and Eitan, 2011; Herremans and Chew, 2016). Most of the features that have been mentioned in this section are cognitive ones (e.g. tonal hierarchy, distance between chords in the tonal space, tonal stability, melodic attraction, etc.). Examples of psychoacoustic features that might have an effect on perceived musical tension include roughness levels, loudness changes, temporal density or tempo changes, among others. As these have not been introduced until now, a question to consider is whether our definition of tension will account for the tension changes produced by psychoacoustic features.

Granot and Eitan (2011, p.129) identify psychoacoustic features with the psychological dimensions of arousal and activity, and so conclude that tension is “not necessarily related to expectation”. Thus, our definition of tension can be completed as follows:

**Defining musical tension**

*Musical tension* refers to the reaction listeners experience as a consequence of the expectations generated by the musical discourse (either when these expectations emerge, cease, change, evolve, are fulfilled, are violated, are lacking, are unclear, concern stability or trigger some non music-related emotional responses), as well as a consequence of any modifications in the music’s acoustic signal that may change the perceived levels of arousal and activity.
2.1.4 | A taxonomy of the approaches to modelling musical tension

As concluded in Chapter 1, this dissertation aims to develop a system capable of generating tonal music that has long-term structure and that matches input tension profiles in real time. To do so, a model of tension, or a collection of models, should be identified as the most suitable for that purpose.

A model of tension will suit our goals if it meets a collection of requirements. The upcoming sections aim to identify these requirements. There may exist a model that refers to musical expectations but does not mention tension at all (e.g. Larson’s (1993; 1997; 2004; 2005) model of musical forces). According to our definition of tension, such a model could still suit our goals. Thus, the first requirement a model of tension should meet, to suit our goals, can be put as:

(i) the model must be consistent with our definition of tension.

A first classification of the approaches to modelling musical tension and/or expectation was already introduced in Section 2.1.3. Recall that this classification, inspired by those in Bigand et al. (1996); Granot and Eitan (2011); Herremans and Chew (2016), distinguishes between pieces of research whose focus is put on either cognitive features, psychoacoustic features or both. Another classification could be proposed from the point of view of the methodology adopted by an approach, where approaches to modelling musical tension and/or expectation could be classified as theoretical accounts, empirical studies or generation models of tension and/or expectation. For the sake of completeness, in this dissertation, both classifications are merged into a single one, as shown in Figure 2.1.

![Figure 2.1: A classification of the approaches to modelling musical tension.](image-url)
In order to determine which approach best suits our goals, Sections 2.2, 2.3 and 2.4 review the scope of theoretical accounts, empirical studies and generation models of musical tension, respectively.

2.2 | Theoretical accounts of musical tension

Theoretical accounts of musical tension consist of well-defined frameworks that include collections of postulates which may help predict how tension behaves in certain scenarios. These accounts usually frame the scope of the concept of musical tension within a specific context. These contexts can go from a general perspective of music psychology to a particular focus on a single musical dimension. Examples of these contexts include the work of Meyer (1956), as a general theory of musical emotions and meaning, the work of Narmour (1990), as a theory of expectation in the context of melody, the work of Swain (1998), as a theory of tension in the context of harmonic rhythm, or the work of Lerdahl and Jackendoff (1983), as a theory of tension in the context of tonal structure. These theories are reviewed in Sections 2.2.1, 2.2.2, 2.2.3 and 2.2.4, respectively, to illustrate the scope of theoretical accounts of musical tension.

2.2.1 | Meyer’s theory of musical emotions and meaning

Meyer (1956) presents a theory of musical emotions and meaning, where expectations play a central role. To Meyer, emotional responses to music are temporary and evanescent feelings that emerge from a cognitive interpretation of musical stimuli. He believes that emotions in music are mostly based on listeners’ implicit musical knowledge,⁴ that they might change from person to person and that they communicate meaning.

Meyer distinguishes between two sources of musical meaning. On one hand, he believes music may acquire meaning from the music itself or through references to non-musical concepts, such as actions, emotional states or character.⁵ On the other hand, he

⁴In spite of this fact, Meyer’s theory does consider some innate processes from Gestalt psychology, such as grouping, closure, good continuation or gap fill.

⁵This idea is similar to the states of “emotional significance” in Lehne and Koelsch’s (2015) model of tension.
also believes that the source of meaning could either concern an intellectual understanding of music or may lay within the music itself, and its internal relationships.\(^6\)

In addition to the different sources of meaning, Meyer distinguishes between three different types of meaning. These types of meaning are attributed to: (i) pre-outcome expectations, such as the feeling of anticipation; (ii) post-outcome realisation or denial of preceding expectations; and (iii) long-term re-evaluation of both previous pre- and post-outcome responses and their relationships.

In Meyer’s theory, where emotion is the starting point of his discussion, it is possible to find a path from emotion, to meaning, to expectation. It is then not a surprise that Meyer thoroughly dissects the concept of expectation in his work. In doing so, he discusses ways in which listeners’ expectations may be violated, ways in which listeners may perceive the emergence of and the departure from uncertainty, ways in which listeners may be influenced by stylistic learned expectations, ways in which some music elements may produce an anchoring effect and many more behaviours concerning expectation. For more detail on his thoughts on musical expectations, see Meyer (1956).

### 2.2.2 | Narmour’s implication-realisation model

Narmour (1990) presents a theory of melodic expectation, known by the name of implication-realisation model. According to Narmour, his model describes listeners’ cognitive response to melodies. The implication-realisation model consists of two components: a top-down one, whose principles concern knowledge acquired through experience, and a bottom-up one, whose principles are unconscious and universal.

---

\(^6\) This distinction has had a crucial impact on empirical studies concerning emotion and meaning in music. For instance, Krumhansl (1997) argues that a basic issue about musical emotions is whether the music itself elicits emotions or just expresses the emotions recognised by listeners in the music. She refers to these positions by the names of “emotivist” and “cognitivist”, respectively. Lehne and Koelsch (2015, p.3) address the same issue, but with a different terminology. They distinguish between “perceived” emotions, meaning those emotions that listeners assume the music is supposed to express, and “subjectively felt” emotions, meaning the listeners’ actual emotional experiences. They illustrate this issue by means of an example:

someone might cognitively acknowledge that a specific movie scene is supposed to induce suspense, without subjectively feeling any suspense.

Similar descriptions of this issue have been given by other scholars, such as Gabrielsson (2001).
Chapter 2. Modelling Musical Tension

The bottom-up component of the implication-realisation model describes the cognition of melodies in terms of implication, realisation and closure, where strong feelings of closure imply weak feelings of melodic continuation.

For the sake of simplicity, we have summarised the main aspects of the bottom-up component in Narmour’s implication-realisation in Figure 2.2. We have designed this figure inspired by the summary of the implication-realisation model provided by Krumhansl (1995).

According to Narmour (1990), the melodic lines that fit the type 1 closure conditions in Figure 2.2 are associated with the strongest sense of closure.

An example of the different types of closure in Figure 2.2 is shown in Table 2.2.

For more detail concerning the implication-realisation model, see Narmour (1990).

2.2.3 | Swain’s dimensions of harmonic rhythm

Swain’s (1998) theory describes harmonic rhythm along the lines of six properties, which he calls dimensions: texture rhythm which records the fastest moving rhythms in any voice; root change density, which records the number of voices that are different between two chords; root/quality rhythm, which records any change of root and/or quality (i.e. major, minor, diminished, or augmented) between chords; bass pitch rhythm, which records the motion in the bass voice; phenomenal rhythm, which records any change in the combination of the tones sounding together, whether or not they pertain to a different chord; and harmonic function, which records patterns of change of the Riemannian functions in the piece.

To calculate tension, Swain (1998) proposes two assumptions. First, that one kind of tension rises with the increase of speed of changes in the model’s six dimensions of harmonic rhythm; that is to say, the more frequent the dimensions change, the tenser, and vice versa. Second, that another kind of tension rises with the increase of discrepancies (i.e. the independence of motions) among the six dimensions of harmonic rhythm; that is

---

7These are the tonic function, the subdominant function and the dominant function, which Swain (1998, p.65) describes as “the stable”, “the mediating” and “the mobile and tense”, respectively.
Chapter 2. Modelling Musical Tension

Theoretical accounts of musical tension

Figure 2.2: A flowchart of the bottom-up component of Narmour's implication-realisation model, for a given three-note melodic line, x-y-z. The flowchart focuses on the intervals between the notes in the given melodic line and the direction (i.e. ascending or descending) between consecutive notes.

Table 2.2: An example of the application of Narmour's model according to Figure 2.2.

<table>
<thead>
<tr>
<th>Type 1</th>
<th>Type 2</th>
<th>Type 3</th>
<th>Type 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strong closure</td>
<td>c4-b4-c5</td>
<td>c4-b4-a4</td>
<td>c4-f4-g4</td>
</tr>
<tr>
<td>Weak closure</td>
<td>c4-g4-d5</td>
<td>c4-g4-b3</td>
<td>c4-d4-e4</td>
</tr>
</tbody>
</table>
to say, the more different the speeds at which the dimensions move, the tenser, and vice versa.

For more detail concerning the analysis and interpretation of the six dimensions of harmonic rhythm, see Swain (2003, 2002).

2.2.4 | Lerdahl and Jackendoff’s Generative Theory of Tonal Music (GTTM)

Lerdahl and Jackendoff (1983) present a theory of tonal music concerning tension, where music structure plays a central role, known by the name of Generative Theory of Tonal Music (GTTM). The theory attempts to characterise musical structures within a hierarchy. To do so, it establishes a collection of principles by which a listener would infer such structures. These principles are given as a system of rules, associated with four main structures: the grouping structure, which organises music into smaller groups, such as motifs, themes, subjects or phrases; the metrical structure, which identifies the most probable patterns of strong and weak beats; the time-span reduction, which expresses the relative structural importance of the musical events; and the prolongational reduction, which reflects qualitative patterns of tension and relaxation in the musical discourse.

GTTM will not be further discussed here as it is reviewed in more detail in Chapter 3.

2.2.5 | A critical look at theoretical accounts of musical tension

Theoretical accounts of musical tension present some advantages and some disadvantages in the context of this dissertation. On one hand, they satisfy the first requirement an approach should meet to suit our goals (i.e. they are consistent with our definition of tension) considering how theoretical accounts of musical tension were defined at the beginning of Section 2.2. On the other hand, most theoretical accounts of musical tension present a qualitative\(^8\) description of the flow of tension in the musical discourse. It

\(^8\)Some authors have investigated the partial quantification of some of the theories reviewed in Section 2.2, such as Rosner and Meyer (1986), Lerdahl (1988), Krumhansl (1995) or Schellenberg (1996), among others.
is true that some of these theories base their postulates and hypotheses upon quantitative research. For instance, the work of Huron (2006), which could be described as a *theoretical account* of musical expectations, supports some of its claims on quantitative measurements provided by several studies. But, in the end, his work is mostly theoretical.

The qualitative nature of *theoretical accounts* of musical tension entails a disadvantage concerning our goals. In order to automatically generate music that matches a given tension profile, using an approach that models tension quantitatively will ease the task.

A direct consequence of the qualitative nature of *theoretical accounts* of musical tension is the difficulty of fully testing them empirically. However, in the context of this dissertation, it is preferred for a candidate model of tension to have been empirically tested and have shown strong correlations against human judgements of perceived musical tension. In this way, the evaluation of our music generation system can be based on certain ground-truth. Thus, two more requirements (i.e. (ii) and (iii)), which a model should meet to suit our goals, can be derived from this section’s discussion:

(i) the model must be consistent with our definition of tension,

(ii) the model must provide a method to calculate quantitative values of tension, and

(iii) the model must have been empirically tested and have shown strong correlations with the degrees of musical tension perceived by human listeners.

### 2.3 Empirical studies of musical tension

*Empirical studies* of musical tension use experimental paradigms in which explicit data is recorded from human subjects concerning the tension they perceive when listening to music. As shown in Figure 2.1, *empirical studies* of musical tension can be classified depending on whether they focus on either *cognitive features*, *psychoacoustic features* or *both*.

The most frequently investigated *cognitive features* concern harmony, melody, rhythm and structure. The main harmonic features include harmonic motion, both vertical (e.g.
distance between chords in the tonal space) and horizontal (e.g. voice-leading paths),
chords’ and notes’ degree of stability in the tonal context, and tonal function. The main
melodic features include intervallic difference, pitch proximity and contour.⁹ The main
rhythmic features include onset frequency, metre and metrical irregularities, and har-
monic rhythm. The main structural features include structural hierarchies and points of
segmentation in the musical discourse.

The most frequently investigated psychoacoustic features concern dynamics, agogics
and timbre. The main dynamic features include the influence of traditional dynamic
markings and loudness differences in a musical stimulus. The main agogics features in-
clude the effect of expressiveness and tempo changes. The main timbral features include
roughness and sensory dissonance, note density, texture, and orchestration.

Scholars usually investigate empirical studies of musical tension from the point of view
of the feature they focus on. Notice, however, that our goals focus on the methods rather
than the features. Therefore, this section reviews empirical studies of musical tension
from the perspective of the methodology they follow. For detailed analyses of the scope
of empirical studies of musical tension from the former point of view, see studies such
as: Bigand et al. (1996), regarding harmonic features; Pearce (2003), regarding melodic
features; Fernández-Sotos et al. (2016), regarding rhythmic features; Krumhansl (1996),
regarding structural features; Ilie and Thompson (2006), regarding dynamic and agogics
features; and Farbood and Price (2017), regarding timbral features.

From the point of view of the methodologies adopted by empirical studies of musical
tension, there exist three main research approaches. These approaches differ in the
type of data collected by a study, which could either be human judgements, physiological
responses or neural activity. These approaches are reviewed in Sections 2.3.1, 2.3 and
2.3.3, respectively, to illustrate the scope of empirical studies of musical tension.

⁹Some authors, such as Granot and Eitan (2011), describe contour as a psychoacoustic feature.
2.3.1 | Studies that record human judgements

Studies that record *human judgements* involve an active participation of human subjects in the experimental phase. Usually, participants are played some musical stimuli and are asked to provide data concerning tension or expectation. There are different ways in which participants can be asked about the given stimuli.

The most common approaches include asking participants to do one of the following:

(a) sing a completion of the given stimuli,

(b) select, from a collection of supplied possibilities, the tension profile that best suits the given stimuli,

(c) rank, using a quantitative scale, the extent to which a collection of supplied possibilities fits the context established by the given stimuli, and

(d) draw the curve described by tension in the given stimuli.

Approach (a) is mostly used in *empirical studies* that focus on melodic expectation. Typically, in the studies that follow this approach, participants first listen to a musical stimulus and then are asked to sing one note, as in Povel (1996), or several notes, as in Carlsen (1981), to continue the stimulus. Although some studies allow participants to sing several notes, their analysis often only focuses on the first note.\(^\text{10}\)

Approach (b) is mostly used in *empirical studies* that investigate the shapes described by perceived musical emotions. Typically, in the studies that follow this approach, participants first listen to a musical stimulus and then are asked to select, from a collection of curves, the one they consider best matches the tension conveyed by the given stimulus. Farbood (2012) composed a collection of stimuli, each of which isolated a particular feature,\(^\text{11}\) and asked participants to select, from seven different curves representing tension, the curve they considered best matched each stimulus.

\(^{10}\)Larson (1997) defines this approach as a “melodic continuation”, which he distinguishes from “melodic completions”, where participants’ entire responses are analysed. He believes that the latter approach captures participants’ melodic expectations in a better way, compared to the former approach.

\(^{11}\)The features examined in her study include onset frequency, tempo, loudness, pitch height, harmonic motion, rhythmic regularity and meter. Notice that these consist of both cognitive and psychoacoustic features, so Farbood’s study will fall in the category labelled as both in Figure 2.1.
Approach (c) is mostly used in empirical studies that investigate stability in the tonal context. This approach is used extensively in the field of music cognition, probably because of the impact of the “probe-tone method”. Krumhansl and Kessler (1982) popularised this method in their investigation of the cognitive representation of harmonic and tonal structure in Western music. They asked a group of participants to rate how well the twelve tones of the chromatic scale followed a scale, a chord and a cadence.

Approach (c) is also used in empirical studies that focus on melodic expectation, such as Krumhansl (1995); Schellenberg (1996); Schmuckler (1989). Pearce (2003) refers to these as “perceptual paradigms”, in the context of melodic expectation.

Approach (d) is mostly used in empirical studies that investigate the shapes described by perceived musical tension. There are two ways in which approach (d) is usually applied: the stop-tension task and the continuous-tension task. In the stop-tension task, the first event in a sequence (e.g. a note or chord) is played and listeners record its degree of tension; then, the first and second events are played and listeners record the degree of tension of the second event; then, the first, second and third events are played and listeners record the degree of tension of the third event, and so on. In the continuous-tension task, listeners are asked to record the degree of tension perceived in real time while the music is being played. For more detail, see Lerdahl and Krumhansl (2007).

2.3.2 | Studies that record physiological responses

Studies that record physiological responses involve a passive participation of human subjects in the experimental phase. Usually, participants are played some musical stimuli and are asked to relax and listen to the music while their physiological responses are being recorded.

---

12The profiles Krumhansl and Kessler obtained were stable. They believe these profiles reflect the tonal hierarchies in major and minor keys and so produced a map of the distances between keys. See Krumhansl (2001) for more detail on the “probe-tone method” and Milne (2013) for more detail on its impact on other existing maps of tonality.

13Much research concerning approach (d) has been carried out using the Continuous Response Digital Interface (CDRI), a device specifically designed for this purpose. See Geringer et al. (2004) for more detail.
The most common physiological\textsuperscript{14} measures include: cardiac changes, respiratory changes, electrodermal changes and vascular changes.

Meyer’s (1956) approaches to meaning had a great impact on and boosted the exploration of musical emotions using physiological responses. That is because differences have been observed between the emotions listeners may “perceive” while listening to music, meaning the emotions they assume music should be expressing, and the emotions they “feel” while listening to music, meaning their actual emotional experiences (Gabrielsson, 2001; Lehne and Koelsch, 2015).

Krumhansl (1997) is a good example to illustrate the points made in this section. She distinguishes between two positions concerning musical emotions, depending on whether music elicits emotion in music or just expresses the emotions recognised by listeners in the music, respectively (for more detail, see footnote 6 on page 21). To support the former, as opposed to much of the research introduced in Section 2.3.1, she recorded physiological changes\textsuperscript{15} listeners experienced while listening to music.

Krumhansl played six excerpts of music which represented sadness, fear and happiness. Two slow excerpts in minor mode, T. Albinoni’s Adagio in G minor for Strings and Orchestra and S. Barber’s Adagio for Strings, Op. 11, were chosen to represent sadness. Two fast and aggressive excerpts, G. Holst’s Mars — the Bringer of War from The Planets

\textsuperscript{14}Some authors, such as Van der Zwaag et al. (2011) or Krumhansl (1997), use the term psychophysiology, instead of just physiology, when investigating the above responses. However, according to Cacioppo et al. (2007, p.5):

Psychophysiology is intimately related to anatomy and physiology but is also concerned with psychological phenomena – the experience and behavior of organisms in the physical and social environment. The complexity added when moving from physiology to psychophysiology includes both the capacity by symbolic systems of representation (e.g., language and mathematics) to communicate and to reflect upon history and experience as well as the social and cultural influences on physiological response and behavior.

Following this line of thought, in this dissertation, the term physiology is preferred over psychophysiology when only referring to the four physiological responses presented above. If it were the case that the recorded data included neural activity, the term psychophysiology may be preferred. Note, however, that neural responses are reviewed in an independent section.

\textsuperscript{15}The measures included cardiac interbeat interval (i.e. cardiovascular arousal), pulse transmission time to the finger, finger pulse amplitude (i.e. amount of blood in the finger’s periphery), pulse transmission time to the ear, respiration intercycle interval (i.e. time between successive inspirations), respiration depth (i.e. the point of maximum inspiration minus the point of maximum expiration), respiration-sinus asynchrony, systolic blood pressure, distolic blood pressure, mean arterial pressure, skin conductance level and temperature on the finger.
and M.Mussorgsky’s *Night on Bare Mountain*, were chosen to represent fear. Two playful and bouncy excepts in major mode, A. Vivaldi’s *Spring* from *The Four Seasons* and H. Alfvén’s *Midsommarvaka*, were chosen to represent happiness. She then compared the listeners’ physiological responses to ratings of the degree of sadness, fear, happiness and tension provided by another group of listeners. Tension correlated most strongly with fear, as well as to some extent with sadness and happiness, especially when these were the main emotion of an excerpt. Because of this, Krumhansl hypothesised that tension would correlate with a broad spectrum of the physiological measures. She confirmed this by computing a factor analysis of the recorded physiological measures. All factors but one, consisting of cardiac interbeat interval, correlated significantly with tension.

### 2.3.3 | Studies that record neural activity

Studies that record *neural activity* involve a passive participation of human subjects in the experimental phase. Usually, participants are played some musical stimuli and are asked to relax and listen to the music while their brain activity is being recorded. In the last decade, the number of these studies has rapidly increased, particularly fostered by the joint efforts of a collection of research centres in Berlin. See Blankertz et al. (2016) for more detail.

There exist two main methods to record participants’ brain activity: fMRI (*functional Magnetic Resonance Imaging*), which records brain activity by detecting changes associated with blood flow, and EEG (*Electroencephalography*), which records the brain’s spontaneous electrical activity by detecting voltage fluctuations within the neurons of the brain.

Sturm et al. (2015) is a good example to illustrate the scope of EEG-based studies that record neural activity with regards to musical tension. In their study, six pieces of music, a sequence of major chords and two excerpts of environmental and instrumental noise were played to a group of participants. Sturm et al. recorded the participants’ brain activity and analysed it against a collection of audio-related features (i.e. related to audio signal properties rather than higher-level musical concepts) extracted from the
musical stimuli, as well as against continuous tension judgements provided by another group of participants.

Sturm et al. found that sharpness, calculated as the mean positive first derivative of the stimulus’s waveform power, was the feature that most strongly correlated with tension ratings.

Lehne et al. (2014) is a good example to illustrate the scope of fMRI-based studies that record neural activity with regards to musical tension. In their study, four pieces of music were played to a group of participants and their activity in the orbitofrontal cortex and the amygdala was recorded. Lehne et al. analysed this activity against continuous tension judgements provided by the same group of participants. The tension judgements were collected twice, before and after the fMRI scanning. Participants were asked to focus “on their subjective experience of tension, and not on the tension they thought that the music was supposed to express” (p.1516). Note the discussion about the source of musical emotion described in footnote 6 on page 21.

Lehne et al. found that musical tension positively correlated with blood oxygen level-dependent signal changes in the left pars orbitals of the inferior frontal gyrus (i.e. the lowest positioned ridge on the prefrontal cortex).

2.3.4 | A critical look at empirical studies of musical tension

Empirical studies of musical tension present some advantages and some disadvantages in the context of this dissertation. On one hand, most of them meet the requirements identified in Sections 2.1.4 and 2.2.5, considering how empirical studies of musical tension were defined at the beginning of Section 2.3. On the other hand, the scope of empirical studies of musical tension is usually narrower than that of other approaches. This issue is mostly due to the difficulties that a broader scope would add to the analysis of the results of an empirical study. That is to say, empirical studies of musical tension usually study the relationship between tension and one or a few number of features. These features normally describe some harmonic, melodic and rhythmic qualities of the studies’ stimuli. The more of these features that are included in the analysis of an empirical stud-
ies, the more difficult it is to conclude the extent to which these features may contribute to modelling musical tension. Hence the scope of empirical studies of musical tension tends to be narrow.

The fact that the scope of most empirical studies of musical tension is usually narrow entails a disadvantage concerning our goals. By only investigating one or a few number of features, it is likely that some of the feelings included in our definition of tension would not be taken into consideration. This may happen because of either one or both of the following reasons: the study only considers features concerning either harmony, melody or rhythm; and/or the features only concern pre-outcome or post-outcome responses.

Recall that our definition of tension conceives musical tension as a direct consequence of musical expectations. Concerning these expectations, many actions are considered in our definition of tension, such as their emergence, fulfilment or violation, among others. But, most importantly, recall that these expectations may refer to the three dimensions we have chosen to concentrate on, harmony, melody and rhythm, and that they refer to both pre- and post-outcome responses. Thus, two more requirements (i.e. (iv) and (v)), which a model should meet to suit our goals, can be derived from this section’s discussion:

(i) the model must be consistent with our definition of tension,

(ii) the model must provide a method to calculate quantitative values of tension,

(iii) the model must have been empirically tested and have shown strong correlations with the degrees of musical tension perceived by human listeners,

(iv) the model components must concern harmony, melody and rhythm, and

(v) the model must consider both pre- and post-outcome responses.

To wrap this section up, there is a final note to be made. From the three approaches concerning empirical studies of musical tension reviewed in Sections 2.3.1, 2.3.2, 2.3.3, the former, studies that record human judgements, is the most used in the literature. This may be because of several reasons. For instance, in terms of apparatus, it may be the
cheapest and easiest to set up. As a consequence, this approach has been developed and improved over several decades. Thus, this dissertation will prefer a model of tension that can be empirically tested using human judgements instead of physiological and neural data. This observation should not be interpreted as a requirement, but rather as a preference.

2.4 Generation models of musical tension

Generation models of musical tension consist of collections of methods capable of predicting quantitative\textsuperscript{16} degrees of tension in the musical discourse. They usually stand upon the findings of other theoretical accounts and empirical studies of musical tension.

There exist many generation models whose main focus is melodic expectation. These include Larson’s (1993; 2004) model of musical forces, Margulis’ (2005) model of melodic expectation and Pearce’s (2005; 2018) information dynamics of music (IDyOM) model, among others. Although these match the above description of generation models of musical tension, they do not fully cover the harmonic and rhythmic elements that may have an effect on the tension perceived by human listeners. That is to say, they do meet requirement (iv), previously proposed in Section 2.3.4 (i.e. the model components must concern harmony, melody and rhythm), and so they will not be further discussed here.

In the literature, there exist three generation models of musical tension that meet all the requirements proposed so far. These are Lerdahl’s (2004) Model of Tonal Tension (MTT), Farbood’s (2012) Parametric-Temporal Model of Musical Tension and Herremans and Chew’s (2016) Tension Ribbons. These models are reviewed below to illustrate the scope of generation models of musical tension.

\textsuperscript{16}Other models that calculate these predictions qualitatively might also be labelled as generation models of musical tension. However, this will mean that pieces of research such as GTTM (Lerdahl and Jackendoff, 1983) will fall within generation models instead of falling within theoretical accounts of musical tension. In this dissertation, because of the requirements discussed so far, GTTM fits better the description of theoretical accounts of musical tension and so generation models are here required to provide a quantitative output.
2.4.1 | Lerdahl’s Model of Tonal Tension (MTT)

As introduced in Section 2.2.4, GTTM (Lerdahl and Jackendoff, 1983) consists of a set of rules to generate hierarchical reductions of a given piece of music. To quantify GTTM’s patterns of tension and relaxation, Lerdahl (2004) developed his Model of Tonal Tension (MTT). To calculate the quantitative degrees of musical tension in a given piece of music using MTT, one must follow three steps:

1. calculating the cognitive distance between the events connected by a GTTM hierarchical representation,

2. calculating, for each event in the piece, the degree of tension generated by surface dissonances, and

3. calculating the degrees of attraction of all voice-leading paths between consecutive chords.

MTT will not be further discussed here as it is reviewed in more detail in Chapter 3.

2.4.2 | Farhood’s Parametric-Temporal Model of Musical Tension

Farhood’s (2012) Parametric-Temporal Model of Musical Tension includes three major components:

- an attentional window, which extracts the current tension trend at a “perceptual moving window in time” (p.415),

- a memory window, which calculates the direction of the tension trend extracted in the immediately preceding attentional window, and

- a collection of weightings, which rate the influence of the musical features included in her model.
In Farbood’s model, tension is defined by the changes of direction of a collection of musical features. To calculate a tension trend, an *attentional window* calculates the slopes of the features over a period of time. For each feature, its slope is defined as that with the best linear fit. The slope of a tension trend is then determined by summing all weighted slopes of each feature over the discrete time defined by the *attentional window*. To model the effect the *memory window* may have on the *attentional window*, the tension trend calculated in an *attentional window* is multiplied by a “memory constant”, $\beta$. If the direction of the slope in the *memory window* is different from the slope in the *attentional window*, $\beta = 1$; otherwise, $\beta$ is some (empirically determined) positive value.

Farbood (2012) used her model to generate predictions of tension of a collection of stimuli. She tried *attentional windows* ranging from 1 to 8 seconds and *memory windows* ranging from 0 to 8 seconds. By comparing these predictions against participants’ judgements of the stimulus’s tension, she concluded that, for both the *attentional* and the *memory window*, the optimal size is 3 seconds, and that the optimal value of $\beta$ is 5. These optimal values correspond to those that best fit the participants’ judgements.

### 2.4.3 | Herremans and Chew’s Tension Ribbons

Herremans and Chew’s (2016) *Tension Ribbons* consist of three methods to quantify and visualise musical tension associated with melodic and harmonic motion. The methods are based on the *spiral array*, a geometric model of tonality developed by Chew (2014).

The *spiral array* represents pitch classes as spatial coordinates within a three-dimensional helix. To locate pitch classes in the helix, they are indexed by their number of perfect fifths from an arbitrarily chosen reference, $c$ at position $[0, 1, 0]$. An increment in the index corresponds to a quarter turn along the spiral. Therefore, notes that are

---

17 These features include: tempo, in terms of beats per minute; onset frequency; pitch height of the melody and the bass, in terms of the position in the scale of the notes in the melody and bass respectively; loudness, measured according to Jehan’s (2005) psychoacoustic loudness model; melodic expectation, measured according to Margulis’ (2005) model of melodic expectation; and harmonic tension, measured according to Lerdahl’s (2004) MTT.

18 The weighting assigned to each feature is estimated from the results of the first study in Farbood (2012): dynamics and pitch height are assigned the highest weight ($weight = 3$) followed by tempo and onset frequency ($weight = 2$), and harmonic tension and pitch height are assigned the lowest weight ($weight = 1$).
positioned one perfect fifth away from each other are separated by quarter turns in the spiral. Four quarter turns result in notes positioned above each other, which represent major thirds. It is also possible to represent chords and keys within the spiral array. Chords are represented by the combinations of their respective pitches in the array. In this way, triads will be represented by non-overlapping triangles. Keys are represented by the combinations of the triangles of their respective tonic, subdominant and dominant chords. See Chew (2002; 2014) for more detail.

To apply the Tension Ribbons to calculate musical tension, a given piece of music must be divided into equal-length windows. For each of these windows, its notes can be represented as a cloud in the spiral array.

The three methods of Tension Ribbons aim to capture different contributions of musical tension with regards to these clouds. These methods are: the cloud diameter, which aims to capture the perceptual distance between the notes in a cloud;\(^{19}\) the cloud momentum, which aims to capture the movement in tonality between clouds;\(^{20}\) and the tensile strain, which aims to capture the perceptual distance between a local and a global key.\(^ {21}\)

To test the capabilities of the Tension Ribbons, Herremans and Chew (2016) calculated the values of the cloud diameter, the cloud momentum and the tensile strain in the pieces of music used in Farbood’s (2012) empirical study of musical tension. They compared the curves described by the three methods against the judgements provided by the participants in Farbood’s study and observed that the three methods do capture, to some extent, some nuances of musical tension.

Despite the promising results of the Tension Ribbons, Herremans and Chew (2016, p.10) conclude that:

> a more thorough empirical study of how the quantitative measures produced

---

19. To do so, the cloud diameter calculates the largest Euclidean distance between any two notes in a cloud.
20. To do so, each cloud is assigned a “centre of effect”. To calculate the value of a cloud’s “centre of effect”, each pitch is assigned a contribution corresponding to the product of its duration by its position in the spiral array divided by the total duration of the pitches in the cloud. The cloud’s “centre of effect” is then calculated as the sum of the contributions assigned to all the pitches in the cloud. To represent the movement in tonality, the cloud momentum is calculated as the Euclidean distance between the “centres of effect” of two clouds. In this way, a change in tonality will result in a larger value of cloud momentum.
21. To do so, the tensile strain calculates the distance between the “centre of effect” of a cloud and the piece’s global key. The global key of a piece is estimated using Chew’s (2014) key detection algorithm.
Chapter 2. Modelling Musical Tension

Generation models of musical tension

by [the Tension Ribbons] correlate with what listeners describe as tension
[is needed] (...) [and] further extensions could take into account features
related to melodic contour, rhythm and timbre.

2.4.4 | A critical look at generation models of musical tension

Generation models of musical tension present some advantages in the context of this
dissertation. Notice that the three models reviewed in Sections 2.4.1, 2.4.2 and 2.4.3
(Lerdahl’s, Farbood’s, and Herremans and Chew’s, respectively), meet the requirements
identified in Sections 2.1.4, 2.2.5 and 2.3.4. Thereby, using a generation model of musical
tension seems to be the best approach to answer our Research Question. But, will any
generation model of musical tension suit our goals?

Apart from meeting the above requirements, for a generation model to suit our goals,
it must also meet two additional requirements (i.e. (vi) and (vii)):

(i) the model must be consistent with our definition of tension,

(ii) the model must provide a method to calculate quantitative values of tension,

(iii) the model must have been empirically tested and have shown strong correlations
with the degrees of musical tension perceived by human listeners,

(iv) the model components must concern harmony, melody and rhythm,

(v) the model must consider both pre- and post-outcome responses,

(vi) the model must take into account a piece’s structure, and

(vii) the model must be applicable in real time.

Lerdahl’s (2004) MTT accounts for a piece’s structure as its calculations of tension
depend on the piece’s hierarchical structure provided by the corresponding GTTM anal-
ysis. Taking into account that Farbood’s (2012) Parametric-Temporal Model of Musical
Tension includes Lerdahl’s hierarchical contribution as its harmonic feature, it can be
concluded that it also accounts for a piece’s structure. On the other hand, Herremans
and Chew's (2016) Tension Ribbons do not explicitly account for a piece's structure. It is true that they have used their model to generate music that has long-term structure and that matches input tension profiles using the system MorpheuS (Herremans and Chew, 2017). However, other aspects of MorpheuS deal with a piece's structure and not the Tension Ribbons model.

Concerning the application of the models in real time, Herremans and Chew's model offers the possibility of defining clouds of notes of short duration. For instance, in the case of their system MorpheuS, which uses the Tension Ribbons, they claim that some modifications will allow its use in real time (Herremans and Chew, 2017, p.12). On the other hand, Farbood's model will not be applicable in real time. Notice that the optimal window sizes in her model were of 3 seconds, which is a too long time-span that could prevent our system from reacting to tension changes in real time. Finally, Lerdahl's model does meet this requirement, since both TPS's and MTT's rules “are essentially algebraic operations” (F. Lerdahl, personal communication, September, 2019). Therefore, the timing of the application of the model's rules would only depend on our system's running time.

2.5 | Which models of tension are best suited to our goals?

Sections 2.2, 2.3 and 2.4 have proposed a collection of requirements that a model of tension should meet to suit our goals. These can be summarised as follows.

For a model of tension to suit our goals it is required that:

(i) it is consistent with our definition of tension,

(ii) it provides a method to calculate quantitative values of tension,

(iii) it has been empirically tested and has shown strong correlations with the degrees of musical tension perceived by human listeners,

---

22MorpheuS identifies musical patterns using Meredith's (2013) pattern-recognition algorithms.
(iv) it consists of rhythmic, melodic and harmonic components,

(v) it considers both pre- and post-outcome responses,

(vi) it takes into account a piece’s structure, and

(vii) it is applicable in real time.

From Section 2.4.4, it can be concluded that Lerdahl’s (2004) Model of Tonal Tension (MTT) may be the best approach to answer our Research Question.

Concerning requirement (i), MTT is presented as a theory that conceives tonal tension as both a pre- and post-outcome response and it is based on rhythmic, melodic and harmonic components. That is to say, MTT is consistent with our definition of tension and so it meets the first requirement. Concerning requirement (ii), MTT’s components “combine to [quantitatively] predict the rise and fall in tension in the course of listening to a tonal passage or piece” (Lerdahl and Krumhansl, 2007, p.319), so MTT meets this requirement. Concerning requirement (iii), MTT has been empirically tested in Lerdahl and Krumhansl (2007) and has shown strong correlations against the degrees of musical tension perceived by human listeners, so MTT meets this requirement. Concerning requirement (iv), MTT’s components consist of a “[a] prolongational structure, [b] a pitch-space model, [c] a surface-tension model, and [d] an attraction model” (Lerdahl and Krumhansl, 2007, p.319), which concern rhythm (a), melody (c and d) and harmony (a, b, c and d), so MTT meets this requirement. Concerning requirement (v), MTT’s attraction and pitch-space components concern pre- and post-outcome responses, respectively, so MTT meets this requirement. Concerning requirement (vi), MTT’s prolongational structure component takes into account a piece’s structure, based on GTTM’s outputs, so MTT meets this requirement. As discussed in Section 2.4.4, MTT is applicable in real time, so MTT meets requirement (vii).

To wrap this section up, there is a final note to be made. It should be noted that our definition of tension also considers psychoacoustic features. However, as pointed out by Granot and Eitan (2011), while many empirical studies of musical tension suggest that psychoacoustic features such as loudness and tempo may have a strong effect in generat-
ing perceived musical tension, some studies report a very weak effect of both parameters. Examples of the latter include Krumhansl (1996) or Burnsed and Sochinski (2001). Because of this reason, this dissertation has not included in the above requirements the need of a model of tension to consider psychoacoustic features. They, however, will be considered in Chapter 6 in line with Farbood’s (2012) findings.

2.6 | Summary and Conclusions

In order to seek answers to our Research Question, Chapter 2 has identified the model of tension best suited to our goals.

Section 2.1 has framed the scope of the concept of musical tension. A key issue was observed concerning the different interpretations of the concept in the literature: some scholars only use the term “tension” to refer to pre-outcome responses, whereas some others only use it to refer to post-outcome responses. Nevertheless, it was concluded that, in the end, both views describe similar feelings when they are interpreted with regards to the stress or energy perceived by human listeners when listening to music. In order to account for this type of descriptions, Section 2.1 generalises the concept of tension by defining it as the reaction listeners experience as a consequence of musical expectations and of changes in the perceived levels of arousal and activity.

Section 2.2 has defined the concept of theoretical accounts of musical tension and has reviewed four main theories to illustrate the scope of the approach. It was concluded that these theories, on their own, may not suit our goals because most of them are entirely qualitative and have not been fully empirically tested.

Section 2.3 has defined the concept of empirical studies of musical tension and has reviewed some of them with regards to the three most common experimental methodologies to illustrate the scope of the approach. It was concluded that these studies may not suit our goals because most of them do not concern the three dimensions we have chosen to concentrate on, harmony, melody and rhythm, and do not consider both pre- and post-outcome responses.

Section 2.4 has defined the concept of generation models of musical tension and has
reviewed three main models to illustrate the scope of the approach. It was concluded that these theories may suit our goals if they concern long-term structure and are applicable in real time.

Finally, Section 2.5 has identified Lerdahl’s (2004) Model of Tonal Tension (MTT) as the model best suited to our goals. It has also concluded that the findings in Farbood (2012) will also be of interest when designing, implementing and evaluating the music generator later in Chapters 6 and 7.
Chapter 3

Lerdahl’s Model of Tonal Tension

*Background, overview and application*

In this chapter, we review Lerdahl’s Model of Tonal Tension (MTT), which was identified in Chapter 2 as the model best suited to our goals. To do so, we will:

1. contextualise Lerdahl’s MTT,
2. review the components in Lerdahl’s MTT,
3. provide an illustrative example of the calculation of musical tension according to Lerdahl’s MTT, and
4. review the evaluation of Lerdahl’s MTT.

The first step is addressed in Section 3.1. The second and third steps are addressed in Sections 3.2, 3.3 and 3.4, each of which reviews one of the three components MTT consists of and provides an example of their application. The fourth step is addressed in Section 3.5.
3.1 | Background

In the early 1970s, inspired by the reformulations proposed by Noam Chomsky in the field of linguistics, Fred Lerdahl and Ray Jackendoff decided to study music in a similar fashion, and so decided to formulate a theory that focused on the following goals:

- generating the structure of a piece of music from the point of view of its musical surface and not from the idea of an ideal fundamental structure,
- not focusing on a particular musical style,
- considering both pitch and rhythm with the same level of importance,
- deriving rules, motivated by psychological principles, that represent cognitive principles of organisation,
- providing structural descriptions that would correspond to those that may be inferred by human listeners, and
- making it possible for the theory to be tested.

At first, Lerdahl and Jackendoff (1977) developed some rules to assign hierarchical structure to pitch events. To do so, they developed a collection of conditions to decide the relative importance of the events in a piece of music, as well as a collection of rules to assign grouping and metrical structures that could be combined to form a time-span segmentation. They used this segmentation to generate a first hierarchy where each event was assigned a location in the time structure.

Lerdahl and Jackendoff concluded that their time-span component was useful but insufficient to generate an accurate analysis of a piece’s structure, so they developed another pitch-related hierarchy to represent qualitative patterns of tonal tension and relaxation.

The combination of the above mentioned rule-based structures is what shapes Lerdahl and Jackendoff’s (1983) Generative Theory of Tonal Music (GTTM). For more detail on
the genesis and architecture of GTTM see Lerdahl (2009). For an overview of the main arguments supporting GTTM’s principles see Lerdahl (2015).

Five years after the publication of GTTM, Lerdahl (1988) developed his Tonal Pitch Space (TPS) in order to quantify “the most important of the stability conditions through computational modeling of empirical data on the tonal hierarchy” (Lerdahl, 2009, p.191). TPS was inspired by early theories that aimed to capture the distance among pitch configurations that may be inferred by listeners. However, Lerdahl observed that most existing theories, at the time, only addressed the relative distance, between musical events, from the point of view of either pitch classes, chords or tonal regions; but they never consider them all. Thus, he concluded that a new model that incorporated all three perspectives was needed.

Inspired by the work of Deutsch and Feroe (1981), who claim that pitch “alphabets” may account for the way listeners perceive musical structures, Lerdahl (1988) developed a “basic space” to investigate the proximity among pitch classes. He then developed a collection of rules to evaluate transitions from the “basic space” of one chord to the “basic space” of another. He did this to be able to investigate the proximity among both chords and their respective tonal regions.

The combination of the above mentioned collection of transition rules is what shapes TPS. TPS was first introduced in Lerdahl (1988); it was later improved in Lerdahl (2004) and its final version was presented in Lerdahl and Krumhansl (2007). TPS has shown connections with many other theories and empirical studies. For more detail, see Lerdahl (1988, pp.337-346) and Lerdahl (2004, ch.2).

Eight years after the publication of (the first version of) TPS, Lerdahl (1996) developed his Model of Tonal Tension (MTT) motivated by “[GTTM’s] vagueness of the stability conditions and lack of a quantitative account of tension” (Lerdahl, 1996, p.322). His first step towards overcoming this issue was to translate the notion of stability into that of tension by assuming that “the more unstable two events are with respect to one another, the greater the tension between them” (Lerdahl, 1996, p.325). From this new interpretation, he developed a harmonic and a melodic model of tension. The former calculated tension both sequentially, as distances between consecutive chords within TPS,
and hierarchically, as TPS distances between chords connected by GTTM’s hierarchical structures. The latter evaluated tension at the level of the melody inspired by Bharucha’s (1984) “anchoring principle”.¹ Because some preliminary analysis did not match the predictions of the harmonic and melodic models, Lerdahl developed an additional method to evaluate the perception of tension caused by “surface dissonance”.

The combination of the above mentioned models of tension is what shapes Lerdahl’s MTT. MTT was first introduced in Lerdahl (1996); it was later improved in Lerdahl (2004) and its final version was presented in Lerdahl and Krumhansl (2007). MTT has shown connections with other theories. For more detail, see Lerdahl (2004, pp.188-192).

Lerdahl's GTTM, TPS and MTT are reviewed in Sections 3.2, 3.3 and 3.4, respectively.

3.2 | A Generative Theory of Tonal Music (GTTM)

The best description of Lerdahl and Jackendoff’s (1983) Generative Theory of Tonal Music (GTTM) may be that of Lerdahl (2004, p.3):

A listener familiar with a musical idiom organizes its sounds into coherent structures. GTTM attempts to characterize those musical structures that are hierarchical and to establish principles by which the listener arrives at them for a work in the Classical tonal idiom. These principles are stated as a musical grammar, or system of rules, that generates the structure that the listener associates with the signal.

He also adds:

GTTM proposes four types of hierarchical structure simultaneously associated with a musical surface. Grouping structure describes the listener’s segmentation of the music into units such as motives, phrases, and sections. Metrical structure assigns a hierarchy of strong and weak beats. Time span reduction, the primary link between rhythm and pitch, establishes the relative structural

¹According to the “anchoring principle”, there is a psychological need for an unstable pitch to be resolved to an immediately subsequent stable pitch. For more detail, see Bharucha (1984); Lerdahl (2004).
importance of events within the rhythmic units of a piece. Prolongational reduction develops a second hierarchy of events in terms of perceived patterns of tension and relaxation.

What does GTTM mean by reducing a piece of music hierarchically? We can grasp an intuitive understanding of the concepts of reduction and hierarchy by means of a simple example. Let us consider Figure 3.1a, which shows the melody of the first phrase of the theme in Mozart’s *Ah vous dirai-je, Maman*, K. 265/300e, popularly known as *Twinkle, twinkle, little star*. In order to hierarchically reduce this phrase, we need to identify the important events in it. For instance, the last note in the penultimate measure in Figure 3.1a, note e, can be considered an embellishment. That is to say, it can be seen as being less important than the notes around it. This note can then be reduced. This is shown in Figure 3.1b, where note e has been absorbed by its preceding note. Notice that, by doing this, all the measures, but the last, now have the same structure consisting of two repeated quarter-notes.

The musical content in Figure 3.1b can still be reduced. For instance, since all measures but the last consist of a repeated note, they can be simplified. This is shown in Figure 3.1c, where the second note in each measure has been absorbed by its respective first note.

A fine-grained reduction can still be done upon Figure 3.1c. Notice that in measures two to four, note a departs from and arrives at note g. That is to say, note a can be considered an embellishment. Likewise, in measures five to eight, the sequence f-e-d could be seen as pushing the music forward towards the last note in the phrase, note c, as if they were resolving into this note. That is to say, the notes in this sequence can be seen as also being some kind of embellishment and so can be reduced. This is shown in Figure 3.1d, which, therefore, only includes the most important notes in Mozart’s *Twinkle, twinkle* theme. The theme’s phrase can then be understood as a journey that starts at note c, passes through note g and ends at note c. By looking at Figure 3.1, from bottom to top, this journey is embellished with new notes.

Finally, the idea of the c-g-c journey can be represented in the form of a basic skele-
ton upon which the theme stands. This is shown in Figure 3.1e. Notice this figure resembles Schenker’s (1979) analytical theory, where it is proposed that tonal music follows the same fundamental structure (i.e. the Ursatz, I-V-I) from which the musical structure of a piece of music can be derived.

![Figure 3.1](image)

**Figure 3.1**: An intuitive reduction of the first phrase of the theme in Mozart’s *Ah vous dirai-je, Maman*, K. 265/300e, with regards to the level of importance of its notes.

The example in Figure 3.1 is a simple way to introduce the concept of reduction as a simplified structure into which we, the listeners, organise the events in a piece of music with regards to their importance. If we compare the notes that are absorbed and those that are kept between consecutive levels in Figure 3.1, we can also sense some kind of hierarchical structure. That is to say, we can sense some system of organisation whose elements are related in such a way that more important elements, the dominating ones, subsume less important elements, the subordinate ones.

GTMM attempts to define a method to characterise musical structures within a hierarchy, similar to the process discussed above with regards to Figure 3.1. To do so, it establishes a collection of principles by which a listener would infer such structures. These principles are given as systems of rules organised along four components: the *grouping structure*, the *metrical structure*, the *time-span reduction* and the *prolongational reduction*. These are reviewed in Sections 3.2.1, 3.2.2, 3.2.3 and 3.2.4, respectively.
3.2.1 | The grouping structure

In order to hierarchically reduce a piece of music, the first step GTTM takes is to organise the piece into smaller groups, such as motifs, themes, subjects, phrases, etc. To do so, it incorporates the *grouping structure*.

GTTM defines a group as a unit into which listeners naturally organise sound signals. In order to segment a piece of music into groups, GTTM’s *grouping structure* incorporates two sets of rules: grouping well-formedness rules (GWFRs) and grouping preference rules (GPRs). The former establish formal principles that patterns of musical events must show to be considered groups. The latter establish formal principles to select, from the possible patterns of groups that can be assigned to the musical discourse, those patterns that would most likely correspond to the listeners’ intuitions.

GWFRs and GPRs are introduced below, although the original form of some of the rules has been summarised. For a full description of the rules, and how they are derived, see Lerdahl and Jackendoff (1983, ch.3).

---

**GTTM’s grouping well-formedness rules (GWFRs)** consist of:

- **GWFR 1** Any contiguous sequence of events can constitute a group.
- **GWFR 2** A piece constitutes a group.
- **GWFR 3** A group may contain smaller groups.
- **GWFR 4** If a group contains part of another group, the former must contain all of the latter.
- **GWFR 5** If a group contains a smaller group, the former must be partitioned into smaller groups.

**GTTM’s grouping preference rules (GPRs)** consist of:

- **GPR 1** Prefer a grouping structure without small groups.
- **GPR 2** If there exist great differences between the intervals of time among
consecutive notes, a boundary between groups may be placed.

**GPR 3** If there exist great changes in intervallic distance, dynamics, articulation or duration among consecutive notes, a boundary between groups may be placed.

**GPR 4** Where boundaries identified by GPRs 2 and 3 are relatively more pronounced, a boundary between larger-level groups may be placed.

**GPR 5** Prefer a symmetrical grouping structure where groups are subdivided into two parts of equal length.

**GPR 6** Prefer a parallel grouping structure if groups show similarity of rhythm, internal grouping or pitch contour.

**GPR 7** Prefer a grouping structure which results in a stable organisation of the time-spans.

By parallel structures, GTTM refers to analyses which are similar or even identical. For instance, in Figure 3.1, we reduced all measures in Figure 3.1b into those in Figure 3.1c in the same way (i.e. all two-quarter-note measures were reduced to one-half-note measures). We refer to the analyses of these measures as being parallel.

Figure 3.2 shows the *grouping structure* of the first phrase of the theme in Mozart’s *Ah vous dirai-je, Maman*, K. 265/300e. The groups are represented through a collection of slurs.

![Figure 3.2: GTTM's grouping structure of the first phrase of the theme in Mozart's Ah vous dirai-je, Maman, K. 265/300e.](image)

In Figure 3.2, *level a* is generated according to GWFR 2; that is to say, the whole
theme is considered a group. Taking into account that the notes in the theme’s melody repeat once in each measure, level c is generated according to GPR 3; that is to say, each measure is considered a group because there is always a jump (i.e. a change in intervallic distance) from the last note in a measure to the first note in the next and so a boundary is placed between measures. Likewise, in the accompaniment, notes follow a descending pattern, except for the first and last two measures. According to GPR 6, parallel groups are preferred, so that is another reason why level c is generated as a sequence of identical groups. Finally, level b is generated according to GPR 5; that is to say, level a is split into two equal-length groups. Note that these senses of symmetry and parallelism match the intuitive subdivision of the theme into two semi-phrases, according to the theme’s contour.

Levels beyond level c would mean generating one-note groups. These are not considered because of GPR 1. That is also the reason why the last two measures are included in the same group at level level c, so that the last measure does not constitute a one-note group on its own.

Notice that the grouping structure in Figure 3.2 matches all GWFRs.

3.2.2 | The metrical structure

In order to hierarchically reduce a piece of music, the second step GTTM takes is to identify the most probable patterns of strong and weak beats. To do so, it incorporates the metrical structure.

GTTM defines a beat as a basic element, without duration, that makes up metrical patterns and is inferred by the listeners from the musical signal. Some of these metrical patterns may be more important than others and so it is possible to organise them into a metrical hierarchy. At a given level of the metrical hierarchy, if a beat is included in a larger level, it is a strong beat. Otherwise, it is a weak beat. Recall that we used a similar analogy in the example in Figure 3.1 to differentiate between more and less important notes in Mozart’s Twinkle, twinkle theme, at different levels of the reduction.

The metrical hierarchy can be represented through a grid consisting of different lev-
els. The levels are normally defined by the length of the time-spans, which represent intervals of time that take place between successive beats.

In order to estimate the metrical hierarchy in a piece of music, GTTM’s *metrical structure* incorporates two sets of rules: metrical well-formedness rules (MWFRs) and metrical preference rules (MPRs). The former establish formal principles that groups of beats must show to consider they form metrical patterns. The latter establish formal principles to select, from the possible metrical patterns that can be assigned to the musical discourse, those patterns that would most likely correspond to the listeners’ intuitions.

MWFRs and MPRs are introduced below, although the original form of some of the rules has been summarised. For a full description of the rules, and how they are derived, see Lerdahl and Jackendoff (1983, ch.4).

**GTTM’s metrical well-formedness rules (MWFRs)** consist of:

- **MWFR 1** Every attack point must be associated with a beat at the smallest metrical level of the metrical structure.
- **MWFR 2** Every beat at a given level must also be a beat at all smaller levels of the metrical structure.
- **MWFR 3** At each metrical level, strong beats are spaced either two or three beats apart.
- **MWFR 4** Each metrical level must consist of equally spaced beats.

**GTTM’s metrical preference rules (MPRs)** consist of:

- **MPR 1** Prefer a parallel metrical structure where groups, or their subgroups, can be constructed as parallel.
- **MPR 2** Weakly prefer a metrical structure in which the strongest beat in a group appears relatively early in the group.
- **MPR 3** Prefer a metrical structure in which beats, in a given level, coincide with the inception of notes that are strong beats of that level.
Chapter 3. Lerdahl’s Model of Tonal Tension  A Generative Theory of Tonal Music (GTTM)

MPR 4 Prefer a metrical structure in which beats that are accentuated in a given level are strong beats of that level.

MPR 5 Prefer a metrical structure in which a relatively strong beat occurs at the inception of either a long note, a long dynamic, a long slur, a long pattern of articulation and/or a long duration of the harmony.

MPR 6 Prefer a metrically stable bass.

MPR 7 Prefer a metrical structure in which cadences are metrically stable.

MPR 8 Prefer a metrical structure in which a suspension is on a stronger beat than its resolution.

MPR 9 Prefer a metrical structure which results in a stable organisation of the time-spans.

MPR 10 Prefer binary metrical structures where at each level every other beat is strong.

MPR 4 concerns accentuated notes. In GTTM, an accent represents stress in the musical discourse. It can be caused by either the emphasis of an event, known as a phenomenal accent, by a point of melodic or harmonic gravity, known as a structural accent, or by a relatively strong beat in the metrical context, known as a metrical accent.

Figure 3.3 shows the metrical structure of the first phrase of the theme in Mozart’s Ah vous dirai-je, Maman, K. 265/300e. The beats are represented in the form of a metrical grid.

In Figure 3.3, level a represents the semi-phrases in the theme, level b represents the whole-note level, level c represents the half-note level, level d represents the quarter-note level and level e represents the sixteenth-note level. An additional level, representing the eighth-note level, could have been included between level d and level e. However, this additional level is not shown in Figure 3.3 for the sake of simplicity.

The time-signature in Mozart’s theme is 2/4; that is to say, listeners will most likely perceive the first beat in each measure as a strong one. Thus, according to MPR 3, level c considers these strong beats, which, in addition, take place at the inception of each
harmonic change in the first semi-phrase, so MPR 5 is also taken into consideration. This structure is then replicated in the second semi-phrase so that both semi-phrases are parallel at this level, as proposed by MPR 1. Likewise, note that level c also accounts for a stable metrical bass and as proposed by MPR 6.

Considering MWFR 1 and MWFR 2, each attack point (i.e. each quarter-note) is included in level d, which leaves, in addition, a parallel distribution of the subgroups within level c, as proposed in MPR 1. Level a and level b represent the strongest beats according to patterns of articulation and harmonic rhythm, as proposed by MPR 5; that is to say, they represent the subdivision of the theme into two semi-phrases, each of which is, in turn, subdivided into parallel motifs.

Figure 3.3 also includes a small representation of level e in the last two measures of the theme. This level follows MWFR 1, as it represents the corresponding attack points that refer to the smallest metrical level. As proposed by Lerdahl and Jackendoff (1983), only the relevant² beats at this point in the theme are shown in Figure 3.3.

²MWFRs may be violated in some cases. More specifically, they may be violated at metrical levels beyond the tactus. The tactus is the most prominent level of the metrical hierarchy. That is to say, it is the one listeners tend to focus primarily on. In Figure 3.3, the tactus may correspond to either level c or level d. That is why the missing beats in level e are omitted in Figure 3.3 and why they can be seen as not being
Chapter 3. Lerdahl’s Model of Tonal Tension  A Generative Theory of Tonal Music (GTTM)

Notice that the metrical structure in Figure 3.3 matches all MWFRs.

3.2.3 | The time-span reduction

In order to hierarchically reduce a piece of music, the third step GTTM takes is to use the grouping and metrical analyses, those introduced in Sections 3.2.1 and 3.2.2, to rank the piece’s content into a first hierarchy. This first hierarchy expresses the relative structural importance of the events in the piece. To produce this hierarchy, GTTM incorporates the time-span reduction component.

Recall the example in Figure 3.1, which was presented at the beginning of Section 3.2 to introduce the concepts of reduction and hierarchy. In this example, the notes in Mozart’s Twinkle, twinkle theme were organised into different levels depending on their degree of importance. In each level, less important notes were absorbed by their surrounding more important notes. These dominance-subordination relations, which are the basis of a hierarchical structure, can be represented in the form of a tree. This is shown in Figure 3.4. For instance, see the notes in Figure 3.4c that are reduced to generate Figure 3.4d. The subordination of these less important notes to their corresponding dominating note can be represented by a short tree branch that connects to the corresponding longer tree branch. This processes is repeated in Figure 3.4 in higher levels of the reduction to represent similar dominance-subordination relations.

The shorter branches in Figures 3.4a and 3.4b are known as right branches. This type of branches signifies the subordination of an event to a preceding event. Similarly, the shorter branches in Figure 3.4c are known as left branches. This type of branches signifies the subordination of an event to a succeeding event.

It should be noted that the tree representations in Figure 3.4c are just examples of one way the subordinate branches can connect to the dominating branches. For instance, the branches of notes f, e and d may connect to each other before they connect to the branch of note c. At this point in the discussion, it is unclear which of the possible connections would be the one inferred by listeners. And so, it is GTTM’s aim to define the rules relevant in the current context.
Figure 3.4: An example of possible hierarchical tree structures from the reduction of the first phrase of the theme in Mozart’s *Ah vous dirai-je, Maman*, K. 265/300e.

to estimate the most appropriate branching connections. Notice, however, that the *time-span reduction* is just a first hierarchy that exclusively focuses on rhythmic characteristics. A second hierarchy will be introduced in Section 3.2.4, where melodic and harmonic characteristics are also taken into consideration.

The branches in Figure 3.4 only concern local connections. These branches can be combined into a single tree that represents all hierarchical relations in a piece of music. In order to estimate this tree, with regards to rhythmic hierarchies, GTTM combines the metrical and grouping analyses, those discussed in Sections 3.2.1 and 3.2.2. This combination represents the segmentation of the time-spans in a piece of music with regards to both the organisation of its events into groups and its metrical hierarchical relations. In GTTM, this representation is known as the *time-span segmentation*.

In order to produce the hierarchical (tree) representation of the rhythmic elements in a piece of music, GTTM’s *time-span reduction* incorporates two sets of rules: time-span reduction well-formedness rules (TSRWFRs) and time-span reduction preference
rules (TSRPRs). The former establish formal principles that time-spans, in a time-span segmentation, must show to be considered part of the reduction. The latter establish formal principles to select, from the possible relations of reduction that can be assigned to the musical discourse, those relations that would most likely correspond to the listeners’ intuitions.

TSRWFRs and TSRPRs are introduced below, although some of the rules have been summarised. For a full description of the rules, and how they are derived, see Lerdahl and Jackendoff (1983, ch.7).

GTMM’s time-span reduction well-formedness rules (TSRWFRs) consist of:

**TSRWFR 1** For every time-span, there is an event that is the head to which all other events in the time-span are subordinated to.

**TSRWFR 2** If a time-span does not contain any other time-spans, the head is whatever event occurs in the time-span.

**TSRWFR 3** If a time-span contains other time-spans, the head of the former is one of the events contained in one of the latter.

**TSRWFR 4** If a cadence is directly subordinated to the head of a time-span, the final is also directly subordinated to the head and the penult is directly subordinated to the final.

By head of a time-span, GTTM refers to the single structurally most important event.

TSRWFR 4 refers to the concept of cadence. In GTTM, a cadence is defined as a sign, or conventional formula, that marks and articulates the end of groups from phrase level to the most global levels of musical structure. The last event of a cadence is labelled “the final” and its preceding event “the penult”. GTTM also defines a cadenced group as a group that at some level of reduction reduces to two elements, the second of which is a cadence.
GTTM’s *time-span reduction* preference rules (TSRPRs) consist of:

1. **TSRPR 1** Prefer as head of a time-span an event that is in a relatively strong metrical position.

2. **TSRPR 2** Prefer as head of a time-span an event that is relatively consonant and closely related to the local tonic.

3. **TSRPR 3** Weakly prefer as head of a time-span an event that has a higher melodic pitch or a lower bass pitch compared to the rest of the events in the time-span.

4. **TSRPR 4** When two or more time-spans can be construed as motivically and/or rhythmically parallel, preferably assign them parallel heads.

5. **TSRPR 5** Prefer as head of a time-span an event that results in more stable choice of metrical structure.

6. **TSRPR 6** Prefer as head of a time-span an event that results in more stable choice of *prolongational reduction* (that is GTTM’s last component, which will be introduced in Section 3.2.4).

7. **TSRPR 7** Prefer as head of a time-span an event that forms a cadence.

8. **TSRPR 8** Prefer as head of a time-span an event that is relatively close to the beginning of the time-span if it can function as the beginning of a cadenced group.

9. **TSRPR 9** Prefer as head of a time-span an event that is at the ending of a cadenced group rather than at the beginning.

As in Section 3.2.1, by parallelism GTTM here refers to similar or even identical analyses. For instance, because of the similarities in Figure 3.4b, their corresponding branching connections are identical. We refer to these types of analysis as parallel.

Figure 3.5 shows the *time-span segmentation* and the *time-span reduction* of the first phrase of the theme in Mozart’s *Ah vous dirai-je, Maman*, K. 265/300e. A collection of brackets represents the *time-span segmentation*, derived from the combination of Figure 3.2.
and Figure 3.3, and the *time-span reduction* is represented in the form of a tree.

![Time-span reduction and segmentation of Mozart's theme](image)

Figure 3.5: GTTM's *time-span reduction* (top) and *time-span segmentation* (bottom) of the first phrase of the theme in Mozart's *Ah vous dirai-je, Maman*, K. 265/300e.

The only time-span at *level a* consists of the whole theme. The head of this span is the last note in the theme. This choice is in line with TSRPR 1, as the last chord is in a strong metrical position, TSRPR 2, as the last chord is a tonic chord, TSRPR 7, as the last chord is included in a cadence, and TSRPR 9, as the last chord is located at the structural ending.

The above arguments also explain why the last chord in the theme is also the head of the second time-span at *level b*. In the case of the first time-span at *level b*, the theme’s first chord is the head. The previous arguments also explain this decision, but, most notably, in this way the theme can be reduced to a I-V-I progression and so the first chord in the theme can be seen as a structural beginning of the progression. Because of this, and considering TSRPR 8, the first chord is selected as the head of the first time-span at *level b*.

Following the above arguments, the head of the first and last time-spans at *level c* are still the first and last chords in the theme, respectively. In the case of the second and third time-spans at *level c*, the first chords in the fourth and sixth measures are the heads,
respectively. That is because they are both tonic chords, which is relevant to TSRPR 2. Likewise, this distribution of the heads is in line with TSRPRs 4 and 5, as it leaves a parallel representation, both from a motivic and a rhythmic perspective, as well as a stable metrical structure, respectively.

Finally, the remaining non-connected events within the time-spans at level $d$ are connected to their respective heads. In this way, notes in every second beat (i.e. a weak beat) are connected to their corresponding first beat (i.e. a strong beat). Also, notice these connections take place between pairs of repeated notes and so parallel analyses are produced (i.e. the distribution of branches is the same within every bar, at the quarter-note level, but also at higher levels in measures 3-4 and 5-6), which are in line with TSRPR 4.

Notice that the time-span reduction in Figure 3.5 matches all TSRWFRs.

### 3.2.4 | The prolongational reduction

In order to hierarchically reduce a piece of music, the fourth and final step GTTM takes is to develop a second hierarchy by taking into account the stability of chords based on features such as local consonance/dissonance, inversion of chords or closeness to the tonic, among others. This second hierarchy, represented again as a tree, reflects qualitative patterns of tension and relaxation in the musical discourse. To produce this hierarchy, GTTM incorporates the prolongational reduction component.

As in Section 3.2.3, the right and left branches in this second hierarchy signify the subordination of an event to a preceding or a succeeding event, respectively. However, in the prolongational reduction, branches carry some more information. Right branches denote a qualitative increase of tension, whereas left branches denote a decrease.

The connections of the branches in the tree representation of the prolongational reduction also show some differences compared to those of the tree representation of the time-span reduction. GTTM’s prolongational reduction distinguishes between three different types of branching connections: the progression, the weak prolongation and the strong prolongation. Progression refers to the connection of the branches of two events that occurs when the harmonic roots of the two events are different (e.g. the events
involve different chords). Weak prolongation refers to the connection of the branches of
two events that occurs when the harmonic roots of the two events are identical but one of
the events is in a less consonant position (e.g. chords that involve different inversions or
different melodic notes). The weak prolongation is represented by a filled-in circle at the
joining of branches. Strong prolongation refers to the connection of the branches of two
events that occurs when all the notes of the two events, and their positions, are identical.
The strong prolongation is represented by an open circle at the joining of branches.

Figure 3.6 shows an example of the types of branching connections in a GTTM's prolongational reduction. This figure includes measures one, two and six from Mozart's Twinkle, twinkle theme used in the running example. In this figure, measure one illustrates a strong prolongation, represented with an open circle at the joining of branches. Despite the fact the notes in the bass line are in different octaves, they both have the same pitch class. Therefore, the two events in the first measure involve the same chord in the same inversion and so they are analysed as identical events. That is why measure one is considered a strong prolongation. Measure two illustrates a weak prolongation, represented with a filled-in circle at the joining of branches. The two events in this measure involve the same chord. However, the chord upon which the first event stands is in first inversion, whereas that of the second is in root position. That is why measure two is considered a weak prolongation. Finally, measure six illustrates a progression, represented with no circle at the joining of branches. Despite the fact the melodic notes of these events are the same, they involve different chords. That is why measure six is considered a progression.

Notice that the types of branching connections in GTTM's prolongational reduction have introduced the concept of prolongation. This concept is borrowed from Schenkerian analysis. In GTTM, prolongations are understood as connections that show how the music progresses. This idea names the prolongational reduction, whose function is to represent “the way music progresses from points of relative tension to points of relative repose” (Lerdahl and Jackendoff, 1983, p.179).

As discussed in Section 3.2.3, the representation of the time-span reduction concerns hierarchical relations with regards to the rhythmic elements in a piece of music. In
the case of the tree representation of the *prolongational reduction*, the tree concerns hierarchical relations with regards to the stability of melody and harmony. In order to determine the stability of the melody and the harmony of a given piece of music, GTTM defines a collection of stability conditions. These represent a scale of stability among pitch configurations derived from the tonal system. According to GTTM's stability conditions:

- right strong prolongations are more stable than right progressions,
- left progressions are more stable than left weak prolongations, which in turn are more stable than left strong prolongations,
- a connection between two events is more stable if they involve or imply a common diatonic collection,
- a connection between two events is melodically more stable if the interval between them is smaller,
- an ascending melodic progression is most stable as a right-branching structure; a descending one is most stable as a left branching structure,
- a connection between two events is harmonically more stable if their roots are closer on the circle-of-fifths, and
a progression that ascends along the circle-of-fifths is most stable as a right branching structure; one that descends along the circle-of-fifths or produces a subordinate-to-dominating relationship is most stable as a left-branching structure.

In order to produce the hierarchical (tree) representation of the melodic and harmonic elements in a piece of music, GTTM's *prolongational reduction* incorporates two sets of rules: *prolongational reduction well-formedness rules* (PRWFRs) and *prolongational reduction preference rules* (PRPRs). The former establish formal principles that prolongations must show to be considered part of the reduction. The latter establish formal principles to select, from the possible relations of prolongation that can be assigned to the musical discourse, those relations that would most likely correspond to the listeners' intuitions.

PRWFRs and PRPRs are introduced below, although the original form of some of the rules has been summarised. For a full description of the rules, and how they are derived, see Lerdahl and Jackendoff (1983, ch.9).

GTTM's *prolongational reduction* well-formedness rules (PRWFRs) consist of:

- **PRWFR 1** There is a single event in each grouping structure that functions as prolongational head.
- **PRWFR 2** A subordinate event can connect to its dominating event as either a strong prolongation, a weak prolongation or a progression.
- **PRWFR 3** An event in a grouping structure is either the prolongational head or a subordinate of an event within the current grouping structure.
- **PRWFR 4** The branches of a prolongational tree must never cross each other.

As in Section 3.2.3, by prolongational head GTTM refers to the single most important event in the current region of prolongational connections.
GTTM’s *prolongational reduction* preference rules (PRPRs) consist of:

**PRPR 1** Prefer as the prolongational head in the current region of prolongational connections an event that is a relatively important one in the time-span reduction.

**PRPR 2** Prefer a prolongational reduction that matches the time-span segmentation.

**PRPR 3** Prefer as the prolongational head in the current region of prolongational connections an event that attaches so as to form a maximally stable prolongational connection, according to the stability conditions, with one of the endpoints of the region.

**PRPR 4** Prefer a prolongational reduction in which the prolongational head in the current region of prolongational connections is a subordinate of the events dominating one of the regions’ endpoints.

**PRPR 5** Prefer a prolongational reduction in which parallel passages receive parallel analyses.

**PRPR 6** Prefer a prolongational reduction that matches the *basic form* and/or the *normative prolongational reduction*.

PRPR 6 concerns two particular structures that are of special interest in GTTM: the *basic form* and the *normative prolongational structure*.

The *basic form*, shown in Figure 3.7, represents the typical structure followed by the most global levels\(^3\) of the *time-span reduction* and the *prolongational reduction* in tonal pieces, and it is similar to Schenker’s *Ursatz* (i.e. I–V–I). The first event in the *basic form* will correspond to an opening I that attaches, as a weak prolongation, to the final event, which will correspond to the closing I. The event in the middle will correspond to a dominant V that attaches to the final I through as a left-branch progression. This is

---

\(^3\)According to Lerdahl and Jackendoff (1983, p.189), the *basic form* not only summarises the content of most tonal pieces but also appears in subordinate grouping levels such as themes and phrases. They claim that “much of the unity of tonal music depends on this fact”.

---

64
because V is part of the cadence and the last I is the dominating event in its environment.

Lerdahl and Jackendoff (1983, pp.198-199) observed that there is a tree pattern that recurs constantly in tonal music. In this tree pattern, which they labelled as the normative prolongational structure, a piece (or phrase) begins in relative repose, it then builds toward tension and, in the end, relaxes into a resolving cadence. They believe that:

\[
\text{this is the most essential way in which the idiom achieves the aesthetic effects of balance and closure, (…) any phrase or section that meets this condition possesses a degree of completeness; any that does not sounds unbalanced or unfulfilled in terms of tension and relaxation.}
\]

The normative prolongational structure consists of the representation of this tensing-relaxing pattern according to the branching connections shown in Figure 3.8. Notice that it starts with a right-branching progression, indicating an rise in tension, and it ends with a cadential preparation and a cadenced group, both as a double left-branching progression, indicating a fall in tension.

Finally, the basic form and the normative prolongational structure may be combined into a single prolongational structure, known as basic form + normative prolongational structure. This is shown in Figure 3.9, which consists of a tree where the basic form is
elaborated so as to fulfill the normative prolongational structure. According to Lerdahl and Jackendoff (1983, p.199), “this configuration feels more complete than either the basic form or normative structure alone”.

The basic form + normative prolongational structure may not always include the second branch on the left in Figure 3.9. This event represents the reprise of opening material in the tonic, which was a compositional strategy often used in the Classical period.

Figure 3.10 shows the prolongational reduction of the first phrase of the theme in Mozart’s *Ah vous dirai-je, Maman*, K. 265/300e. The prolongations are represented by branches in a tree with different types of nodes. Figure 3.10 also includes the labels of the chords in the theme. These labels have been double-checked by an expert musicologist (personal communication, February, 2021).

![Figure 3.10: GTTM's prolongational reduction and chord labels of the first phrase of the theme in Mozart's *Ah vous dirai-je, Maman*, K. 265/300e.](image)

Notice that the prolongational reduction tree in Figure 3.10 is almost identical to that of the time-span reduction in Figure 3.5. They only differ on the branches in the penul-

---

4It is unclear whether the first chord in measure five is dominating the second one or the other way around. This has been preferred in Figure 3.10 to match the time-span reduction in Figure 3.5. However, in an equally good analysis, measure five could have parallel prolongations to those in measure seven, because of the duplication of the frequency of chord changes in the theme’s second phrase.
timate measure. In this way, except for the last measure, the prolongational reduction matches PRPRs 1 and 2. Likewise, considering the theme’s parallelisms, it also matches PRPR 5.

The prolongational branches in the penultimate measure differ from those in the time-span reduction so that they match PRPR 3, as the dominant chord, V, is relatively more important than its preceding chord, ii; but, most importantly, the reduction now matches PRPR 6. Notice how, except for measures five and six, the prolongational reduction follow a basic form + normative prolongational structure.

Finally, notice that the prolongational reduction in Figure 3.10 matches all PRWFRs.

3.3 | Tonal Pitch Space (TPS)

Lerdahl’s (2004) Tonal Pitch Space (TPS) presents a model of pitch space that accounts for perceived proximity among three dimensions, pitches, chords and regions, and treats all three within one single framework. To do so, TPS is developed at three different levels of analysis: the pitch-class level, the chordal level and the regional level. These are reviewed in Sections 3.3.1, 3.3.2 and 3.3.2, respectively.

3.3.1 | The pitch-class level

TPS fundamental construct is its basic space, which could be described as a hierarchy of stability among pitch configurations. Given a chord label, TPS’s basic space will assign its pitch classes into five different spaces:

(a) octave,

(b) fifths,

(c) triadic,

(d) diatonic, and

(e) chromatic.
For example, Figure 3.11 shows the basic space of the tonic chord in the key of C major (i.e I/C). The chord’s root, pitch class c, is assigned the octave space (level a), the chord’s fifth, pitch class g, is assigned the fifths space (level b), the chord’s third, pitch class e, is assigned the triadic space (level c),\(^5\) the remaining pitch classes that are included in C major’s diatonic scale, pitch classes d, f, a and b, are assigned the diatonic space (level d) and the remaining pitch classes that are included in the chromatic scale, pitch classes c\(^\#\)/d\(^b\), d\(^\#\)/e\(^b\), f\(^\#\)/g\(^b\), g\(^\#\)/a\(^b\) and a\(^\#\)/b\(^b\), are assigned the chromatic space (level e).

\[\begin{array}{c|cccccccccc}
\text{level a} & x \\
\text{level b} & x & x \\
\text{level c} & x & x & x \\
\text{level d} & x & x & x & x & x & x & x & x & x & x \\
\text{level e} & c & c\(^\#\)/d\(^b\) & d & d\(^\#\)/e\(^b\) & e & f & f\(^\#\)/g\(^b\) & g & g\(^\#\)/a\(^b\) & a & a\(^\#\)/b\(^b\) & b \\
\end{array}\]

Figure 3.11: TPS’s basic space oriented toward I/C.

Notice that the levels in the basic space are cumulative as pitch classes assigned to a particular level are also found in the lesser stable levels.

TPS’s basic space can also be represented by assigning a degree of embedding to each level. This will result in a more compact representation of the basic space, as seen in Figure 3.12.

\[\begin{array}{cccccccccc}
c & c\(^\#\)/d\(^b\) & d & d\(^\#\)/e\(^b\) & e & f & f\(^\#\)/g\(^b\) & g & g\(^\#\)/a\(^b\) & a & a\(^\#\)/b\(^b\) & b \\
5 & 1 & 2 & 1 & 3 & 2 & 1 & 4 & 1 & 2 & 1 & 2 \\
\end{array}\]

Figure 3.12: TPS’s basic space (compact version) oriented toward I/C.

To account for pitch-class proximity, one can calculate the distance between pitch classes in the basic space both vertically and horizontally. Given two pitch classes, x and y, the vertical distance between them will be given by the difference of their respective degrees of embedding (as those shown in Figure 3.12). To calculate the horizontal distance between them, one must count the number of steps that separate them at a given

\(^5\)In seventh chords, the seventh tone is also assigned this space.
level of the basic space. That is to say, there exists a different value of horizontal distance between two pitch classes for each level in the basic space.

3.3.2 | The chordal level

A transition between two chords, $x$ and $y$, within the same tonal region, can be represented by the horizontal steps that the basic space of $x$ needs to take to reach the basic space of $y$.

For example, Figure 3.13 illustrates the transition between the chord of C major and the chord of G major, both in the key of C major.

As seen in Figure 3.13, in the chord of C major, level $a$ is assigned to pitch class $c$ (see ($x_a$) in Figure 3.13), whereas, in the chord of G major, level $a$ is assigned to pitch class $g$. In Figure 3.13, this transition is denoted by an arrow at level $a$, where $x_a$, at pitch class $c$, is moved to $x_a$, at pitch class $g$. Level $a$, at pitch class $c$, will no longer be assigned a pitch class in the basic space of the chord of G major. This is denoted by ($x_a$) in Figure 3.13.

In the chord of C major, level $b$ is assigned to pitch class $g$ (see $x(b)$ in Figure 3.13), whereas, in the chord of G major, level $b$ should be assigned to pitch class $d$. In Figure 3.13, this transition is denoted by an arrow at level $b$, where $x_b$, at pitch class $g$, is moved to $x_b$, at pitch class $d$. Level $b$, at pitch class $g$, will still be assigned a pitch class because of its newly assigned $x_a$, which recall is present in all levels below it. This is denoted by $x(b)$ in Figure 3.13.

In the chord of C major, level $c$ is assigned to pitch class $e$ (see $x_c$ in Figure 3.13), whereas, in the chord of G major, level $c$ should be assigned to pitch class $b$. In Figure 3.13,
this transition is denoted by an arrow at level c, where $x_c$, at pitch class e, is moved to $x_e$, at pitch class b. Level c, at pitch class e, will no longer be assigned a pitch class in the basic space of the chord of G major. This is denoted by $(x_c)$ in Figure 3.13.

Notice that all the above transitions, those denoted by arrows in Figure 3.13, correspond to steps in the diatonic circle-of-fifths: at level a, there is the transition $c \rightarrow g$; at level b, there is the transition $g \rightarrow d$; and, at level c, there is the transition $e \rightarrow b$. Notice how each of these transitions corresponds to one step to the right in the diatonic circle-of-fifths (see Figure 3.15).

![Figure 3.14: Major-key diatonic circle-of-fifths (chord label format).](image)

![Figure 3.15: C major’s diatonic circle-of-fifths (roots format).](image)

Based on the above observation concerning the motion by fifths in the basic space, TPS develops the following rule to calculate the distance between chords within the same tonal region:

$$\delta(x \rightarrow y) = j + k$$  \hspace{1cm} (3.1)

where $\delta(x \rightarrow y)$ refers to the distance between chords $x$ and $y$, $j$ refers to the shortest number of steps, from $x$ to $y$, in the diatonic circle-of-fifths and $k$ refers to the total difference between distinctive pitch classes in the basic space of $y$ compared to those in the basic space of $x$.

For instance, let us calculate $\delta(I/C \rightarrow V/C)$. As shown in Figure 3.14, in the major-key diatonic circle-of-fifths, $V$ is one step to the right from I, so $j = 1$. To calculate $k$, let us
consider the differences of all twelve weightings from the basic spaces of both chords and add together only those that are greater in the case of the destination chord (i.e. V/C).

The basic space oriented toward I/C is shown in Figure 3.12. The basic space oriented toward V/C is shown in Figure 3.16.

\[
\begin{array}{ccccccccccc}
 \text{c} & \text{c}^\#/\text{d} & \text{d} & \text{d}^\#/\text{e} & \text{e} & \text{f} & \text{f}^\#/\text{g} & \text{g} & \text{g}^\#/\text{a} & \text{a} & \text{a}^\#/\text{b} & \text{b} \\
 2 & 1 & 4 & 1 & 2 & 2 & 1 & 5 & 1 & 2 & 1 & 3 \\
\end{array}
\]

Figure 3.16: TPS’s basic space (compact version) oriented toward V/C.

Therefore:

\[
k = (2 - 5) + (1 - 1) + (4 - 2) + (1 - 1) + (2 - 3) + (2 - 2) + (1 - 1) + (5 - 4) + (1 - 1) + (2 - 2) + (1 - 1) + (3 - 2) = 2 + 1 + 1 = 4
\]

Notice that those pairs where the subtraction was either negative or zero have been struck through. Thus:

\[
\delta(\text{I/C} \rightarrow \text{V/C}) = j + k = 1 + 4 = 5.
\]

So far, the idea of “chordal step” (i.e. distance between two chords) has only been accounted for based on the diatonic circle-of-fifths. However, Lerdahl (2004, p.57) claims that “root motion by thirds is next in proximity to motion by fifths”. In this way, he develops a diatonic circle-of-thirds, similar to that in Figure 3.14, where each step represents an interval of a third instead of an interval of a fifth. By combining these two diatonic circles, Lerdahl arrives at the graphical representation of the chordal space, which is shown in Figure 3.17.

In Figure 3.17, the vertical axis represents the diatonic circle-of-fifths and the horizontal axis the diatonic circle-of-thirds. The chords in the latter circle are more widely spaced, which is reflected by a more widely spaced horizontal axis in the chordal space. The area enclosed in Figure 3.17 represents the centre of the space, known as the “chordal core”.

71
Figure 3.17: Graphical representation of the chordal space.

The chordal space (i.e. Figure 3.17) is one of TPS’s key contributions, as it provides us with a clear and simple representation of the inter-chord distances in the tonal system. Likewise, it is useful to represent harmonic progressions as paths through the space.

3.3.3 | The regional level

In the example in Section 3.3.2, the last chord is a G major chord labelled as V in the key of C major. But, what if this chord is labelled as a tonic chord in the key of G major (i.e. I/G)? To account for transitions between chords in different regions, TPS incorporates a new rule:

\[
\delta(x \rightarrow y) = i + j + k
\]  

(3.2)

where \(\delta(x \rightarrow y)\), \(j\) and \(k\) are borrowed from rule 3.1 and \(i\) refers to the shortest number of steps, from the key \(x\) is in to the key \(y\) is in, in the chromatic circle-of-fifths.

For instance, let us calculate \(\delta(I/C \rightarrow I/G)\). As the key of G major is one step to the right from the key of C major in the chromatic circle-of-fifths, \(i = 1\) (see Figure 3.19).

As the roots of I/C and I/G are still c and g, respectively, \(j = 1\), as shown in Figure 3.15. Finally, if calculating the differences between the basic spaces of I/C and I/G, which are shown in Figure 3.12 and Figure 3.20, respectively, \(k = 5\).

Therefore:

\[
\delta(I/C \rightarrow I/G) = i + j + k = 1 + 1 + 5 = 7.
\]
Notice that, in Section 3.3.2 and here, two different values of distance have been calculated for the chord transition C major → G major. In the first example, G major is labelled as V/C and its distance to the tonic of C major is equal to 5. In the second example, G major is labelled as I/V and its distance to the tonic of C major is equal to 7. Does this mean listeners may perceive a different distance between chords depending on their label?

Labelling a G major chord as either V/C or as I/V will not change how listeners perceive the chord. Chords should be labelled according to the musical context. If a G major chord is played in a piece of music where the key of C major was previously established, the G major chord would probably be perceived by the listeners as V/C. However, if in that same piece there is a modulation towards the key of G major, listeners may perceive a new G major chord as a I/V if the new key has been established.

According to Lerdahl (2004, p.74), Rules 3.1 and 3.2 require the least cognitive effort and so TPS defines a new rule concerning this issue:

---

Figure 3.18: Major-key chromatic circle-of-fifths (key label format).

Figure 3.19: Major-key chromatic circle-of-fifths (roots format).

Figure 3.20: TPS's basic space (compact version) oriented toward I/G.
principle of the shortest path

“The pitch-space distance between two events is preferably calculated to the smallest value”.

In the same way Section 3.3.2 developed the chordal space, Lerdahl also developed a graphical representation of the regional space. To do so, he calculated the closest keys given a major key as reference. Let I/I be the tonic chord of a reference key, the tonic chords of its closest keys are I/V, I/IV, i/vi and i/i (their distance to I/I, δ, is all equal to 7). Lerdahl placed on one axis the tonics related to the references by the circle-of-fifths (see V and IV in Figure 3.18) and the relative and parallel minor relationships (vi and i) on the other axis. By recursively calculating all the chords whose distance, δ, is equal to 7 from these tonic chords, he generated the whole regional space, which is shown in Figure 3.21.

\[
\begin{array}{ccccccc}
\sharp ii & \sharp IV & \sharp iv & VI & vi & I & i \\
\sharp vi & VII & vii & II & ii & IV & iv \\
\sharp i & III & iii & V & v & bVII & bvi \\
\sharp iv & VI & vi & I & i & bIII & biii \\
vii & II & ii & IV & iv & bVI & bvi \\
iii & V & v & bVII & bvi & bII & bii \\
vI & I & i & bIII & biii & V & v \\
\end{array}
\]

Figure 3.21: Graphical representation of the regional space.

In Figure 3.21, the elements are equidistant because the vertical and horizontal distance between them, δ, is the same in all directions. The area enclosed in Figure 3.21 represents the original keys before any transpositions of δ were calculated.

\(^{6}\text{Notice that the Roman numeral labels in the regional space represent keys, as opposed to the chordal space where they represented chord degrees.}\)
The *regional space* (i.e. Figure 3.21) is another of TPS’s key contributions, as it provides us with a clear and simple representation of the inter-key distances in the tonal system. Likewise, it is useful to represent harmonic progressions across different keys as paths through the space.

### 3.4 | Model of Tonal Tension (MTT)

Lerdahl’s (2004) Model of Tonal Tension (MTT) presents a collection of methods to quantify the patterns of tension and relaxation, in a piece of tonal music, according to GTTM’s and TPS’s findings.

MTT consists of a harmonic sequential model, a harmonic hierarchical model and a melodic model. These are reviewed in Sections 3.4.1, 3.4.2 and 3.4.3, respectively. Lastly, MTT’s empirical evaluation is reviewed in Section 3.5.

#### 3.4.1 | The harmonic sequential model

According to Lerdahl (2004, p.143), “listeners understand music both sequentially and hierarchically”. Concerning the former, he developed the following rule to predict the degree of tonal tension in a given piece of music:

\[
T_{seq}(y) = \delta(x_{prec} \rightarrow y) + T_{diss}(y) \tag{3.3}
\]

where \(T_{seq}(y)\) refers to the degree of *sequential tension* at \(y\), which represents the target chord; \(\delta(x_{prec} \rightarrow y)\) refers to the TPS distance between \(x_{prec}\), the chord that immediately precedes \(y\), and \(y\), calculated as in Rule 3.2; and \(T_{diss}(y)\) represents the degree of *surface tension* at \(y\), calculated according to the following rule:

\[
T_{diss}(y) = sc.dg + inv. + nh.t. \tag{3.4}
\]
where \( sc.dg \), which stands for scale degree, is equal to 1 if \( y \)’s melodic voice corresponds to the chord’s third or fifth pitch class, 0 otherwise; \( inv. \), which stands for inversion, is equal to 2 if \( y \) is in first or second inversion, 0 otherwise; and \( nh.t. \), which stands for non-harmonic tones, is equal to 4 for every pitch class in \( y \) that is a chromatic non-chordal tone (i.e. a pitch class from the chromatic scale that does not belong to the current diatonic scale), to 3 for every pitch class in \( y \) that is a diatonic non-chordal tone (i.e. a pitch class from the current diatonic scale that does not belong to \( y \)), to 1 if \( y \) is labelled as a seventh chord and \( y \) includes a seventh, 0 otherwise.

MTT’s sequential tension, \( T_{\text{seq}} \), aims to predict the degree of tonal tension from one event to the next. As seen in Rule 3.3, its calculation considers two components: the distance between two consecutive chords within TPS, \( \delta \), and the degree of surface tension at the target chord, \( T_{\text{diss}} \), which aims to estimate its degree of surface dissonance. According to Lerdahl and Krumhansl (2007, p.334):

\[
\text{tones} (\ldots) \text{ differ in their degree of perceived dissonance depending on intervallic structure, metrical position, duration, loudness, timbre, and textural location.}
\]

Thereby, Rule 3.4 is only an approximation of the degree of perceived dissonance at the target chord.

In order to illustrate the calculation of MTT’s sequential tension, \( T_{\text{seq}} \), let us consider again the first phrase of the theme in Mozart’s Ah vous dirai-je, Maman, K. 265/300e. Its prolongational reduction, already presented in Figure 3.10, is shown again in Figure 3.22. This last figure assigns, in addition, a number to each event in the theme.

The calculation of the degrees of sequential tension in Figure 3.22 is shown in Table 3.1. For each event in Figure 3.22, \( y \), Table 3.1 includes the number assigned to its preceding event, \( x_{\text{prec}} \). Following Rule 3.2, the values of \( i \), \( j \) and \( k \), concerning the transition \( x_{\text{prec}} \rightarrow y \), are calculated and are added up to calculate the distance between each pair of events, \( \delta(x_{\text{prec}} \rightarrow y) \). Following Rule 3.4, Table 3.1 also includes the weightings asso-
Figure 3.22: GTTM’s *prolongational reduction* and chord labels of the first phrase of the theme in Mozart’s *Ah vous dirai-je, Maman*, K. 265/300e (numbered version).

It should be noted that the first row in Table 3.1 includes an $x_{\text{prec}}$ event equal to 0 so that the first element in the theme, $y = 1$, gets assigned a value of $T_{\text{seq}}$. For this purpose, all the variables concerning $x_{\text{prec}} = 0$ are assigned a value of zero, and so $T_{\text{seq}}(y = 1) = 0$.

### 3.4.2 | The harmonic hierarchical model

In order to account for the listeners’ hierarchical understanding of music, Lerdahl developed another rule to predict the degree of tonal tension in a given piece of music. This rule is phrased as follows:

$$T_{\text{hier}}(y) = \delta(x_{\text{dom}} \rightarrow y) + T_{\text{diss}}(y) + T_{\text{inh}}(x_{\text{dom}})$$  \hspace{1cm} (3.5)
where \( T_{\text{hier}}(y) \) refers to the degree of *hierarchical tension* at \( y \), which represents the target chord; \( \delta(x_{\text{dom}} \rightarrow y) \) refers to the TPS distance between \( x_{\text{dom}} \), the chord that immediately dominates \( y \) in GTTM’s *prolongational reduction* tree, and \( y \), calculated as in Rule 3.2; \( T_{\text{diss}}(y) \) represents the degree of *surface tension* at \( y \), calculated as in Rule 3.4; and \( T_{\text{inh}}(x_{\text{dom}}) \) refers to the contribution of tension inherited by \( y \) from the branching connections in GTTM’s *prolongational reduction* tree that dominate \( x_{\text{dom}} \), calculated according to the following rule:

\[
T_{\text{inh}}(x_{\text{dom}}) = \sum_{i=1}^{n} \delta(z_{i+1} \rightarrow z_i) \tag{3.6}
\]

where \( z_i \) refers to the chords, \( z_1, z_2, z_3, \ldots, z_n \), that dominate \( x_{\text{dom}} \) in GTTM’s *prolongational reduction* tree; \(^7\) and \( \delta(z_{i+1} \rightarrow z_i) \) refers to the TPS distance between \( z_{i+1} \) and \( z_i \), calculated as in Rule 3.2.

Table 3.1: Stepwise calculation of the degrees of *sequential tension* in the first phrase of the theme in Mozart’s *Ah vous dirai-je, Maman*, K. 265/300e.

<table>
<thead>
<tr>
<th>( x_{\text{prec}} )</th>
<th>( y )</th>
<th>( \delta(x_{\text{prec}} \rightarrow y) )</th>
<th>( T_{\text{diss}}(y) )</th>
<th>( T_{\text{seq}}(y) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>0 0 0 0</td>
<td>0 0 0 0</td>
<td>0 0 0 0</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>0 0 0 0</td>
<td>0 0 0 0</td>
<td>0 0 0 0</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>0 0 0 0</td>
<td>1 2 0 0</td>
<td>3 3 3 3</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>0 0 0 0</td>
<td>1 0 0 0</td>
<td>1 1 1 1</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>0 1 4 5</td>
<td>1 0 0 0</td>
<td>1 1 1 1</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>0 0 0 0</td>
<td>1 2 0 0</td>
<td>3 3 3 3</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>0 1 4 5</td>
<td>1 2 0 0</td>
<td>3 3 3 3</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>0 0 0 0</td>
<td>1 0 0 0</td>
<td>1 1 1 1</td>
</tr>
<tr>
<td>8</td>
<td>9</td>
<td>0 2 7 9</td>
<td>1 0 0 0</td>
<td>1 1 1 1</td>
</tr>
<tr>
<td>9</td>
<td>10</td>
<td>0 3 4 7</td>
<td>1 0 0 0</td>
<td>1 1 1 1</td>
</tr>
<tr>
<td>10</td>
<td>11</td>
<td>0 2 6 8</td>
<td>1 0 0 0</td>
<td>1 1 1 1</td>
</tr>
<tr>
<td>11</td>
<td>12</td>
<td>0 3 4 7</td>
<td>1 0 0 0</td>
<td>1 1 1 1</td>
</tr>
<tr>
<td>12</td>
<td>13</td>
<td>0 1 4 5</td>
<td>0 2 0 0</td>
<td>2 2 2 2</td>
</tr>
<tr>
<td>13</td>
<td>14</td>
<td>0 1 4 5</td>
<td>1 0 0 0</td>
<td>1 1 1 1</td>
</tr>
<tr>
<td>14</td>
<td>15</td>
<td>0 0 0 0</td>
<td>0 0 3 3</td>
<td>3 3 3 3</td>
</tr>
<tr>
<td>15</td>
<td>16</td>
<td>0 1 4 5</td>
<td>0 0 0 0</td>
<td>0 0 0 0</td>
</tr>
</tbody>
</table>

\(^7\)That is to say, in the *prolongational* tree, \( x_{\text{dom}} \) is dominated by \( z_1 \), which is in turn dominated by \( z_2 \), which is in turn dominated by \( z_3 \) and so on.
For instance, in Figure 3.22, event 10 is immediately dominated by event 9, which is in turn dominated by event 11, which is in turn dominated by event 16. According to Rule 3.5, the degree of hierarchical tension at event 10 will correspond to:

$$T_{\text{hier}}(y = 10) = \delta(x_{\text{dom}} = 9 \rightarrow y = 10) + T_{\text{diss}}(y = 10) + T_{\text{inh}}(x_{\text{dom}} = 9)$$

where, according to Rule 3.6, $T_{\text{inh}}(x_{\text{dom}} = 9)$ will correspond to:

$$T_{\text{inh}}(x_{\text{dom}} = 9) = \delta(z_1 = 11 \rightarrow x_{\text{dom}} = 9) + \delta(z_2 = 16 \rightarrow z_1 = 11)$$

For the sake of easing the calculation of $T_{\text{inh}}$, Figure 3.22 includes the values of TPS distance, $\delta$, of each node in the prolongational tree. By taking a look at the nodes where the branches of the pairs of events 11 − 9 and 16 − 11 connect in Figure 3.22, one arrives at:

$$T_{\text{inh}}(x_{\text{dom}} = 9) = 8 + 0 = 8$$

The calculation of the degrees of hierarchical tension in Figure 3.22 is shown in Table 3.2. For each event in Figure 3.22, $y$, Table 3.2 includes the number assigned to its dominating event, $x_{\text{dom}}$. Following Rule 3.2, the values of $i$, $j$ and $k$, concerning the transition $x_{\text{dom}} \rightarrow y$, are calculated and are added up to calculate the distance between each pair of events, $\delta(x_{\text{dom}} \rightarrow y)$. Table 3.2 includes the degree of surface dissonance for the second event in each pair of events, $T_{\text{diss}}(y)$. These values are borrowed from Table 3.1. Table 3.2 also includes the numbers assigned to the events $z_{l+1}$ and $z_l$, those that dominate each $x_{\text{dom}}$. When an event $y$ is dominated by the global head of the prolongational tree in Figure 3.22 (i.e. if $x_{\text{dom}} = 16$), $y$ is assigned a non-existent $z_l$ in Table 3.2, as there is no event in the tree that dominates over the global head. Following Rule 3.6, a value of inherited contribution of tension is calculated for the first event in each pair of events, $T_{\text{inh}}(x_{\text{dom}})$. Finally, following Rule 3.5, Table 3.2 calculates the degree of hierarchical tension at the second event in each pair of events, $T_{\text{hier}}(y)$.

As in Table 3.1, Table 3.2 includes a row where a reference event is equal to 0. In this case, this event corresponds to $x_{\text{dom}}$ in the table’s last row. Notice how this value allows us to assign the last element in the theme, $y = 16$, with a degree of $T_{\text{hier}}$ equal to zero.
Table 3.2: Stepwise calculation of the degrees of hierarchical tension in the first phrase of the theme in Mozart's *Ah vous dirai-je, Maman*, K. 265/300e.

<table>
<thead>
<tr>
<th>y</th>
<th>$x_{dom}$</th>
<th>$\delta(x_{dom} \rightarrow y)$</th>
<th>$T_{dom}(y)$</th>
<th>$T_{inh}(x_{dom})$</th>
<th>$T_{hier}(y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>16 0 0 0 0 0</td>
<td>0</td>
<td>none</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1 0 0 0 0 0</td>
<td>0</td>
<td>16</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1 0 0 0 0 0</td>
<td>3</td>
<td>16</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>3 0 0 0 0 0</td>
<td>1</td>
<td>1,16</td>
<td>0+0=0</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>7 0 1 4 5 1</td>
<td>1</td>
<td>1,16</td>
<td>0+0=0</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>5 0 0 0 0 0</td>
<td>3</td>
<td>7,16</td>
<td>5+0+0=5</td>
<td>8</td>
</tr>
<tr>
<td>7</td>
<td>1 0 0 0 0 0</td>
<td>3</td>
<td>16</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>7 0 0 0 0 0</td>
<td>1</td>
<td>1,16</td>
<td>0+0=0</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>11 0 2 6 8 1</td>
<td>1</td>
<td>16</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td>9 0 3 4 7 1</td>
<td>1</td>
<td>11,16</td>
<td>8+0=8</td>
<td>16</td>
</tr>
<tr>
<td>11</td>
<td>16 0 0 0 0 0</td>
<td>1</td>
<td>none</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>11 0 3 4 7 1</td>
<td>1</td>
<td>16</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>13</td>
<td>14 0 1 4 5 2</td>
<td>1</td>
<td>16</td>
<td>5</td>
<td>12</td>
</tr>
<tr>
<td>14</td>
<td>16 0 1 4 5 1</td>
<td>none</td>
<td>0</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>15</td>
<td>11 0 0 0 0 0</td>
<td>3</td>
<td>16</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>16</td>
<td>0 0 0 0 0 0</td>
<td>none</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

### 3.4.3 | The melodic model

The basic construct of Lerdahl's melodic model is its anchoring space. This space aims to capture a pitch class’s anchoring strength in a given key. To build a key’s anchoring space, one should take the basic space of the key’s tonic chord and omit the fifths space (i.e. level b). According to Lerdahl (2004, pp.161-162), “the calculations work better if the fifth level (...) is suppressed for melodic modelling”.

For instance, to calculate the anchoring space in the key of C major, the fifths space in Figure 3.12 is omitted. To do so, the weighting assigned to pitch class g is transformed from a 4 to a 3, and, consequently, the weighting assigned to pitch class c is transformed from a 5 to a 4. This process is shown below in Figure 3.23.

![Figure 3.23: TPS's anchoring space (compact version) in the key of C major.](image)

To calculate the degree of melodic attraction between two pitch classes, in a given key, MTT develops the following rule:
\[
\alpha_{mel}(p_1 \rightarrow p_2) = \frac{s_2}{s_1} \cdot \frac{1}{n^2}
\]  
(3.7)

where \(\alpha_{mel}(p_1 \rightarrow p_2)\) refers to the degree of melodic attraction between pitch classes \(p_1\) and \(p_2\), being \(p_1 \neq p_2\); \(s_2\) refers to the weighting of \(p_2\) in the current anchoring space; \(s_1\) refers to the weighting of \(p_1\) in the current anchoring space; and \(n\) refers to the interval, in semitones, between \(p_1\) and \(p_2\).

To calculate the total degree of attraction between two consecutive chords, as a result of the combination of the voice-leading melodic attractions, MTT develops the following rule:

\[
\alpha(x_{prec} \rightarrow y) = c \cdot \frac{\sum_{a=1}^{b} \alpha_{mel}(x_{prec,a} \rightarrow y_a)}{\delta(x_{prec} \rightarrow y)}
\]  
(3.8)

where \(\alpha(x_{prec} \rightarrow y)\) refers to the total degree of attraction between two consecutive chords, \(x_{prec}\) and \(y\); \(c\) is a constant equal to 10; \(\alpha_{mel}(x_{prec,a} \rightarrow y_a)\) refers to the degree of melodic attraction between \(x_{prec,a}'s\) and \(y_a's\) respective pitch classes at voice \(a\), calculated as in Rule 3.7; \(a,...,b\) refer to the voices that may exist between \(x_{prec}\) and \(y\); and \(\delta(x_{prec} \rightarrow y)\) refers to the distance between chords \(x_{prec}\) and \(y\) calculated as in Rule 3.2.

The calculation of the degrees of total attraction, \(\alpha\), in Figure 3.22 is shown in Table 3.3 according to Rule 3.8.

From each pair of consecutive events in Figure 3.22, \(x_{prec} \rightarrow y\), Table 3.3 extracts two voices: a voice corresponding to the top voice in Figure 3.22 (melody) and voice corresponding to the bottom voice in Figure 3.22 (bass). For each of these voices, Table 3.3 includes their pitch classes, \(p_1\) and \(p_2\), their respective anchoring weightings according to the anchoring space in Figure 3.23, \(s_1\) and \(s_2\), and the distance between \(p_1\) and \(p_2\) in semitones, \(n\). Following Rule 3.7, the degree of melodic attraction, \(\alpha_{mel}(p_1 \rightarrow p_2)\), is calculated both in the melody and the bass. When a pair of events consists of a repeated pitch class (i.e. \(p_1 = p_2\)), the pair is assigned a null \(\alpha_{mel}\). That is because the distance that separates them, \(n\), will be equal to zero (notice that \(n\) is included in the
### Table 3.3: Stepwise calculation of the degrees of attraction in the first phrase of the theme in Mozart's Ah vous dirai-je, Maman, K. 265/300e.

<table>
<thead>
<tr>
<th>transition</th>
<th>voice</th>
<th>melodic attraction</th>
<th>total attraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 → 2</td>
<td>melody</td>
<td>c 4 c 4 0 null</td>
<td>0 null</td>
</tr>
<tr>
<td></td>
<td>bass</td>
<td>c 4 c 4 12 4/4-1/12</td>
<td></td>
</tr>
<tr>
<td>2 → 3</td>
<td>melody</td>
<td>c 4 g 3 7 3/4-1/7²</td>
<td>0 null</td>
</tr>
<tr>
<td></td>
<td>bass</td>
<td>c 4 e 3 4 3/4-1/4 ²</td>
<td></td>
</tr>
<tr>
<td>3 → 4</td>
<td>melody</td>
<td>g 3 g 3 0 null</td>
<td>0 null</td>
</tr>
<tr>
<td></td>
<td>bass</td>
<td>e 3 c 4 4 4/3-1/4 ²</td>
<td></td>
</tr>
<tr>
<td>4 → 5</td>
<td>melody</td>
<td>g 3 a 2 2 2/3-1/2²</td>
<td>5 .373</td>
</tr>
<tr>
<td></td>
<td>bass</td>
<td>c 4 f 2 5 2/4-1/5²</td>
<td></td>
</tr>
<tr>
<td>5 → 6</td>
<td>melody</td>
<td>a 2 a 2 0 null</td>
<td>0 null</td>
</tr>
<tr>
<td></td>
<td>bass</td>
<td>f 2 c 4 5 4/2-1/5²</td>
<td></td>
</tr>
<tr>
<td>6 → 7</td>
<td>melody</td>
<td>a 2 g 3 2 3/2-1/2²</td>
<td>5 .844</td>
</tr>
<tr>
<td></td>
<td>bass</td>
<td>c 4 e 3 4 3/4-1/4²</td>
<td></td>
</tr>
<tr>
<td>7 → 8</td>
<td>melody</td>
<td>g 3 g 3 0 null</td>
<td>0 null</td>
</tr>
<tr>
<td></td>
<td>bass</td>
<td>e 3 c 4 4 4/3-1/4²</td>
<td></td>
</tr>
<tr>
<td>8 → 9</td>
<td>melody</td>
<td>g 3 f 2 2 2/3-1/2²</td>
<td>9 .324</td>
</tr>
<tr>
<td></td>
<td>bass</td>
<td>c 4 d 2 2 2/4-1/2²</td>
<td></td>
</tr>
<tr>
<td>9 → 10</td>
<td>melody</td>
<td>f 2 f 2 0 null</td>
<td>7 .159</td>
</tr>
<tr>
<td></td>
<td>bass</td>
<td>d 2 b 2 3 2/2-1/3²</td>
<td></td>
</tr>
<tr>
<td>10 → 11</td>
<td>melody</td>
<td>f 2 e 3 1 3/2-1/1²</td>
<td>8 .4375</td>
</tr>
<tr>
<td></td>
<td>bass</td>
<td>b 2 c 4 1 4/2-1/1²</td>
<td></td>
</tr>
<tr>
<td>11 → 12</td>
<td>melody</td>
<td>e 3 e 3 0 null</td>
<td>7 .079</td>
</tr>
<tr>
<td></td>
<td>bass</td>
<td>c 4 a 2 3 2/4-1/3²</td>
<td></td>
</tr>
<tr>
<td>12 → 13</td>
<td>melody</td>
<td>e 3 d 2 2 2/3-1/2²</td>
<td>5 .458</td>
</tr>
<tr>
<td></td>
<td>bass</td>
<td>f 2 f 2 2 2/2-1/2²</td>
<td></td>
</tr>
<tr>
<td>13 → 14</td>
<td>melody</td>
<td>d 2 d 2 0 null</td>
<td>5 .75</td>
</tr>
<tr>
<td></td>
<td>bass</td>
<td>f 2 g 3 2 3/2-1/4²</td>
<td></td>
</tr>
<tr>
<td>14 → 15</td>
<td>melody</td>
<td>d 2 e 3 2 3/2-1/4²</td>
<td>0 null</td>
</tr>
<tr>
<td></td>
<td>bass</td>
<td>g 3 g 3 0 null</td>
<td></td>
</tr>
<tr>
<td>15 → 16</td>
<td>melody</td>
<td>e 3 c 4 4 4/3-1/4²</td>
<td>5 .221</td>
</tr>
<tr>
<td></td>
<td>bass</td>
<td>g 3 c 4 7 4/3-1/7²</td>
<td></td>
</tr>
</tbody>
</table>

Denominator in Rule 3.7. Table 3.3 also includes the TPS distance between the events in each pair, \(\delta(x_{\text{prec}} \rightarrow y)\). These distances are borrowed from Table 3.1. Finally, following Rule 3.8, Table 3.3 calculates the final degree of total attraction for each pair of events, \(\alpha(x_{\text{prec}} \rightarrow y)\). When the distance between the events in a pair, \(\delta\), is equal to zero, the pair is assigned a null total attraction, as in Lerdahl (2004). ⁸

---

⁸In Lerdahl and Krumhansl (2007), concerning the null issue, they repeat the value of \(\delta\) from the point at which the harmony last changed so that it is different from zero. Note, however, that, in the case of Figure 3.22, this modification will not be valid in most of the null pairs.
3.5 | Empirical evaluation

Lerdahl and Krumhansl (2007) carried out an empirical study of musical tension to investigate the correlation between the degrees of musical tension predicted by MTT and those perceived by human listeners. The participants and materials, the experimental procedure, the results and the main conclusions are discussed in Sections 3.5.1, 3.5.2, 3.5.3 and 3.5.4, respectively.

3.5.1 | Participants and materials

Lerdahl and Krumhansl recruited two groups of 14 and 15 participants trained in the Western tonal tradition, respectively, and asked them to record the degree of tension they perceive while listening to three pieces of music. The 14-participant group rated Wagner’s Grail theme from Parsifal and Bach’s chorale Christus, der ist mein Leben, BWV 95, and the 15-participant group rated a reduction of Chopin’s Prelude No. 9 in E major, Opus 28.

3.5.2 | Procedure

The first group of participants was asked to perform both a stop-tension and a continuous-tension task, which recall were introduced in Section 2.3.1. In the stop-tension task, participants listened to the first event in each piece and rated its degree of tension; then they listened to the first and second events in the piece and rated the degree of tension of the second event; then they listened to the first, second and third events in the piece and rated the degree of tension of the third event; and so on. According to Lerdahl and Krumhansl, both tasks entail some advantages and disadvantages. On one hand, the stop-tension task records a response that can be precisely associated with every event in the piece, but is highly time-consuming. On the other hand, the continuous-tension is less time-consuming and provides real-time tension responses, but there will be a lag time in the judgements provided by the participants which will have to be corrected.
Krumhansl (1996) carried out a study with a similar methodology where participants varied in the extent of their musical training. However, she concluded that training had little effect on the tension judged by the participants. Because of this conclusion, Lerdahl and Krumhansl (2007) only focused on musically trained participants and did not analyse their judgements from the point of view of the effect musical training may have had on the perception of musical tension.

3.5.3 | Results

Lerdahl and Krumhansl calculated a collection of regressions to investigate whether the judgements of tension provided by the participants fitted MTT’s predictions of tension. They calculated different regressions for the judgements of tension of each piece in order to arrive at “an explanatory account of why the model succeeds or fails at any given point in the analysis” (Lerdahl and Krumhansl, 2007, p.340). In these regressions, they tried different independent variables to better understand how well MTT’s theory matches experimental data.

In the first regression Lerdahl and Krumhansl calculated, they analysed the judgements of tension against the degrees of sequential tension, \( T_{seq} \), calculated as in Rule 3.3, in Wagner’s Grail theme from Parsifal in the stop-tension task. In all the regressions, participants’ judgements of tension were considered the dependent variable. In this first regression, \( T_{seq} \) was the independent variable.

The fit of the first regression was quite poor: \( R^2 = .08, df = (1, 7), R^2_{adj} = -.05 \) and \( p = .46 \). These results include: a value of \( R^2 \) as a goodness-of-fit measure (i.e. proportion of variation) of MTT’s predictions; the degrees of freedom, \( df \), associated with \( R^2 \), the first of which indicates the number of predictor (i.e. independent) variables there are in the regression and the second of which corresponds to the data points in the regression minus the first degree of freedom minus one; an adjusted value of \( R^2 \), \( R^2_{adj} \), so that it could be compared with other models for similar data that have different degrees of freedom;\(^9\) and the regression’s \( p \) value(s), which, by convention, indicates that \( R^2 \) may

---

\(^9\)According to Ratner (2009, p.240), “The \( R^2 \) is typically interpreted as ‘the percent of variation in one variable explained by the other variable’ or ‘the percent of variation shared between the two variables’.
be significant when \( p < .05 \).

Because of the poor fit of the first regression, Lerdahl and Krumhansl concluded that listeners did not hear Wagner’s *Grail* theme from *Parsifal* in a sequential manner. And so they continued the analysis with another regression.

In the second regression Lerdahl and Krumhansl calculated, attraction, \( \alpha \), calculated as in Rule 3.8, was the independent variable. The fit of this regression was improved, compared to that of the first one, but still was not good: \( R^2 = .35 \), \( df = (1, 7) \), \( R^2_{adj} = .26 \) and \( p = .09 \).

In order to achieve a best-fit solution, Lerdahl and Krumhansl explored the addition of sequential tension values, those used in the first regression, to the attraction values, those used in the second regression. To do so, they calculated a multiple regression where \( T_{seq} \) and \( \alpha \) were the independent variables. In this third regression, the fit was better than that of the first regression, but not better than the second one: \( R^2 = .35 \), \( df = (2, 6) \), \( R^2_{adj} = .13 \), \( \beta_t = .02 \), \( \beta_{t} = .96 \), \( \beta_{\alpha} = .58 \) and \( p_{\alpha} = .17 \). These results now include the coefficients in the linear model, \( \beta_t \) and \( \beta_{\alpha} \), of each predictor variable, \( T_{seq} \) and \( \alpha \), respectively. Likewise, each predictor variable includes its own \( p \) value.

Because of the poor fit of the first three regressions, Lerdahl and Krumhansl concluded that the strict sequential treatment of tension did not contribute to the fit of the data and so they abandoned the sequential tension approach. Instead, they decided to consider a hierarchical tension approach.

In order to explore the hierarchical approach, Lerdahl and Krumhansl calculated a fourth regression where hierarchical tension, \( T_{hier} \), calculated as in Rule 3.5, was the independent variable. Results achieved a better correlation than previous analyses: \( R^2 = .43 \), \( df = (1, 7) \), \( R^2_{adj} = .35 \) and \( p = .056 \).

The correlation of the fourth regression was better than that of previous regressions. However, it was still statistically not significant (i.e. \( p > .05 \)). Thus, Lerdahl and Krumhansl explored a new approach and calculated a fifth regression where \( T_{hier} \) and

\[ [\text{whereas}] \text{ the } R^2_{adj} \text{ penalises the statistic when unnecessary variables are included in the model. Specifically, the } R^2_{adj} \text{ adjusts the } R^2 \text{ for the sample size and the number of variables in the regression model. Therefore, the } R^2_{adj} \text{ allows for (...) a comparison between models with different numbers of variables and different sample sizes. Unlike } R^2, \text{ the } R^2_{adj} \text{ does not necessarily increase, if a predictor variable is added to a model} \]
\( \alpha \) were the independent variables. In this fifth regression, the correlation was the best and was statistically significant: \( R^2 = .97, df = (2, 6), R^2_{adj} = .97, \beta_t = .79, p_t < .0001, \beta_\alpha = .58 \) and \( p_\alpha < .0001 \).

From the analysis of the Grail theme data, Lerdahl and Krumhansl drew two conclusions: first, \( \alpha \) must be incorporated into the predictions; and, second, listeners perceive tension hierarchically more than sequentially. Following these conclusions, Lerdahl and Krumhansl analysed the data of the remaining pieces used in the study only using the last multiple regression, that where hierarchical tension, \( T_{hier} \), and attraction, \( \alpha \), are the predictor variables.

In their analysis of the judgements of tension provided by the participants concerning Bach’s chorale Christus, der ist mein Leben, BWV 95, the correlation with MTT’s predictions was strong and statistically significant: \( R^2 = .79, df = (2, 38), R^2_{adj} = .78, \beta_t = .67, p_t < .0001, \beta_\alpha = .47 \) and \( p_\alpha < .0001 \).

In their analysis of the judgements of tension provided by the participants concerning Chopin’s Prelude No. 9 in E major, Opus 28, the correlation with MTT’s predictions was poorer than those of the previous pieces: \( R^2 = .42, df = (2, 44), R^2_{adj} = .4, \beta_t = .62, p_t < .0001, \beta_\alpha = .11 \) and \( p_\alpha = .34 \). What is more, the correlation with \( \alpha \) was not statistically significant (i.e. \( p_\alpha > .05 \)).

Lerdahl and Krumhansl wondered what could give a better prediction of the participants’ judgements of tension of the Chopin prelude. They observed that the tension judgements followed the rise and fall of the contour of the melody. Therefore, they calculated a new variable: melodic contour, \( c \). To do so, they assigned pitch height in imitation of scale degrees (this can be done either diatonically or chromatically). That is to say, let \( c_4 \) be a reference pitch class. Its degree of melodic contour, \( c \), will be assigned a value of zero. Then, the degree of melodic contour assigned to the next chromatic note, \( c#_4 \), will be equal to 1. The degree assigned to \( d_4 \) will be equal to 2. And so on.

Lerdahl and Krumhansl observed that the correlation of melodic contour alone with the judgements of Chopin prelude was surprisingly robust: \( R^2 = .66, df = (1, 45), R^2_{adj} = .65 \).\(^{10}\)

\(^{10}\)Lerdahl and Krumhansl (2007, p.344) claim that the fact that this correlation was also strong is “all the more impressive” because, even though a correlation tends to decrease the more data points are considered, both \( T_{hier} \) and \( \alpha \) were statistically significant.
.65 and \( p < .0001 \). Therefore, they calculated a new multiple regression where \( T_{\text{hier}}, \alpha \) and \( c \) were the independent variables. Results showed that the incorporation of contour, \( c \), improved the correlation and that all the variables were statistically significant: \( R^2 = .67, \; df = (3, 43), \; R^2_{\text{adj}} = .65, \; \beta_t = .33, \; p_t = .003, \; \beta_\alpha = .22, \; p_\alpha = .02, \; \beta_c = .57 \) and \( p_c < .0001 \).

Lerdahl and Krumhansl re-calculated the regressions of Wagner’s and Bach’s pieces using the melodic contour factor. Interestingly, these new correlations were not statistically significant. Lerdahl and Krumhansl believe this may be a consequence of the melodic contours of these pieces moving in more intricate patterns.

3.5.4 | Conclusions

Two main conclusions can be drawn from Lerdahl and Krumhansl’s (2007) evaluation. First, results point to a good efficacy of the combination of MTT’s harmonic hierarchical model and melodic model to predict the degrees of tonal tension in a given piece of music. This, however, does not mean that MTT’s harmonic sequential model should be disregarded. In fact, Lerdahl and Krumhansl believe that since attraction contributes to the overall perceived tension, MTT reflects a balance of hierarchical and sequential listening.

Second, MTT offers a way to investigate the efficacy of pitch-reduction theories through the measurement of tonal tension. According to Lerdahl and Krumhansl (2007), listeners’ “unpremeditated awareness” does not concern a piece’s hierarchical structure but the patterns of tension and relaxation that arise from this structure. As a consequence, Lerdahl and Krumhansl (2007, p.356) believe that “the empirical validity of different prolongational analyses can be explored through the tension patterns they represent”. This last conclusion is of great importance in the present research as it provides us with a way to test the accuracy of generated GTTM structures. This approach was ratified in Lerdahl (2011) and will be used in the evaluation methodology introduced in Chapter 7.
3.6 | Summary and Conclusions

In order to seek answers to our Research Question, Chapter 3 has reviewed Lerdahl’s Model of Tonal Tension (MTT).

Section 3.1 has contextualised Lerdahl’s MTT. To do so, it has identified its components, Generative Theory of Tonal Music (GTTM), Tonal Pitch Space (TPS) and Model of Tonal Tension (MTT), and has discussed why these were derived.

Section 3.2 has reviewed GTTM’s components. To do so, the rhythmic components (grouping structure and metrical structure) and the reduction components (time-span reduction and prolongational reduction) were introduced after defining the main concepts they respectively consist of. In order to understand how each component is generated, its rules (well-formedness rules and preference rules) were introduced. Finally, to illustrate the application of GTTM in a given piece of music, an example was provided based on the first phrase of the theme in Mozart’s Ah vous dirai-je, Maman, K. 265/300e, upon which each GTTM component was generated.

Section 3.3 has reviewed TPS’s components. To do so, the Rules concerning TPS’s three different levels (pitch-class level, chordal level and regional level) were introduced and their application was illustrated with small examples.

Section 3.4 has reviewed MTT’s components. To do so, the Rules concerning MTT’s three different models (harmonic sequential model, the harmonic hierarchical model and melodic model) were introduced. To illustrate the application of MTT in a given piece of music, an example was provided based again on Mozart’s theme, upon which MTT’s models were applied.

Finally, Section 3.5 has reviewed MTT’s evaluation methodology and the main results in Lerdahl and Krumhansl (2007). These point to a good efficacy of the combination of MTT’s harmonic hierarchical model and melodic model to predict the degrees of tonal tension in a given piece of music. Likewise, they support the use of MTT’s tension calculations as a way to test the accuracy of generated GTTM structures.
Chapter 4

Automatically Calculating Tonal Tension

Challenges, implementation and evaluation

In this chapter, we show how to fully automate Lerdahl’s Model of Tonal Tension (MTT). To do so, we will:

(1) identify the challenges in Lerdahl’s MTT that may hinder its complete automation,

(2) design and implement a system capable of automatically calculating distances between chords within Lerdahl’s Tonal Pitch Space (TPS),

(3) design and implement a system capable of automatically calculating tension according to Lerdahl’s MTT, and

(4) evaluate the accuracy of the implemented systems.

The first step is addressed in Section 4.1, the second step is addressed in Section 4.2, the third step is addressed in Section 4.3 and the fourth step is addressed in Section 4.4.
4.1 | Why is it so challenging to fully automate Lerdahl’s Model of Tonal Tension?

In the last decade, there have been many attempts to automate Lerdahl’s work. These have been developed as part of different research projects, such as the study of the effect of chord structure in the perception of tension (Henry, 2017), the study of harmonic similarity (De Haas et al., 2013), the analysis of classical (Sakamoto et al., 2016) and jazz (Fukunari et al., 2016) harmonies or the automatic detection of cadences (Matsubara et al., 2018), among others.

Most existing attempts to automating Lerdahl’s work, such as those mentioned above, show at least one of the following limitations:

- only implementing MTT’s sequential component, $T_{seq}$, and not considering its hierarchical component, $T_{hier}$ (e.g. Henry (2017); Yoo and Lee (2006)); according to Lerdahl and Krumhansl (2007), the latter generally translates into more accurate predictions of tonal tension,

- not considering chord transitions across regions when calculating distances within TPS (e.g. De Haas et al. (2013)),

- not implementing MTT’s latest version (Lerdahl and Krumhansl, 2007) but a previous one (Lerdahl, 1988, 1996, 2004) (e.g. Fukunari et al. (2016); Matsubara et al. (2018); Sakamoto et al. (2016); Williams (2007)), when the former has proven more effective than the latter, and/or

- designing a straightforward implementation of MTT’s components that does not differentiate between enharmonic keys (e.g. Chang (2013)), which can lead to wrong values of TPS distance.

Despite the fact that Lerdahl’s original Model of Tonal Tension (MTT) was published more than two decades ago, an open source full implementation is not available yet. Why is that? There are four main challenges. These are discussed in Sections 4.1.1, 4.1.2, 4.1.3 and 4.1.4.
Chapter 4. Automatically Calculating Tonal Tension

4.1.1 | Calculating \( i \) in the Tonal Pitch Space

Chapter 3 introduced Rule 3.2 to calculate the distance between two chords across regions (i.e. \( \delta = i + j + k \)). In this rule, \( i \) refers to the shortest number of steps, from the key of the first chord to that of the second chord, in the chromatic circle-of-fifths. Prima facie, implementing a method to automatically calculate \( i \) may seem straightforward. However, how would this implementation deal with enharmonic keys?

Let us, for instance, consider the keys of C\(^\#\) major and D\(\flat\) major, which are enharmonic keys. If manually calculating the value of \( i \) from the key of C major to the above keys, one will get that \( i(I/C \rightarrow I/C^\#) = 7 \) whereas \( i(I/C \rightarrow I/D^\flat) = 5 \), as these correspond to the number of steps in the chromatic circle-of-fifths to the right and to the left, respectively, starting at I/C. Despite the fact that C\(^\#\) and D\(\flat\) share the same location in the chromatic circle-of-fifths, TPS considers that, starting at I/C, the former is reached by moving to the right in the circle, whereas the latter is reached by moving to the left. Why is that?

According to Schoenberg (1969), the keys of C\(^\#\) major and D\(\flat\) major are in different tonal regions, with regards to the key of C major, and so their respective values of inter-key distance are different. This view aims to illustrate that a modulation from the key of C major to the key of C\(^\#\) major would follow a longer path than that of a modulation to the key of D\(\flat\) major. That is to say, we follow different directions because we seek to reach the exact key label and not an enharmonic one. For instance, in a neighbour-key C major \( \rightarrow \) C\(^\#\) major modulation, we will do:

\[
C \rightarrow G \rightarrow D \rightarrow A \rightarrow E \rightarrow B \rightarrow F^\# \rightarrow C^\#
\]

which is a longer path compared to a neighbour-key C major \( \rightarrow \) D\(\flat\) major modulation:

\[
C \rightarrow F \rightarrow B^\flat \rightarrow E^\flat \rightarrow A^\flat \rightarrow D^\flat
\]

TPS adopts Schoenberg’s approach and that is why \( i(I/C \rightarrow I/C^\#) \) is different from \( i(I/C \rightarrow I/D^\flat) \). This is discussed in more detail in Section 4.2.1.

When calculating the number of steps between two keys in the chromatic circle-of-fifths, one will get two values: a value corresponding to the steps taken from the first
key to the right and another value corresponding to the steps taken from the first key to
the left. According to Rule 3.2, from the two possible numbers of steps in the chromatic
circle of fifths (i.e. steps to the right and steps to the left), \( i \) is defined as the shortest
one. In the above example, both \( C\# \) major and \( D\flat \) major share the same location in
the chromatic circle-of-fifths. Thus, a straightforward implementation of Rule 3.2 would
assign the distances from the key of \( C \) major to the keys of \( C\# \) major and \( D\flat \) major with
a value of \( i \) equal to 5, as it is the shortest value to the location the last two chords
share in the chromatic circle-of-fifths. However, this value will be wrong in the case of
\( i(I/C \rightarrow I/C\#) \), as it will be equal to 7, as discussed above.

The above issue will arise whenever the calculation of the value of \( i \) concerns an
enharmonic key and so this is the first challenge to be overcome when implementing
Lerdahl’s work.

4.1.2 | Calculating \( j \) in the Tonal Pitch Space

In Rule 3.2, \( j \) refers to the shortest number of steps, between two chords, in the diatonic
circle-of-fifths. Again, prima facie, implementing a method to automatically calculating
\( j \) may seem straightforward. However, how would this implementation deal with chords
from different regions?

Let us, for instance, consider the chords of \( C \) major and \( D\flat \) major, which are in dif-
ferent regions. Notice that in the diatonic circle-of-fifths centred at \( C \) major, which was
shown in Figure 3.15, the chord of \( D\flat \) major is not included. How would one calculate
the value of \( j \) between these two chords? Likewise, notice that the chord of \( D\flat \) major
may be labelled in many different ways. For instance, with regards to the key of \( C \) major,
it could be labelled as the tonic of the key of \( D\flat \) major, \( I/D\flat \), but also as the Neapolitan
chord, \( b\)II/C. How can we make sure that the values of \( j \) of the two chord labels in this
particular case are different?

The above issue will arise whenever the calculation of the value of \( j \) concerns chords
from different regions and so this is the second challenge to be overcome when imple-

\[1]\text{Notice this issue concerns TPS’s “principle of the shortest path”}.
menting Lerdahl's work.

4.1.3 | Representing the prolongational reduction

Chapter 3 introduced Rule 3.6 to calculate the inherited degree of tension, $T_{inh}$, that will contribute to the calculation of the total degree of hierarchical tension, $T_{hier}$. In order to calculate $T_{inh}$ within a given piece of music, the piece's *prolongational reduction* must be computationally represented in such a way it is accessible.

One of the most well-known computational representations of GTTM's components may be that of the Interactive GTTM Analyser (IGA)\(^2\) (Hamanaka and Tojo, 2009). IGA is a system designed to automatically generate GTTM's components from a given piece of music. To do so, IGA re-formalises GTTM's rules using numerical expressions, with adjustable control parameters, which weight the priority of application of each rule.\(^3\)

IGA's inputs and outputs are in MusicXML format (Good, 2001). To represent GTTM hierarchies, IGA uses XML's "element trees".\(^4\) Despite the fact XML's "element trees" may be practical to automatically calculating $T_{inh}$, their representation is rather obscure, sometimes hardly accessible and not intuitive at all. Thus, a new way of representing the *prolongational reduction* is needed and so this is the third challenge to be overcome when implementing Lerdahl's work.

4.1.4 | Calculating distances within the Tonal Pitch Space across regions

One of the key differences between TPS’s versions is that, originally (Lerdahl, 2004), to calculate the distance between two chords in different regions, one had to draw a

---

\(^2\)http://www.gttm.jp/

\(^3\)IGA’s current available version is based on the ATTA algorithm (as in Automatic Time-Span Tree Analyser) (Hamanaka et al., 2006) and the exGTTM algorithm (as in machine-executable extension of GTTM) (Hamanaka and Tojo, 2009). These algorithms are quasi-automatic, as they need from some preliminary manual intervention in order to automatically generate GTTM analyses given a piece of music. Hamanaka et al. (2007) investigated the full automation of these algorithms and implemented a new one, FATTA (as in Full Automatic Time-Span Tree Analyser). However, according to M. Hamanaka (personal communication, June, 2020), FATTA is computationally very expensive and it usually requires more than twenty-four hours to produce a GTTM analysis of a single piece of music.

\(^4\)For instance, see: https://docs.python.org/3/library/xml.etree.elementtree.html
path through contiguous regions in the regional space until both chords were connected. In this way, the calculation of the total distance between the two chords was split into the calculation of a collection of intermediate distances. For instance, in Lerdahl (2004, p.68), the distance between the tonic chords of the keys of C major and D♭ minor is calculated as:

\[
\delta(I/C \rightarrow i/d♭) = \\
\delta_1(I/C \rightarrow i/c) + \delta_2(i/c \rightarrow I/Ab) + \delta_3(I/Ab \rightarrow I/D♭) + \delta_4(I/D♭ \rightarrow i/d♭) = \\
7 + 9 + 7 + 7 = 30
\]

It was later discovered in TPS’s final version (Lerdahl and Krumhansl, 2007), that directly calculating the distance as in Rule 3.2 (i.e. \(\delta = i + j + k\)) is more accurate than the above splitting approach. This conclusion was confirmed by F. Lerdahl (personal communication, September, 2019).

Concerning the automation of Lerdahl’s work, the above conclusion may not exactly be an implementation challenge. However, it has been included here as such because, despite the fact that using one version of TPS or the other could imply critical differences between the values of distance between two chords, at the time of writing, some authors still use TPS’s preliminary version. Thus, just by acknowledging that TPS’s latest version may produce more accurate values of distance between chords, this section’s challenge is already overcome, as it is now clear that Rule 3.2 must be applied to calculate any distance within TPS.

### 4.2 | AutoTPS

**Automating Lerdahl’s Tonal Pitch Space**

This section introduces *AutoTPS* (as in **A**utomatic **T**onal **P**itch **S**pace), our publicly available system capable of automatically calculating distances within TPS.

*AutoTPS* must be fed with three inputs: \(K, x_1/y_1\) and \(x_2/y_2\). The first element, \(K\), refers to the current local key. For instance, \(C\) represents the key of C major whereas \(c\) represents the key of C minor. The last two elements, \(x_1/y_1\) and \(x_2/y_2\), refer to the original
and the destination chords, respectively, between which the distance is to be calculated. They must be input in Roman numeral notation (i.e. upper case denotes major chord degrees, X, and major keys, Y, and lower case denotes minor chord degrees, x, and minor keys, y). For instance, the diatonic fourth degree of a major and a minor key will be represented by IV/I and iv/i, respectively. Altered chord degrees and keys can also be input into AutoTPS. To do so, they must be preceded by the corresponding symbol, ♭ or ♯ or n (instead of ♮). For instance, the Neapolitan chord may be written as ♭II/I. Likewise, augmented and diminished chords should be followed by + and − symbols. For instance, the diatonic seventh degree in a major key must be written as vii-/I instead of vii/I. Finally, AutoTPS also considers diatonic seventh-chords, as the incorporation of sevenths may affect the calculation of TPS’s k. For instance, a dominant seventh chord in a major and a minor key will be represented by V7/I and V7/i, respectively.

To calculate the distance between two chords, x₁/y₁ and x₂/y₂, in a given key, K, AutoTPS follows Rule 3.2. That is to say, it calculates the values of i, j and k and adds them up together. The computational approach followed to calculate these values is discussed in Sections 4.2.1, 4.2.2 and 4.2.3, respectively.

4.2.1 | Automatically calculating i in the Tonal Pitch Space

Given two chords, x₁/y₁ and x₂/y₂, to calculate the value of i between them, AutoTPS must calculate the shortest number of steps that separate y₁ and y₂ in the chromatic circle-of-fifths.

As introduced in Section 4.1.1, the implementation of the calculation i poses a challenge because of y₁’s possible enharmonic nature. To overcome this challenge, AutoTPS incorporates two different circles-of-fifths. These are shown in Figures 4.1 and 4.2.

Figure 4.1 shows AutoTPS’s “ascending” (Schoenberg, 1983) chromatic circle-of-fifths. To build it, one must consider ascending perfect fifth-interval steps moving to the right, starting at C, and so only sharps will be considered. Figure 4.2 shows AutoTPS’s “descending” chromatic circle-of-fifths. To build it, one must consider descending perfect fifth-interval steps moving to the left, starting at C, and so only flats will be considered.
Given an original and a destination key, $y_1$ and $y_2$, to calculate the corresponding value of $i$, AutoTPS takes the following steps:

- if the roots of both $y_1$ and $y_2$ are in one of the circles-of-fifths:
  
  (1) if the circle is the “ascending” one, calculate the number of steps to the right from $y_1$ to $y_2$,
  
  (2) otherwise, if the circle is the “descending” one, calculate the number of steps to the left from $y_1$ to $y_2$.

- otherwise:
  
  (1) identify the circle which includes $y_2$,
  
  (2) in this circle, find the location of $y_1$’s enharmonic key and locate $y_1$ at this location,
  
  (3) if the circle is the “ascending” one, calculate the number of steps to the right from $y_1$ to $y_2$,
  
  (4) otherwise, if the circle is the “descending” one, calculate the number of steps to the left from $y_1$ to $y_2$.

Recall the example given in Section 4.1.1 concerning the enharmonic keys of $C^\#$ major and $D^\flat$ major. According to the condition in the above first bullet point, both
the keys of C♯ major and C major can be found in the “ascending” chromatic circle-of-fifths (i.e. Figure 4.1), where C♯ is located seven steps to the right from C. Both the keys of D♭ major and C major can be found in the “descending” chromatic circle-of-fifths (i.e. Figure 4.2), where D♭ major is located five steps to the left from C. Therefore, \( i(I/C \rightarrow I/C_2) = 7 \) and \( i(I/C \rightarrow I/D♭) = 5 \).

If considering G major instead of C major as the original key in the above example, to calculate the value of \( i \) to the key of D♭ major, one must follow the second condition in the above second bullet point, since D♭ and G cannot be found in the same circle. According to this second condition, the key of D♭ major can be found in the “descending” chromatic circle-of-fifths. The enharmonic key of G major is that of A♭♭ major, and D♭ major is located six steps to the left from A♭♭ major. Therefore, \( i(I/G \rightarrow I/D♭) = 6 \).

4.2.2 | Automatically calculating \( j \) in the Tonal Pitch Space

Given two chords, \( x_1/y_1 \) and \( x_2/y_2 \), to calculate the value of \( j \) between them, AutoTPS must calculate the shortest number of steps that separate them in the diatonic circle-of-fifths.

As introduced in Section 4.1.2, the implementation of the calculation \( j \) poses a challenge if \( x_1/y_1 \) and \( x_2/y_2 \) are not in the same region. To overcome this challenge, AutoTPS takes the following steps to calculate \( j \):

1. generating the diatonic spaces of the tonic chords of the keys between \( x_1/y_1 \) and \( x_2/y_2 \),

2. moving four steps to the right or to the left starting at the root of \( x_1/y_1 \) in the first space and reaching \( z \),

3. moving to \( z \) in the last space if it is included there,

4. otherwise, moving to \( z \) in the next space if it is included there,

5. repeating steps (3) and (4) until \( z = x_2/y_2 \) is reached, and
(6) calculating $j$ as the total number transitions, between different notes, in the previous steps.

*AutoTPS* applies the above steps twice, once taking steps to the right and another taking steps to the left. In the end, the final value of $j$ is calculated as the shortest between these two, as proposed by Rule 3.2.

Figures 4.3, 4.4 and 4.5 include three examples to illustrate the calculation of $j$ using the above steps. Notice that the destination chords in these figures, $I/\flat$, $I/C\#$ and $\flat\II/C$, respectively, concern different chord labels of the same enharmonic chord.

Figure 4.3 illustrates the calculation of $j(I/C \rightarrow I/\flat)$. Notice this figure includes the diatonic spaces of the tonic chords between the keys of $C$ major and $\flat$ major. From pitch class $c$, the root of the key of $C$ major, four steps are taken to the left so as to arrive at pitch class $f$. Since this pitch class is in the diatonic space of $I/\flat$, *AutoTPS* transitions to $f$ in the last space. By taking steps to the left several times, one arrives at pitch class $d\flat$, which is the root of the destination chord. The final value of $j$ will be equal to the number of arrows in Figure 4.3; that is to say, $j(I/C \rightarrow I/\flat) = 5$. *AutoTPS* would have also calculated $j$ taking steps to the right from pitch class $c$. The value of $j$ following this direction will be greater and so, for the sake of clarity, its calculation is not included in Figure 4.3.

Figure 4.4 illustrates the calculation of $j(I/C \rightarrow I/C\#)$. Notice this figure includes the diatonic spaces of the tonic chords between the keys of $C$ major and $C\#$ major. From pitch class $c$, the root of the key of $C$ major, four steps are taken to the right so as to arrive at pitch class $g$. Since this pitch class is not in the diatonic space of $I/C\#$, but it is in the next diatonic space, *AutoTPS* transitions to $g$ in $I/G$'s space. By taking steps to the right several times, one arrives at pitch class $c\#$, which is the root of the destination chord. The final value of $j$ will be equal to the number of arrows in Figure 4.4; that is to say, $j(I/C \rightarrow I/C\#) = 7$. *AutoTPS* would have also calculated $j$ taking steps to the left from pitch class $c$. The value of $j$ following this direction will be greater and so, for the sake of clarity, its calculation is not included in Figure 4.4.

Figure 4.5 illustrates the calculation of $j(I/C \rightarrow \flat\II/C)$. Notice this figure includes the
diatonic spaces of both chords. From pitch class c, the root of the key of C major, four steps are taken to the right so as to arrive at pitch class g. Since this pitch class is in the diatonic space of ♭II/C, AutoTPS transitions to g in the last space. By taking another four steps to the right, one arrives at pitch class d♭, which is the root of the destination chord.

The final value of j will be equal to the number of arrows in Figure 4.5; that is to say, $j(I/C\rightarrow ♭II/C) = 2$. AutoTPS would have also calculated j taking steps to the left from pitch class c. The value of j following this direction will be greater and so, for the sake of clarity, its calculation is not included in Figure 4.5.

How does AutoTPS know which keys to consider when calculating j? For instance, in Figure 4.4, only the keys to the right from C major, in the chromatic circle-of-fifths (i.e. G, D, A, E, B and F♯), are considered to reach the key of C♯ major; but, why not those to the left of C major in the chromatic circle-of-fifths (i.e. F, B♭, E♭ and A♭)?

The keys that must be included between the original and the destination keys come defined by the previously calculated i. If AutoTPS calculated i using the “ascending” diatonic circle-of-fifths, the keys to be considered to calculate j are those to the right from the original key, in the chromatic circle-of-fifths, as shown in Figure 4.4. Otherwise, if AutoTPS calculated i using the “descending” diatonic circle-of-fifths, the keys to be considered to calculate j are those to the left from the original key, in the chromatic circle-of-fifths, as shown in Figure 4.3.

### 4.2.3 | Automatically calculating k in the Tonal Pitch Space

Given two chords, $x_1/y_1$ and $x_2/y_2$, to calculate the value of k between them, AutoTPS generates the basic spaces of both chords. It then calculates the difference of all twelve weightings in the basic spaces and adds together the difference of those that are greater in the case of $x_2/y_2$.

For instance, let us consider again the keys of C♯ major and D♭ major. The basic space oriented toward the tonic chords of these keys, which note is the same, is shown in Figure 4.6.

The basic space oriented toward I/C was already shown in Figure 3.12. It is shown
Chapter 4. Automatically Calculating Tonal Tension

Figure 4.3: An example of the calculation of $j(I/C \rightarrow I/D_b)$ using AutoTPS.

Figure 4.4: An example of the calculation of $j(I/C \rightarrow I/C♯)$ using AutoTPS.

Figure 4.5: An example of the calculation of $j(I/C \rightarrow bII/C)$ using AutoTPS.
Chapter 4. Automatically Calculating Tonal Tension

Figure 4.6: TPS's basic space (compact version) oriented toward the tonic chord of C♯ major or D♭ major.

Figure 4.7: TPS's basic space (compact version) oriented toward I/C (repeated figure).

again in Figure 4.7.

Therefore, to calculate the value of \( k \) when transitioning from I/C to either I/C♯ or I/D♭, AutoTPS will do:

\[
k = (2 - 5) + (5 - 1) + (1 - 2) + (2 - 1) + (1 - 3) + (3 - 2) + (2 - 1) + (1 - 4) + (4 - 1) + (1 - 2) + (2 - 1) + (1 - 2) = 4 + 1 + 1 + 1 + 3 + 1 = 11
\]

Notice that those pairs where the subtraction was either negative or zero have been struck through.

4.3 | AuToTen

Automating Lerdahl's Model of Tonal Tension

This section introduces AuToTen (as in Automatic Tonal Tension), our publicly available\(^5\) system capable of automatically calculating tonal tension according to Lerdahl’s MTT.

Given a piece of music, AuToTen must be fed with four inputs: the score of the piece in MusicXML format, a list of the piece’s key and chord labels (in the notation format required by TPS) and the piece’s prolongational reduction.

To calculate the degrees of MTT’s hierarchical tension, \( T_{\text{hier}} \), and attraction, \( \alpha \), throughout a given piece, AuToTen follows Rules 3.5 and 3.8, respectively. The computational approach followed to calculate these values is discussed in Sections 4.3.1, 4.3.2 and 4.3.3.

\(^5\)https://doi.org/10.21964/ou.rd.13026578.v1
4.3.1 | A representation of the prolongational reduction

The last input AuToTen is fed with is a piece’s *prolongational reduction*. This GTTM component will be used to calculate the corresponding degrees of inherited tension, \( T_{inh} \), that will contribute to the final values of hierarchical tension, \( T_{hier} \). However, as discussed in Section 4.1.3, representing the *prolongational reduction* in such a way that it is readable and accessible by computational means may pose a challenge. To overcome this challenge, AuToTen incorporates a novel method to generate a matrix that incorporates the hierarchical relations embedded in GTTM’s *prolongational reduction*. This matrix is called the *prolongational matrix*.

Given a piece of music whose *prolongational reduction* includes \( l \) events, AuToTen generates an empty matrix consisting of \( l \) rows and \( l \) columns. To fill the matrix, in order to generate the piece’s prolongational matrix, AuToTen first identifies the highest event in the piece’s *prolongational reduction* (i.e. the event that dominates them all and so has no branches above it in the tree representation). From the events in the piece (i.e. 1, 2, ..., \( l \)), let \( m \) be the event identified as the highest one. In the prolongational matrix, the element where the \( m \)th row and the \( m \)th column intersect is assigned the value 0. Next, AuToTen identifies the events in the piece that directly attach to event \( m \). For each of these events, \( n \), the element where the \( n \)th row and the \( m \)th column intersect is assigned the value 1. AuToTen then identifies the events in the piece that directly attach to events \( n \). For each of these events, \( o \), the element where the \( o \)th row and the \( n \)th column intersect is assigned the value 2. And so on.

Notice that in the prolongational matrix generated by AuToTen, each row contains a single value whereas the columns may contain more than one value or even no value at all. In this way, one can easily interpret the information in the prolongational matrix: an event \( x \), in a given piece of music, attaches to event \( y \) in the prolongational reduction if the intersection between the \( x \)th row and the \( y \)th column in the matrix is non-empty (i.e. it contains an integer).

To illustrate the generation of a *prolongational matrix*, Figure 4.8a shows the prolongational matrix associated with the *prolongational reduction* of the first phrase of the
theme in Mozart’s *Ah vous dirai-je, Maman*, K. 265/300e, which was previously shown in Figure 3.22. The theme is shown again in Figure 4.8b. For the sake of clarity, the rows and the column in Figure 4.8a are numbered according to the numbers assigned to the events in Figure 4.8b.

![Prolongational matrix](image)

**(a) Prolongational matrix**

![Prolongational reduction](image)

**(b) Prolongational reduction**

Figure 4.8: Prolongational matrix, (a), associated with the *prolongational reduction* of the theme in Mozart's *Ah vous dirai-je, Maman*, K. 265/300e, (b).
4.3.2 | Automatic calculation of tonal tension

In order to automatically calculate the degrees of hierarchical tension, $T_{\text{hier}}$, in a given piece of music, according to Lerdahl’s MTT, AuToTen takes the following steps:

1. calculating the degree of surface tension, $T_{\text{diss}}$, according to Rule 3.4, for each event in the piece,

2. calculating the distance, $\delta$, according to Rule 3.2, from each of the identified events to the one that immediately dominates it in the piece’s prolongational matrix,

3. calculating the inherited contribution to tension, $T_{\text{inh}}$, according to Rule 3.6, for each of the identified events by calculating all distances, $\delta$, of the connected events in the piece’s prolongational matrix until the matrix’s value 0 is reached, and

4. calculating the final degree of hierarchical tension, $T_{\text{hier}}$, according to Rule 3.5, associated with each of the identified events by adding together the corresponding results of the previous three steps.

In order to automatically calculate the degrees of attraction, $\alpha$, in a given piece of music, according to Lerdahl’s MTT, AuToTen takes the following steps:

6. calculating the anchoring space, as in Figure 3.23, associated with the identified events,

7. extracting all the voices associated with each identified event in the piece’s MusicXML input score,

8. calculating the degree of melodic attraction, $\alpha_{\text{mel}}$, according to Rule 3.7, associated with the extracted voices for each identified event, and

9. calculating the final degree of attraction, $\alpha$, according to Rule 3.8, associated with each of the identified events by operating upon the sum of the values calculated in the previous step.
4.3.3 | Automatic feature extraction

The implementation of AuToTen described in Section 4.3.2 is limited by the inputs, which must be manually calculated. Despite this fact, AuToTen will allow us to implement a music generator to answer our Research Question. In order to automate the calculation of the inputs, AuToTen incorporates some additional algorithms. These algorithms have been incorporated into AuToTen’s implementation so that users can easily extract tonal tension from a piece of music without needing to manually calculate any inputs except for the piece’s prolongational reduction, which, as mentioned in Section 4.1.3, could be calculated using the Interactive GTTM Analyser (IGA). Therefore, the algorithms provide the community with a better tool that may lead to more impactful new research.

The additional algorithms incorporated in AuToTen are applied to whole pieces of music to automatically extract the data that AuToTen needs to calculate tension and attraction. In this way, the additional algorithms improve AuToTen’s analytical abilities. These algorithms, however, will not be further used in this dissertation. That is because the music generation system introduced in Chapter 6 does not need to analyse whole pieces of music but only small chord transitions, which can be handled by AuToTen without needing the additional algorithms. Therefore, this section only summarises the additional algorithms. For more detail concerning their implementation, see AuToTen’s public repository.

AuToTen’s additional algorithms include:

- a method to automatically generate AuToTen’s input list of offsets of the notes in piece of music from the piece’s metrical reduction in MusicXML format (as calculated by IGA),

- a method to automatically estimate\(^6\) AuToTen’s input list of key and chord labels in a given piece of music,

---

\(^6\)Given a piece of music, this additional method estimates the piece's key using music21’s (Cuthbert and Ariza, 2010) analyze function. This function is also used to estimate the most suitable key for each measure. To estimate different versions of the piece’s chord labels, music21’s chordify function is applied twice: once under the piece’s global key and another time under each measure’s estimated key. These two applications will allow us to better identify non-diatonic chords, such as secondary dominants.
a method to automatically generate AuToTen’s input prolongational matrix from the prolongational reduction of a give piece in MusicXML format (as calculated by IGA), and

a method to automatically estimate the heights\(^7\) of the branches in the prolongational reduction when only given the corresponding prolongational matrix, so that the matrix can easily be transformed into a tree representation.

4.4 | Computational evaluation

To test the accuracy of the data calculated by AutoTPS and AuToTen, compared to what one would get when manually applying Lerdahl’s TPS and MTT, respectively, we carried out a total of one hundred test cases. Half of these test cases concern the evaluation of AutoTPS and the other half concern the evaluation of AuToTen. These are discussed in Sections 4.4.1 and 4.4.2, respectively.

4.4.1 | Evaluating AutoTPS’s accuracy

The fifty test cases used to evaluate AutoTPS’s accuracy include thirty typical close-chord transitions annotated from Lerdahl (2004), five distant-chord transitions provided by F. Lerdahl (personal communication) and fifteen atypical chord transitions that we calculated manually.

The ground-truths of the test cases are shown in Table 4.1. For each test case, Table 4.1 includes an original chord from which to transition and a destination chord, both relative to the key of C major, as well as the values of \(i\), \(j\), \(k\) and \(\delta\), calculated according to Rule 3.2, and the corresponding reference pointing out where the ground-truth was transcribed from.

The values of \(i\), \(j\), \(k\) and \(\delta\) automatically calculated by AutoTPS for all the fifty chord transitions in Table 4.1 agreed with the annotated ground-truths.

\(^7\)This approach is based on the calculation of GTTM-related salience, as proposed by Marsden et al. (2018).
4.4.2 | Evaluating AuToTen’s accuracy

The fifty test cases used to evaluate AuToTen’s accuracy include the nine events in the Grail theme from Wagner’s Parsifal and the forty-one events in Bach’s chorale Christus, der ist mein Leben. These pieces of music, their numbered events and chord labels are shown in Figures 4.9 and 4.10, respectively.

Figure 4.9: Grail theme from Wagner’s Parsifal, and its chord labels, transcribed from Lerdahl and Krumhansl (2007, p.343).

Figure 4.10: Bach’s chorale Christus, der ist mein Leben, and its chord labels, transcribed from Lerdahl and Krumhansl (2007, pp.344-345).

The ground-truths of the test cases concerning Wagner’s and Bach’s pieces are shown in Tables 4.2 and 4.3, respectively. For each test case, these tables include the corresponding event number and its degrees of surface tension, $T_{diss}$, hierarchical tension, $T_{hier}$, and
attraction, $\alpha$. These values have been transcribed from Lerdahl and Krumhansl (2007) where they were calculated according to Rules 3.4, 3.5 and 3.8, respectively.

All nine values of $T_{diss}$, $T_{hier}$ and $\alpha$ calculated by AuToTen for the transitions in Figure 4.9 agreed with the ground-truths in Table 4.2. However, only thirty-eight out of the forty-one calculations made by AuToTen for the transitions in Figure 4.10 agreed with the ground-truths in Table 4.3. That is to say, there exist three discrepancies between AuToTen’s calculations and the ground-truth data. The discrepancies concern events 14, 35 and 36 in Table 4.3. Interestingly, these events share the same chord label, $vii^\circ$. According to Lerdahl (2004, pp.77-78), TPS presents some “issues of quantification”. One of these issues deals with the difference between a chord’s label and the harmonic function it may play. For instance, depending on the context, a chord labelled as $vii^\circ$ could act as a dominant function and so it is better treated as $V^7$ with a missing root. Notice this is the case of the three events mentioned above, events 14, 35 and 36. What is more, the sevenths in events 35 and 36 act as passing notes, so they should not be considered within the triadic level when calculating the chord’s basic space. That is to say, in order to get the same results as in Table 4.3, when feeding AuToTen with the chord labels in Bach’s chorale, event 14 should be transcribed as $V^7$, instead of $vii^\circ$, whereas events 35 and 36 should be transcribed as $V$.

Once the above “issues of quantifications” were dealt with, all forty-one of $T_{diss}$, $T_{hier}$ and $\alpha$ calculated by AuToTen for the transitions in Figure 4.10 agreed with the ground-truths in Table 4.3.

### 4.5 Summary and Conclusions

In order to seek answers to our Research Question, Chapter 4 has fully automated Lerdahl’s Model of Tonal Tension (MTT).

Section 4.1 has identified the main issues that posed a challenge in previous attempts of the automation of MTT. These challenges concern the automatic calculation of the number of steps within enharmonic keys in the chromatic circle-of-fifths, the automatic calculation of the number of steps within chords from different regions in the diatonic
circle-of-fifths, the representation of the prolongational reduction in a more readable and accessible way and the calculation of distances within the Tonal Pitch Space (TPS) across regions.

Section 4.2 has introduced AutoTPS, a system capable of automatically calculating distances between chords within Lerdahl’s TPS. AutoTPS implements two novel components to overcome the first two implementation challenges introduced in Section 4.1. The first component implements a method that calculates the number of steps in the chromatic circle-of-fifths by considering two circles, an ascending one, which only includes keys with sharps, and a descending one, which only includes keys with flats. The second component implements a method that calculates the number of steps in the diatonic circle-of-fifths by a drawing a path through neighbour regions and their respective diatonic circles.

Section 4.3 has introduced AuToTen, a system capable of automatically calculating tension according to Lerdahl’s Model of Tonal Tension (MTT). AuToTen implements two novel components to overcome the last two implementation challenges introduced in Section 4.1. The first component automatically generates a matrix that represents the hierarchical relations in a piece of music in a more readable and accessible manner. The second component implements the calculation of tonal tension according to MTT’s latest version.

Finally, Section 4.4 has reviewed the computational evaluation of AutoTPS and AuToTen. The accuracy of these systems was evaluated by comparing their calculations against a collection of one hundred test cases, fifty per system. All values automatically calculated by AutoTPS and AuToTen agreed with the annotated ground-truths.
Table 4.1: Test cases used to evaluate the accuracy of AutoTPS.

<table>
<thead>
<tr>
<th>Original chord</th>
<th>Destination chord</th>
<th>$i$</th>
<th>$j$</th>
<th>$k$</th>
<th>$s$</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>I/I</td>
<td>IV/I</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>Lerdahl (2004, p.56)</td>
</tr>
<tr>
<td>I/I</td>
<td>vii/I</td>
<td>0</td>
<td>2</td>
<td>6</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>I/I</td>
<td>iii/I</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>vi/I</td>
<td>vi/I</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>vi/I</td>
<td>IV/I</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>I/I</td>
<td>V/I</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>I/I</td>
<td>V/ii</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>7</td>
<td>Lerdahl (2004, p.61)</td>
</tr>
<tr>
<td>I/I</td>
<td>I/VI</td>
<td>3</td>
<td>5</td>
<td>8</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>I/I</td>
<td>i/i</td>
<td>1</td>
<td>2</td>
<td>7</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>I/I</td>
<td>i/iii</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>I/I</td>
<td>i/IV</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>I/I</td>
<td>vi/v</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>I/I</td>
<td>v/v</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>I/I</td>
<td>vii/v</td>
<td>0</td>
<td>2</td>
<td>7</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>I/I</td>
<td>III/I</td>
<td>0</td>
<td>2</td>
<td>8</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>I/I</td>
<td>V/V</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>V/V</td>
<td>I/V</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>i/i</td>
<td>V/I</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>7</td>
<td>Lerdahl (2004, p.145)</td>
</tr>
<tr>
<td>I/I</td>
<td>I/V</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>V/I</td>
<td>I/V</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>Lerdahl (2004, p.148)</td>
</tr>
<tr>
<td>V/I</td>
<td>V/V</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>Lerdahl (2004, p.156)</td>
</tr>
<tr>
<td>V/V</td>
<td>I/II</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>I/I</td>
<td>i/IV</td>
<td>0</td>
<td>6</td>
<td>11</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>I/I</td>
<td>II/I</td>
<td>1</td>
<td>7</td>
<td>11</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>I/I</td>
<td>I/III</td>
<td>5</td>
<td>6</td>
<td>11</td>
<td>21</td>
<td>F. Lerdahl (personal communication, September, 2019)</td>
</tr>
<tr>
<td>I/I</td>
<td>V'/ii</td>
<td>1</td>
<td>3</td>
<td>7</td>
<td>11</td>
<td>F. Lerdahl (personal communication, April, 2020)</td>
</tr>
<tr>
<td>I/I</td>
<td>I/VI</td>
<td>4</td>
<td>3</td>
<td>9</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>vii/iv</td>
<td>i/IV</td>
<td>3</td>
<td>2</td>
<td>9</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>i/iv</td>
<td>I/i</td>
<td>1</td>
<td>3</td>
<td>7</td>
<td>11</td>
<td>manual calculation</td>
</tr>
<tr>
<td>VII/V</td>
<td>iii/iI</td>
<td>2</td>
<td>2</td>
<td>8</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>VII/vI</td>
<td>vii/IV</td>
<td>6</td>
<td>6</td>
<td>11</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>vii/I</td>
<td>III/vI</td>
<td>1</td>
<td>2</td>
<td>9</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>i/IV</td>
<td>i/III</td>
<td>1</td>
<td>2</td>
<td>9</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>i/IV</td>
<td>II/IV</td>
<td>1</td>
<td>3</td>
<td>9</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>VII/vI</td>
<td>VI/I</td>
<td>4</td>
<td>6</td>
<td>7</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>III/v</td>
<td>vii/VII</td>
<td>2</td>
<td>5</td>
<td>8</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>i/ii</td>
<td>III/iI</td>
<td>1</td>
<td>3</td>
<td>7</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>I/i</td>
<td>ii/i</td>
<td>4</td>
<td>1</td>
<td>5</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>VI/iI</td>
<td>iii/vI</td>
<td>5</td>
<td>3</td>
<td>8</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>VII/vII</td>
<td>I/I</td>
<td>5</td>
<td>0</td>
<td>3</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>i/vI</td>
<td>V/V</td>
<td>1</td>
<td>1</td>
<td>8</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>V/IV</td>
<td>V/VII</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>8</td>
<td></td>
</tr>
</tbody>
</table>
Table 4.2: Test cases used to evaluate the accuracy of AuToTen, which include the degrees of surface tension, hierarchical tension and attraction in the Grail theme from Wagner’s Parsifal transcribed from Lerdahl and Krumhansl (2007, p.343).

<table>
<thead>
<tr>
<th>Event</th>
<th>$T_{diss}$</th>
<th>$T_{hier}$</th>
<th>$\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>.33</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>7</td>
<td>1.44</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>15</td>
<td>null</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>15</td>
<td>.34</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>21</td>
<td>null</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>27</td>
<td>null</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>22</td>
<td>1.26</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>8</td>
<td>8.0</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>1</td>
<td>0.</td>
</tr>
</tbody>
</table>

Table 4.3: Test cases used to evaluate the accuracy of AuToTen, which include the degrees of surface tension, hierarchical tension and attraction in Bach’s chorale Christus, der ist mein Leben transcribed from Lerdahl and Krumhansl (2007, pp.344-345).

<table>
<thead>
<tr>
<th>Event</th>
<th>$T_{diss}$</th>
<th>$T_{hier}$</th>
<th>$\alpha$</th>
<th>Event</th>
<th>$T_{diss}$</th>
<th>$T_{hier}$</th>
<th>$\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.0</td>
<td>21</td>
<td>6</td>
<td>19</td>
<td>5.84</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1.6</td>
<td>22</td>
<td>3</td>
<td>17</td>
<td>5.5</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>8</td>
<td>2.25</td>
<td>23</td>
<td>1</td>
<td>6</td>
<td>3.49</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>15</td>
<td>7.75</td>
<td>24</td>
<td>1</td>
<td>6</td>
<td>null</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>15</td>
<td>1.59</td>
<td>25</td>
<td>1</td>
<td>7</td>
<td>4.59</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>5</td>
<td>4.4</td>
<td>26</td>
<td>4</td>
<td>17</td>
<td>1.</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>11</td>
<td>null</td>
<td>27</td>
<td>3</td>
<td>16</td>
<td>1.85</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>7</td>
<td>null</td>
<td>28</td>
<td>1</td>
<td>19</td>
<td>.68</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>5</td>
<td>null</td>
<td>29</td>
<td>4</td>
<td>22</td>
<td>1.43</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>1</td>
<td>1.5</td>
<td>30</td>
<td>2</td>
<td>20</td>
<td>1.87</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>6</td>
<td>4.5</td>
<td>31</td>
<td>0</td>
<td>18</td>
<td>null</td>
</tr>
<tr>
<td>12</td>
<td>3</td>
<td>3</td>
<td>.64</td>
<td>32</td>
<td>1</td>
<td>13</td>
<td>1.18</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>9</td>
<td>1.12</td>
<td>33</td>
<td>0</td>
<td>7</td>
<td>1.46</td>
</tr>
<tr>
<td>14</td>
<td>3</td>
<td>9</td>
<td>8.75</td>
<td>34</td>
<td>1</td>
<td>1</td>
<td>1.93</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>1</td>
<td>0.38</td>
<td>35</td>
<td>6</td>
<td>11</td>
<td>5.5</td>
</tr>
<tr>
<td>16</td>
<td>3</td>
<td>3</td>
<td>null</td>
<td>36</td>
<td>3</td>
<td>8</td>
<td>8.5</td>
</tr>
<tr>
<td>17</td>
<td>4</td>
<td>9</td>
<td>5.82</td>
<td>37</td>
<td>3</td>
<td>3</td>
<td>1.89</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>6</td>
<td>4.82</td>
<td>38</td>
<td>2</td>
<td>14</td>
<td>4.75</td>
</tr>
<tr>
<td>19</td>
<td>1</td>
<td>1</td>
<td>.22</td>
<td>39</td>
<td>1</td>
<td>6</td>
<td>5.32</td>
</tr>
<tr>
<td>20</td>
<td>3</td>
<td>3</td>
<td>.24</td>
<td>40</td>
<td>2</td>
<td>7</td>
<td>8.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>41</td>
<td>0</td>
<td>0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

111
In this chapter, we identify the computational generation methods best suited to our goals. To do so, we will:

1. define and frame the scope of the concept of Automatic Music Generation,

2. design a taxonomy of the computational methods in the field of Automatic Music Generation,

3. identify the computational methods in the field of Automatic Music Generation best suited to our goals, and

4. determine how to apply the identified computational methods to address our Research Question.

The first and second steps are addressed in Section 5.1, the third step is addressed in Sections 5.2 and the fourth step is addressed in Section 5.3.
5.1 Framing the scope of the field of Automatic Music Generation

In music composition, one can define specific rules and frameworks to partially or totally automate the compositional process. This method of composition is often referred to as *Algorithmic Composition* (Ariza, 2005; Essl, 2012; Fernández and Vico, 2013; Nierhaus, 2009).

The history of *Algorithmic Composition* goes back long before the modern era. Already in ancient Greece, Pythagoras, Ptolemy and Plato theorised about utilising formal instructions to create music (Burkholder et al., 1996). In the eleventh century, G. D’Arezzo developed a system to map the vowels of a Latin text to the notes in a scale to generate a melody (Nierhaus, 2009). In the seventeenth century, J. S. Bach composed a piece of music, his *14 Canons*, BWV 1087, which incorporated a musical riddle. The player was supposed to manipulate the piece’s subjects to generate new musical material (Essl, 2012). In the eighteenth century, musical dice games became very popular. Using these, one could generate random pieces of music from pre-composed materials which were stochastically ordered by the score of some dice (Hedges, 1978). In the twentieth century, two events fostered the growth of the study and development of *Algorithmic Composition*. First, the publication of Schillinger’s (1949) system of musical composition, the first of its kind to fully describe the application of mathematical logic to music composition (Essl, 2012). Second, the publication of the *Illiac Suite*, which brought computers into the picture. As a direct consequence, a vast amount of new algorithmic techniques were later explored during the avant-garde period.

5.1.1 Defining the concept of Automatic Music Generation

Over the past sixty years, the development of computers has fostered the development, study and application of new *Algorithmic Composition* techniques. What is more, computers have allowed the automation of processes that directly implement that of *Algorithmic Composition*. These include the extraction, processing and analysis of musical data, the
automation of algorithmic calculations and decision-making, the capability of playing back algorithmically generated music, the interaction with other environments while generating music (e.g. video games), the automatic arrangement of musical elements (e.g. voices or patterns) and many more.

Many different terms have been used in the literature to refer to the application of Algorithmic Composition techniques using computers. These include Computer Music, Procedural Composition or Computer-Aided Composition, among others. What is more, some scholars, such as Fernández and Vico (2013) or Simoni (2003), have decided to take for granted that, in the twenty-first century, the term Algorithmic Composition implies the use of computers.

The term Computer-Aided Composition became very popular in the music computing community in the last two decades. However, Ariza (2005, p.1) argues that this term “lacks the specificity of using generative algorithms”. In fact, the process of music composition using Digital Audio Workstations (e.g. Ableton Live, Cubase, Logic Pro, Pro Tools, etc.) is a form of computer-aided composition, which Reuter (2021) argues is currently one of the most common compositional approaches. However, that is not what Computer-Aided Composition is intended to mean. Ariza (2005) alternatively proposes the term Computer-Aided Algorithmic Composition. This term, however, has not been widely accepted by the community. Instead, the terms Automatic Music Generation and Automatic Music Composition have gained more popularity. Although they are not as explicit as Computer-Aided Algorithmic Composition, they both imply the use of computers by incorporating the word “automatic”. Likewise, the former, by incorporating the word “generation”, also implies the use of generation techniques. Hence, in this dissertation, Automatic Music Generation is defined as follows:

**Defining Automatic Music Generation**

*Automatic Music Generation* is the application of Algorithmic Composition generation techniques by computational means.
5.1.2 | A taxonomy of the main methods in the field

The study of the field of *Automatic Music Generation* can be approached from four different points of view:

- from the *artistic* perspective, one could focus on the musical genres the generated music falls into (e.g. serialism, stochastic music, aleatoric composition, etc.),

- from the *functional* perspective, one could focus on the purposes or goals the algorithm is aimed at (e.g. generating a full composition, an arrangement or just a melody, writing music to match a given narrative, composing with regards to timbre, tweaking a piece of music according to different degrees of complexity, etc.),

- from the *dimensional* perspective, one could focus on the musical elements considered in the generation (e.g. a piece’s granularity,\(^1\) its groove,\(^2\) its directionality,\(^3\) etc.), and

- from the perspective of the *methods*, one could focus on the algorithmic technique used in the generation (e.g. musical grammars, neural networks, Markov models, etc.).

Recall that this chapter started by presenting a sequence of steps to address our goals. These steps concern the *methods* in the field of *Automatic Music Generation*. Thereby, this chapter will exclusively approach the field of *Automatic Music Generation* from the point of view of the *methods*, that is the last one introduced above. For detailed reviews of the field of *Automatic Music Composition*, see: Doornbusch (2002); Essl (2012), concerning the *artistic point of view*; see Herremans et al. (2017); Liu and Ting (2016), concerning the *functional point of view*; and see Plut and Pasquier (2020), concerning the *dimensional point of view*.

The most common *methods* used in the field of *Automatic Music Generation* are shown in Figure 5.1. This figure presents a taxonomy which we have designed inspired by those

---
\(^{1}\) i.e. how the piece’s individual elements are combined to create complex groups of *features*.

\(^{2}\) i.e. how regular is the distribution of the piece’s musical events in time.

\(^{3}\) i.e. whether the generation manipulates musical elements in a vertical or horizontal way.
in Briot et al. (2020); Carnovalini and Rodà (2020); Fernández and Vico (2013); Lopez-Rincon et al. (2018); Nierhaus (2009); Tatar and Pasquier (2019); Whorley and Conklin (2016b).

5.2 Which computational methods are best suited to our goals?

From the methods shown in Figure 5.1, which ones may best suit the generation of music considering Lerdahl’s Model of Tonal Tension (MTT)?

Let us begin by discussing whether our Research Question could be best answered by a non-AI or an AI method. The former are reviewed in Section 5.2.1 and the latter in Sections 5.2.2 and 5.2.3.

5.2.1 Non-AI generation methods

Within non-AI generation methods, Figure 5.1 includes self-similarity and cellular automata methods.

Self-similarity methods entail the generation of music following patterns which are similar to others patterns set as a benchmark. Their study began as a consequence of the findings in Voss and Clarke (1978). They reported that the spectral density of musical audio signals was, in general, approximately inversely proportional to their frequencies (i.e. a \( \frac{1}{f} \) distribution). Interestingly, Voss and Clarke also found that human listeners tended to find random music more pleasant when it was composed following the patterns of \( \frac{1}{f} \) noise. Since then, other self-similar musical patterns have been used for the generation of music, such as fractals. For more detail, see Díaz-Jerez (2000).

Cellular automata methods are used to model systems whose features change over time. The temporal development of such systems is represented by a space filled with discrete cells. The state of the cells change according to specific transition rules, which are applied simultaneously to all cells. It is these transition rules what defines the behaviour of a cellular automaton (Miranda, 2007). The definition of the transition rules
Figure 5.1: A taxonomy of the main methods in the field of Automatic Music Generation.
will determine the states to which the cells could evolve. For a detailed review of cellular automata systems that can evolve to uniform, steady or chaotic states, see Bilotta et al. (2000).

Note that, in general, the inputs given to both self-similarity and cellular automata methods just concern mathematical knowledge (i.e. pattern-related) and not explicit musical-related knowledge. As a consequence, self-similarity methods often lack possibilities for intervention and structuring (Nierhaus, 2009) and cellular automata methods tend to be better suited to sound synthesis rather than Algorithmic Composition (Miranda, 2007). Thus, taxonomies such as that of Figure 5.1 consider that these methods do not actually concern an “intelligent” process and so are labelled as non-AI based.

Considering what has been discussed above, it is not a surprise that non-AI generation methods serve better as inspiration to composers, or as raw material, rather than being used as independent generation systems (Bidlack, 1992; Fernández and Vico, 2013). Because of this, non-AI generation methods do not seem to be an appropriate method to address our goals. For detailed reviews concerning self-similarity and cellular automata methods, see Pareyon (2011) and Miranda (2007), respectively.

5.2.2 Soft computing methods

Within AI generation methods, Figure 5.1 includes soft-computing and symbolic AI methods. Soft-computing methods generate music through models which, unlike others, are tolerant to imprecision and uncertainty. In Figure 5.1, soft-computing methods include deep learning, evolutionary and statistical methods. These are reviewed below in Sections 5.2.2.1, 5.2.2.2 and 5.2.2.3, respectively.

5.2.2.1 Deep learning methods

Deep learning methods (i.e. deep artificial neural networks (Briot and Pachet, 2020)), when applied to music, automatically extract music-related information from a music corpus. This information will consist of multiple levels of abstraction which are processed by multiple layers, thus taking a form similar to that of a biological neural network.
What all deep learning methods have in common is their representation of knowledge through neural networks. However, they differ in four key aspects:

(i) how one evaluates their performance,

(ii) how one optimises their search strategy,

(iii) how they acquire knowledge, and

(iv) how they are implemented.

Concerning (i), the methods are usually evaluated using a cost function which calculates the deviation between predicted and expected values. The cost function is often specific to the domain. The type of cost function to use will depend on the type of the expected output or on the type of cost (i.e. loss) in the problem at hand (e.g. mean squared error, cross-entropy, etc.), among other things. For more detail, see Briot et al. (2020).

Concerning (ii), the search of the musical parameters that best suit the method depends on an optimisation algorithm, or a collection of algorithms. For more detail, see Goodfellow et al. (2016, ch. 8).

Concerning (iii), there are different types of learning strategies depending on the nature of the experience conveyed by the input data. A deep learning system can be trained on data which include an expected answer for a given problem (supervised learning) and it will typically predict new answers taking into account the expected input. However, the system can also be trained on data which do not include an expected answer (unsupervised learning), so the system will extract patterns from the input data to be able to guess what the output should be.

Concerning (iv), there are different types of architectures that could be used to implement a deep learning system. Briot et al. (2020, pp.52-53) distinguish between four main types of deep learning architectures: feedforward, autoencoder, restricted Boltzmann machine and recurrent network.

Deep learning methods usually require large amount of data. According to Herremans et al. (2017, p.24), “the field [of Automatic Music Generation] has seen an ongoing need
for more data”. And, what is more, there exists a critical lack of data concerning musical tension. Research projects based on musical tension either collect a short number of human judgements (e.g. Bigand and Parncutt (1999); Bigand et al. (1996); Farbood (2011); Krumhansl (1996); Lerdahl and Krumhansl (2007); Lopes et al. (2016)) or calculate a short number of tension profiles according to a theoretical model of tension (e.g. Herremans and Chew (2017); Mizutani and Iwami (2017); Mizutani et al. (2018); Mizutani and Nakata (2020); Verstraelen (2019); Yoo and Lee (2006)). In fact, in Halac (2019), which presents an extensive review of existing databases in music composition, there are none that explicitly focus on tension. Because of this, deep learning methods are not currently a good approach to address our goals, but could be in the future if better corpora could be built. For a detailed review of deep learning methods in the field of Automatic Music Generation, see Briot et al. (2020).

5.2.2.2 | Evolutionary methods

Evolutionary methods entail the improvement of a set of candidate solutions to a given problem (i.e. population-based approach) until certain criteria are met. These methods usually generate an original set of musical cues that might fulfil the needs of a given task, such as, for example, melodic improvisation or harmonisation. These cues are evaluated using a fitness function which examines how close each cue is from being considered an optimal solution to the given task. In an iterative process, the set of musical cues are narrowed down until the most appropriate cue is generated. To do so, cues are copied a number of times proportional to how well they fit an optimal solution. This decreases the number of possible output cues. The process is repeated until the collection of best fitted cues is found.

It may seem that evolutionary methods could be the appropriate candidates to seek answers to our Research Question, since GTTM’s rules and MTT’s equations seem to be suitable for the implementation of an automatic fitness function. However, note that according to MTT, the total degree of tension that might be conveyed by a piece of music involves a multiple regression over two features, hierarchical tension, $T_{\text{hier}}$, and
total attraction, \( \alpha \); that is to say, MTT does not provide a method to calculate a final tension profile given a piece of music. Therefore, a fitness function which is implemented considering MTT’s features independently might fail to produce, in the end, the desired degree of tension. Thus, *evolutionary* methods do not seem to be an appropriate method to address our goals, considering Lerdahl’s MTT. For a detailed review of *evolutionary* methods, see Loughran and O’Neill (2020).

5.2.2.3 | *Statistical* methods

*Statistical* methods entail the generation of musical events according to probabilities assigned to a specific context. They usually involve three steps (Whorley and Conklin, 2016b): assigning probability distributions to the musical events, sampling from the probability distributions to generate new musical fragments and optimising the algorithms that establish what the most probable sequences are.

*Statistical* methods present some advantages and some disadvantages with regards to our goals. On one hand, they may be the easiest and most appropriate approach when considering the generation of music in real time. On the other hand, they often lack the capacity to represent and process long-term musical structures (Briot and Pachet, 2020; Lerdahl, 2015; Papadopoulos and Wiggins, 1999; Whorley and Conklin, 2016a). To overcome the latter, scholars often use the following strategies: using pattern discovery algorithms to design a template structure upon which to base the generation (Collins and Laney, 2017; Collins et al., 2016; Padilla and Conklin, 2018; Whorley and Conklin, 2016a) and/or searching for low cross-entropy\(^4\) solutions (Herremans et al., 2014, 2015; Padilla and Conklin, 2018; Whorley and Conklin, 2016b; Whorley et al., 2007).

The cross-entropy strategy is often used together with a multiple viewpoints representation of the input corpus. The multiple viewpoints system transforms music into a more abstract representation (see Conklin and Witten (1995) for more detail). According to Hall and Pearce (2021, p.223):

----

\(^4\)Cross-entropy is a theoretic measure that allows us to ensure fair comparisons between different generated solutions. Given a sequence of musical events \( e_1, \ldots, e_n \), cross-entropy, \( H \), can be approximated by \( H = -\frac{1}{n} \sum_{i=1}^{n} \log_2 P(e_i|c) \), where \( n \) is the number of events in the sequence and \( P(e_i|c) \) is the probability of event \( e_i \), given its context \( c \), in the current statistical model (Whorley and Conklin, 2016b).
viewpoints allow the representation and prediction of pitch structure in music, not only by absolute pitch, but also by interval, scale degree or contour, and the prediction of temporal structures.

The multiple viewpoints system allows for a derivation of probabilistic dependencies that are then used to calculate the degrees of cross-entropy of the generated solutions. In this dissertation, the probabilistic dependencies will be determined by tension-related features and so using multiple viewpoints and cross-entropy calculations will not be necessary. For a detailed review of statistical methods, see Whorley et al. (2013). More specifically, see the reviews in Conklin (2003) and Whorley and Conklin (2016b) concerning the tasks of style imitation and harmonisation, respectively.

On the other hand, the template strategy can be applied to address our goals, particularly to generate music that follows specific GTTM hierarchical structures. Likewise, by using templates, we would not have to use pattern discovery algorithms but just create the templates ourselves based on GTTM’s rules. A music generation system could then be implemented where GTTM-hierarchical-structure templates are used to generate music with long-term structure. In this way, statistical methods could be applied to the templates to generate musical cues that satisfy the given context (i.e. an input tension profile). This strategy is discussed in more detail in Chapter 6.

5.2.3 | Symbolic AI methods

Symbolic AI methods entail the representation of musical knowledge using symbols from a specific language and the derivation of rules that determine how these symbols should be organised to generate a musical output. They have been used in the past to generate music in real time (Fernández and Vico, 2013).

In Figure 5.1, symbolic AI methods include rule-based methods and generative grammars. These are reviewed below in Sections 5.2.3.1 and 5.2.3.2, respectively.
5.2.3.1 | Rule-based methods

Rule-based methods\(^5\) can incorporate any representation of the musical surface, as long as it uses symbols. The combination of these symbols is then constrained by a collection of rules so that a set of possible musical outputs is generated. Finally, an algorithm, or a collection of them, will be applied to select the most suitable sequence from all the calculated possibilities.

According to Anders and Miranda (2011), rule-based systems consist of three components: a music representation, which concerns information regarding the building blocks of a score (e.g. pitches, notes or voices); a constraint formalism, which specifies how rules are defined and should be applied to the music representation; and a search strategy, which involves the application of a specific algorithm, or collection of algorithms, to achieve the system’s goal.

Music representations can adopt many different forms. In the last decades, some musical toolkits, such as Huron’s (2002) humdum or MIT’s music21 (Cuthbert and Ariza, 2010), have become very popular as they already incorporate tools to process the symbolic representation of music. For more details on music representations see Anders (2007); Simonetta (2018); Wiggins et al. (1993).

Constraint formalisms are represented by both conditional instructions and structures. The most common formalisms are often presented as: numeric relations, which are the natural translation of a collection of rules given a numeric music representation; and pattern mechanisms, which focus on the existence of ordered elements of an a priori defined structure. For more detail on constraint formalisms, see Anders (2007); Anders and Miranda (2011).

Search strategies consist of a set of algorithms to find possible solutions to the problems defined by the constraint formalisms. The collection of possible solutions define

---

\(^5\)The terms rule- and constraint- based are often used in the literature interchangeably to refer to the same type of Automatic Music Generation systems (e.g. Carnovalini and Rodà (2020)). However, some scholars prefer to define specific boundaries to delimit what falls within each concept and what does not. For instance, Nierhaus (2009, p.237) considers that rules “are formulated as “if-then” conditions”, whereas constraints “are conditions (...) that limit the scope of possible parameters”. He also considers that it is possible to define constraints as a special class of rules used to exclude objects in the musical discourse. Thus, he uses the term rule-based to refer to systems based on both rules and constraints. For the sake of simplicity, we follow Nierhaus’ approach and so we group rules and constraints within the term rule-based.
the search space, which can be searched in many different ways (e.g. from left to right, from top to bottom, vertically, horizontally, etc.) (Anders and Miranda, 2011; Nierhaus, 2009). For more detail from the point of view of search algorithms in AI, in general, see Korf’s (2010) review. For more detail concerning Automatic Music Generation, see Nierhaus (2009, pp.231-233) and Anders and Miranda (2011), which together review more than a dozen search strategies.

Lerdahl’s work has its own representations and is governed by rules, either for the definition of a piece’s structure (GTTM), for the calculation of the cognitive distance between musical events (TPS) or for the calculation of the contribution of these events to the total degree of perceived tension (MTT). Therefore, rule-based methods seem to be an appropriate approach to address our goals. Notice that a music generation system could be implemented where a search space is filled based on TPS’s spaces (i.e. the chordal and regional spaces - Figures 3.17 and 3.21, respectively) from which music is generated so that it matches input tension profiles.

5.2.3.2 | Generative grammars

Generative grammars are represented by a pre-defined language. Such languages formulate their internal organisations through relationships which are analogous to the relations in natural languages. In this way, instead of applying a search algorithm, the generative rules are defined as syntactic relations which are applied a number of times to produce a musical output.

In the field of linguistics, the theory of syntax aims to represent the structures of sentences in a given language. Such sentences consist of collections of symbols (i.e. linguistic expressions) that follow a specific structure governed by certain rules. Therefore, a generative grammar can be defined as a set of rules capable of producing well-formed expressions in the context of a language, where well-formedness refers to correctness according to the speakers of the language (Nierhaus, 2009).

In general, the application of generative grammars to the process of music composition can be split into three main steps (Fernández and Vico, 2013): defining a set of rules,
designing a mapping strategy between the rules and the musical objects to be generated and implementing a protocol to control which rules should be activated at each stage of the compositional process. For more details on the implementation of the previous three steps, see Abdallah et al. (2016); Holtzman (1980); Roads (1978).

In this section, we focus on generative grammars to generate scores rather than sounds. The most basic component in a generative grammar consists of a collection of characters, or graphemes, that constitute the grammar’s notation system. The notation system could, for instance, include letters from the alphabet to represent notes or chords (e.g. a, b, c... I, V...), numbers to represent octaves or intervals (e.g. 1, 2, 3...), mathematical symbols to represent sharps and flats (e.g +, -), and many more symbols. The symbols used to define the grammar’s rules, known as tokens, consist of sequences of graphemes (e.g. c+, d4, vii, phrase 1, consequent ...).

A vocabulary of tokens, V, can be defined as the collection of all possible tokens. Within V, it is possible to distinguish between terminal, T, and non-terminal, N, tokens. The former corresponds to the tokens which may be found in the final output of the score. The latter corresponds to the intermediate tokens which will be replaced by other characters before arriving at the final output. Note that N includes the starting token, which is known by the root term, Σ.

The goal of a generative grammar is to transform the first non-terminal, that is the root, into a sequence of terminals. These transformations are, in essence, rewriting tokens. Rewriting or production rules can be denoted:

\[ \text{[left hand-side tokens]} \rightarrow \text{[right hand-side tokens]} \]

Production rules are defined as the replacement of the left hand-side tokens by the right hand-side tokens. An example is shown in Figure 5.2, where a collection of production rules are used to generate the ABA form typical in the classical sonata.\(^6\) In this

\(^6\)ABA is not a typical description of the sonata form. According to Jacobson (2016), the sonata form is a modified binary form “developed out of the binary, or two-part, form prominent in the music of the 17th and early 18th centuries”. Here, however, we are borrowing Holtzman’s (1980) ABA description of the sonata form, which he uses to illustrate the possibilities of generative grammars.
figure, the statement \( x \_ y \rightarrow [z] \) means to only produce ‘z’ if ‘x’ occurs before ‘y’.

It should be noted that the design of GTTM was inspired by Chomsky’s grammatical approach. What is more, according to Lerdahl (1992, p.102):

\[
\text{GTTM} \text{ predicts structural descriptions from musical surfaces by means of a set of rules that ideally corresponds to [a] “listening grammar”}
\]

By “listening grammars” Lerdahl (1992) means a listener’s more or less unconscious mental representation of a piece of music, as opposed to “generative grammars”, which he considers the conscious representation of the events in a piece of music and their organisation. Therefore, generative grammars seem a natural way to address our goals. That is because a music generation system could be implemented where musical structure is defined by a collection of generative grammars based on GTTM’s grammatical principles.

<table>
<thead>
<tr>
<th>[SONATA]</th>
<th>( \rightarrow ) [A, B, A]</th>
</tr>
</thead>
<tbody>
<tr>
<td>[B]</td>
<td>( \rightarrow ) [DEVELOPMENT]</td>
</tr>
<tr>
<td>[A]</td>
<td>( \rightarrow ) [theme1 in KEY, theme2 in KEY]</td>
</tr>
<tr>
<td>[theme1 in KEY]</td>
<td>( \rightarrow ) [theme1 in tonic]</td>
</tr>
<tr>
<td>[theme2 in KEY _ DEVELOPMENT]</td>
<td>( \rightarrow ) [theme2 in dominant _ DEVELOPMENT]</td>
</tr>
<tr>
<td>[DEVELOPMENT _ theme2 in KEY]</td>
<td>( \rightarrow ) [DEVELOPMENT _ theme2 in tonic]</td>
</tr>
<tr>
<td>[DEVELOPMENT]</td>
<td>( \rightarrow ) [modulation of themes]</td>
</tr>
</tbody>
</table>

Figure 5.2: A grammar example, transcribed from Holtzman (1980, p.55), that generates the ABA form typical in the classical sonata.

### 5.3 | Hybridisation

In Sections 5.2.2.3, 5.2.3.1 and 5.2.3.2, it was concluded that the computational methods best suited to our goals are statistical methods, rule-based methods and generative grammars, respectively. In the field of Automatic Music Generation, the combination of
several computational methods into the same generation system is known as *hybridisation*. But, how can we implement these methods simultaneously to design a music generation system that addresses our goals?

In order to generate tonal music that has long-term structure, we could benefit from GTTM's pre-defined structures, such as the *basic form* and/or the *normative structure*. These structures could be used to design template hierarchical trees that the newly generated music should follow. Recall that the use of templates was suggested in Section 5.2.2.3. These template hierarchical trees can be then transformed into sequences of chord labels. These sequences must follow GTTM's rules, so that the requirements of GTTM's pre-defined structures are satisfied. Recall that GTTM's approach was presented as being grammatical in Section 5.2.3.2, so the transformation of the templates into the harmonic sequences can be implemented in the form of a *generative grammar*. For these sequences to truly follow GTTM's pre-defined structures, the generated chord labels connected by the branches of the template trees must satisfy the requirements of GTTM's *prolongational reduction*. These requirements can be calculated as values of TPS distance between the generated chord labels. As there may exist many different chord labels that could fit the template trees, their TPS distances can be used to weight the generation of the chord labels in the *generative grammar*. That is to say, we would be implementing a *statistical generative grammar*. In order for the generated sequences of chord labels to match input tension profiles, the melodic and harmonic content to be generated against the labels must follow MTT's rules. Likewise, the generated content must also still match GTTM's rules, so that the sense of long-term structure is maintained. Therefore, the generation of melodic and harmonic content against the chord labels could follow a *rule-based* approach based on MTT's and GTTM's rules. Again, there may exist endless melodic and harmonic possibilities to be generated against the chord labels. As before, MTT's and GTTM's rules could be used to weight the generation of melodic and harmonic content in the *rule-based* approach. That is to say, we would be implementing a *statistical rule-based* method.

In conclusion, an appropriate form of *hybridisation* to address our goals is that of *generative grammars* combined with *statistical* methods, as well as *rule-based* methods com-
bined with statistical methods. This approach is discussed in more detail in Chapter 6.

5.4 | Summary and Conclusions

In order to seek answers to our Research Question, Chapter 5 has identified the computational methods best suited to our goals.

Section 5.1 has defined the concept of Automatic Music Generation as the application of algorithmic composition techniques by computational means. To frame its scope, Section 5.1 has proposed a taxonomy that includes the main AI and non-AI methods used in the field at present times. The taxonomy is inspired by those in Briot et al. (2020); Carnovalini and Rodà (2020); Fernández and Vico (2013); Lopez-Rincon et al. (2018); Nierhaus (2009); Tatar and Pasquier (2019); Whorley and Conklin (2016b).

Section 5.2 has shown that AI-based methods are better suited to our goals than non-AI ones. This conclusion was drawn after observing that the latter methods often lack possibilities for intervention and structuring, and are better suited to sound synthesis rather than Automatic Music Generation. From the AI-based methods, statistical methods have shown to be suited to real-time generation. This section has also determined that rule-based methods and generative grammars are the methods best suited to implement a music generation system based on Lerdahl’s work.

Finally, Section 5.3 has introduced the concept of hybridisation. It has also concluded that an appropriate form of hybridisation according to our goals is that of generative grammars combined with statistical methods, as well as rule-based methods combined with statistical methods.
Chapter 6

The Music Generator

*Architecture, design and implementation*

In this chapter, we discuss the design and implementation of a system capable of generating tonal music with long-term structure that matches input tension profiles in real time. To do so, we will:

1. determine the compositional strategy best suited to our goals,
2. design the architecture of a music generation system to answer our Research Question,
3. design and implement the components of the music generation system, and
4. provide an illustrative example of the steps implemented by each component of the music generation system.

The first and second steps are addressed in Section 6.1 and the third and fourth steps are addressed in Sections 6.2, 6.3, 6.4 and 6.5, each of which concerns a component of the music generation system.
6.1 | *autognomus*

The music generator

This chapter introduces *autognomus* (as in **Auto**matic **Tension**-**Oriented** **Gene**rate**r of **Mus**ic), our publicly available\(^1\) system developed in order to answer our Research Question.

Section 6.1.1 identifies the compositional strategy best suited to our goals. Based on this strategy, Section 6.1.2 covers the design architecture of *autognomus*. Finally, Section 6.1.3 introduces how *autognomus* works.

6.1.1 | Compositional Strategy

The field in which the use of real-time adaptive music is most widespread is probably that of video games. But, how do games adapt their music to a given narrative?

In the early years of game design, games used to abruptly transition between pre-composed tracks, which were normally associated with certain locations or actions, triggered by a sudden change in the narrative. For example, this type of transition is used in *Pokémon Crystal* (GameFreak, 2000), where the game’s main theme will abruptly transition into the battle theme whenever a new “pokémon” appears through the grass.

Over the years, new transitioning techniques were developed. Two of the most popular techniques are cross-fading and horizontal mixing. The former concerns fading out the current track while fading in a new one. For example, this type of transition is used in *World of Warcraft* (Blizzard-Entertainment, 2004), where tracks associated with different worlds cross-fade when travelling from one world to another. The latter concerns transitioning between two main themes in the game by playing a short track written in such a way that it would smoothly connect both main themes. For example, this type of transition is used in *Monkey Island 2* (LucasArts, 1991), where buildings are associated with different themes and a specific new track is played when moving from one building to another, followed by the new building’s theme.

\(^1\)https://doi.org/10.21964/ou.rd.15028599.v1
The three transitioning techniques described above have been used in many games. In most games that use these techniques, a human composer usually writes a (big) collection of tracks associated with different elements in the game's narrative. It is by transitioning between the different tracks that the game's music avoids boredom. It is then the composer's responsibility to write the tracks in such a way there is some coherence among them and so some sense of long-term structure may be sensed while playing the game. However, this strategy poses a disadvantage in our context. That is because having to generate a (big) collection of tracks, among which there is some sort of structural similarity, may be a very complex and time-consuming task. Thus, a new adaptive technique is needed.

The simplest transitioning technique used in games, relevant to our context, may be that of vertical mixing. This technique concerns transforming a single main track according to the real-time changes of the game's narrative. For example, this type of transition is used in *Banjo-Kazooie* (Rareware, 1998), where each of its worlds is associated with one main theme that gets transformed whenever the player goes underwater, goes into a cavern or bumps into an enemy, among other things (see TVtropes (2013) for more detail). The theme's transformations include small re-scorings of the theme's melody, small re-arrangements of the theme's chords or small re-orchestrations of the theme's instrumentation, among others. Unlike the previous transitioning techniques, vertical mixing seems an appropriate strategy to follow in our context.

### 6.1.2 | TAnDeM: autognomus's architecture

Inspired by *Banjo-Kazooie*'s compositional approach, we have identified four main tasks within the application of the vertical mixing technique. Given a narrative or a particular scenario, the following tasks will happen in a sequential manner:

1. generating a main theme, consisting of a melody and an underlying harmony, that suits the given narrative,

2. arranging the theme's harmony so that it suits the given narrative,
(3) developing new material, similar to the theme, so that boredom is avoided and long-term structure is achieved, and

(4) morphing the theme and its developments, both their melody and harmony, so that they match changes in the given narrative in real time.

The above tasks are independent processes. Therefore, we conceive autognomus as a music generation system that consists of four components or sub-systems, each of which will tackle one of the above tasks. Notice that, in the above tasks, the only one that concerns a real-time process is the last one.

Figure 6.1 presents autognomus’s TAnDeM architecture (as in Themer, Arranger, Developer and Morpher).

![Figure 6.1: A diagram of autognomus’s TAnDeM architecture.]

As seen in Figure 6.1, autognomus consists of four sub-systems. The first sub-system, Themer, generates melodic and harmonic sequences which will be interpreted as the main theme in a given narrative. The second sub-system, Arranger, generates the appropriate arrangement of the voices in the harmony of the newly generated theme. Common arrangement techniques in tonal music include playing the chords as blocks, breaking the chords into arpeggios or generating contrapuntal passages, among others. The third sub-system, Developer, generates developments of the newly generated theme. Common development techniques in tonal music include re-organising the notes in the theme’s melody, re-harmonising the theme or generating new material that is structurally similar to that of the theme, among others. The fourth and final sub-system, Morpher, transforms the generated theme so that it matches input tension profiles in real time.
6.1.3 | How does autognomus work?

Let us imagine we want to use autognomus to automatically generate music for a real-time interactive experience, such as, for instance, a video game. How would we do this?

First, we would run the Themer. This sub-system would generate the main theme in our video game. To do so, we would have to feed the Themer with a collection of input parameters. These include:

- the number of measures we want the theme to consist of,
- a tempo marking in beats per minute (BPM),
- the time-signature and key signature to be used in the theme, and
- the character of the theme.

Themer would generate a theme as a sequence of chord labels and a melody against them. In Themer’s interface, we can listen to the generated materials. We can also re-run some steps of the generation, in case we do not like what we are hearing. Once we are satisfied with the generated theme, Themer will save it so that it can be used by the rest of the sub-systems.

Second, we would run the Arranger. This sub-system would organise the voicing of the chords in the generated theme. We would have to feed the Arranger with Themer’s outputs. Arranger would then generate the corresponding arrangement of the chords in blocks and it will save it so that it can be used by the rest of the sub-systems. In the current version of autognomus, we cannot choose between different types of arrangements. This is something we want to implement in the future.

Third, we would run the Developer, but only if we want the generated theme to be developed. We would have to feed the Developer with Themer’s outputs. Developer would then generate a collection of developments (i.e. independent compositions) and it will save them so that they can be used by the rest of the sub-systems. The developments are twice as long as the input theme. That is to say, if the input theme lasts eight measures,
each of the generated developments would last sixteen measures. In order for the generated developments to be used in our video game, they would have to be arranged. That is to say, we would have to run each development through the Arranger, as described above in the second step.

Fourth, we would run the Morpher. We would have to feed the Morpher with Themer’s, Arranger’s and Developer’s outputs. Morpher would then loop the generated materials. It would play the generated theme twice, followed by one of the developments, and this process would repeat ad infinitum. In our video game, this is the music that would be played whenever we are in a safe place. Let us imagine that, in our video game, there are monsters who want to capture us. We could then quantify the degree of tension in the game as inversely related to the distance to the monsters. That is to say, the closer we get to a monster, the tenser, and vice versa. Morpher also considers a tension threshold. This threshold can be interpreted here as the minimum degree of tension for us to start to get worried about being captured by a monster. That is to say, if the current degree of tension is below the threshold, Morpher would interpret that we are in a safe place. And so it would continue playing the looped theme and its developments. However, if the current degree of tension is above the threshold, Morpher would transform the theme or the development so that it matches the current degree of tension.

The implementation of autognomus’s Themer, Arranger, Developer and Morpher is presented in Sections 6.2, 6.3, 6.4 and 6.5, respectively.

6.2 | Themer

The generator of themes

To meet our goals, the music generated by autognomus must have long-term structure. To do so, autognomus incorporates the Themer, which aims to generate a theme whose structure is well-defined and so can be used as reference.

In order to generate a theme with a well-defined structure, Themer takes advantage of GTTM’s components and their rules. That is to say, instead of applying GTTM as an analytical tool to an existing piece of music, Themer applies GTTM to constrain the
Themer implements the generation of themes in seven steps. These are discussed in Sections 6.2.1 to 6.2.7.

6.2.1 | Step 1: defining the input parameters

The first step in the implementation of Themer is to define the following parameters: the number of measures, which could either be 4, 6, 8, 10, or 12; tempo, which could be any integer value representing beats per minute (BPM); time signature, which could either be $\frac{2}{4}$, $\frac{3}{4}$, $\frac{4}{4}$, $\frac{6}{8}$, $\frac{9}{8}$, or $\frac{12}{8}$; key, which could be any of the thirty keys that exist in the tonal context;\(^2\) and character, which could either be adagio, andante, allegro or presto.\(^3\)

6.2.1.1 | An example of the application of Themer’s Step 1

To illustrate the application of Themer’s Step 1, Figure 6.2 shows a template for a theme generated considering the parameters implemented in Themer by default.

\[
\text{\textit{allegro}} \quad \text{\textit{\texttt{j}} = 60}
\]

\[
\begin{array}{cccccccc}
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\end{array}
\]

Figure 6.2: An example of the application of Themer’s Step 1.

Notice that Figure 6.2 consists of an empty score where the number of measures is equal to 8, tempo is equal to 60 BPM, the time signature is $\frac{2}{4}$, the key signature is that of C major/A minor and the character is allegro.

---

\(^2\)There exist fifteen different key signatures: seven that include sharps, seven that include flats and one with no sharps nor flats. Likewise, there exist two keys, a major and a minor one, for each key signature. Thus, there exist thirty keys in total.

\(^3\)According to Randel (2003, entry: Performance marks, definition: 1), Italian performance markings are used to indicate the speed of the music, but also its character. We have adopted this view and that is why character is assigned an Italian performance marking. To differentiate tempo from character, the latter will correlate with note density. In this way, if two themes share the same tempo, one of them may convey a slower or faster movement compared to the other depending on their respective characters.
6.2.2 | Step 2: generating the theme’s structure

The second step in the implementation of Themer is to generate the structure of the theme. To do so, Themer implements a generative grammar. The implementation of the grammar is described in Sections 6.2.2.1 to 6.2.2.4.

6.2.2.1 | The theme as a complete phrase

Themer conceives the theme to be generated as a complete phrase. In order to generate its structure, the theme is split into two semi-phrases. The first semi-phrase will be labelled as the prolongational beginning (PB) and the second semi-phrase will be labelled as the cadential structure (CS). The generation of these two semi-phrases is defined by the following production rule:

\[ \text{[theme]} \rightarrow \text{[PB]} + \text{[CS]} \]  

(6.1)

What do PB and CS represent? These structures, and their names, have been adapted from GTTM. PB represents the prolongations of the beginning in GTTM’s basic form + normative structure, whereas CS represents the prolongations of its cadence. Recall that this structure was introduced in Figure 3.9. It is shown again in Figure 6.3 to illustrate what PB and CS represent.

Figure 6.3: GTTM’s basic form + normative structure including its prolongational beginning (PB) and its cadential structure (CS).
6.2.2.2 | The theme’s hierarchical structure

As discussed in Section 6.2.1, the number of measures allowed in Themer are all even numbers. Considering this, PB and CS will be both assigned the same number of measures, that is half of the number of measures in the theme each. In this way, when generating the grouping structures in the theme, according to GTTM’s rules, the theme’s whole phrase will be perceived as a group itself, as proposed by GWFR 2. PB’s and CS’s semi-phrases will be perceived as smaller groups contained in the whole-phrase larger group, as proposed by GWFRs 3, 4 and 5. Likewise, by splitting the theme into two equal-length semi-phrases, these will be perceived as symmetrical grouping structures, as proposed by GPR 5.

As discussed in Section 6.2.1, the maximum number of measures allowed in Themer is twelve. This number of measures will correspond to a six-measure PB and a six-measure CS. For the sake of simplicity, we have constrained the minimum duration of a chord in PB and in CS to be equal to one measure. Thus, for six-measure semi-phrases, Themer must consider at least six different chords within PB and another six different chords within CS. Notice, however, that only three different chords are considered in Figure 6.3 within PB and another three within CS. In order to consider more chords in the theme to be generated, new branches must be added to Figure 6.3.

Although there are many possible ways to add new branches to Figure 6.3, the simplest one that best matches GTTM’s principles is shown in Figure 6.4. But, why is this the simplest way?

In Figure 6.3, PB consists of three chords: a first tonic chord, its reprise and a final chord. As discussed in Section 3.2.4, the reprise chord is an optional one. Thereby, new chords in PB will attach better to PB’s final chord. In order to generate a six-chord PB structure, which recall is the maximum number of chords there may exist in PB, at least four new branches should be able to attach to PB’s final chord. In this way, a PB structure without a reprise chord would consist of its first chord, its final chord and four extra chords that attach to PB’s final chord. These four extra chords may either be left- or right-branching subordinates. However, defining the level of the prolongational
reduction to which these chords belong would be a challenging task. An easier alternative is to generate a new chord, between PB’s reprise and final chord, that directly attaches to PB’s first chord. This would leave a PB structure that consists of four or three chords, depending on whether PB includes a reprise chord or not. So, to generate a six-chord PB structure, both PB’s final chord and the newly generated intermediate chord may include subordinate branching connections. For the sake of simplicity, we constrain the maximum number of subordinate connections, in each direction (i.e. left and right), to not be greater than one. In this way, a PB structure may now include a first chord, a reprise chord, an intermediate chord, a left-branching subordinate to the intermediate chord, a right-branching subordinate to the intermediate chord, a final chord, a left-branching subordinate to the final chord and a right-branching subordinate to the final chord. Notice that this new collection of chords includes up to eight chords and so it is possible to generate a six-chord PB structure. This is shown in Figure 6.4 (events 1 to 8).

Similarly, in Figure 6.3, CS consists of three chords: a first subdominant chord, a dominant chord and a final tonic chord. According to GTTM, there are two main chords that might also be included in CS: first, a cadential 6\(^4\) chord that directly attaches to CS’s dominant chord right after its subdominant chord; and, second, a progression of subdominant chords whose branches attach to one another until they attach to CS’s subdominant chord. In order to generate a six-chord CS structure, the maximum number of newly generated subdominant chords would be equal to three. In this way, a CS structure may now include a three-subdominant-chord progression, CS’s original subdominant chord, a cadential 6\(^4\) chord, its dominant chord and the final tonic chord. Notice that this new collection of chords includes up to seven chords and, even if CS does not include a cadential 6\(^4\) chord, it is possible to generate a six-chord CS structure. This is shown in Figure 6.4 (events 9 to 15).

Notice that the hierarchical structure shown in Figure 6.4 only refers to chords per time-span. That is to say, it does not focus on the branching connections that would exist within each time-span concerning the melody. In other words, this type of hierarchical structure concerns the level of the reduction where each time-span is reduced to a single chord. For instance, in the reduction of Mozart’s Twinkle, twinkle theme shown in
Figure 6.4: A prolongational structure including all of Themer’s possible chords, which are labelled in Themer as: (1) First Chord, (2) Reprise Chord, (3) Left Subother Chord, (4) Other Chord, (5) Right Subother Chord, (6) Left Subnormative Chord, (7) Normative Chord, (8) Right Subnormative Chord, (9) $\text{Sub}_3$ Subdominant Chord, (10) $\text{Sub}_2$ Subdominant Chord, (11) $\text{Sub}_1$ Subdominant Chord, (12) Subdominant Chord, (13) Cadential $6_4$ Chord, (14) Dominant Chord and (15) Last Chord.

Figure 3.4, this type of hierarchical structure refers to the level of the reduction shown in Figure 3.4c. For the sake of simplicity, this type of hierarchical structure is the one that this and the rest of autognomus’s sub-systems will be considering. By doing this, we are assuming that the notes in the melody in each time-span can be reduced to the head of the span. If autognomus manages to match the intended hierarchical structures, this decision is then supported by PRWFR 1.

6.2.2.3 | The theme’s harmonic templates

Many different combinations of the chords in Figure 6.4 are possible depending on the number of measures in PB and in CS. For the sake of simplicity, we have implemented two pre-defined templates of valid combinations of chords in PB and in CS in Themer. The templates include the combinations of chords that match the following constraints:

- every chord must last at least one measure,
- PB must always include a First Chord and a Normative Chord and CS must always include a Dominant Chord and a Last Chord (so that GTTM’s normative structure
is satisfied),

- any **Left Sub- Chord** must be followed by its corresponding dominating chord and any **Right Sub- Chord** must be preceded by its corresponding dominating chord, and

- the distribution of chords must be symmetrical (so that GPR 5 and MPR 1 are satisfied).

The chord-templates of the possible distribution of chords in PB and in CS, pre-defined in *Themer* according to the above constraints, are shown in Tables 6.1 and 6.2, respectively.

<table>
<thead>
<tr>
<th>Measure</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>N/2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>First</td>
<td>Normative</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>First</td>
<td>Reprise</td>
<td>Normative</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>First</td>
<td>Left Subnorm</td>
<td>Normative</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>First</td>
<td>Normal</td>
<td>Right Subnormal</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>First</td>
<td>Reprise</td>
<td>Left Subnormal</td>
<td>Normative</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>First</td>
<td>Reprise</td>
<td>Right Subnormal</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>First</td>
<td>Left Subnormal</td>
<td>Normative</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>First</td>
<td>Reprise</td>
<td>Other</td>
<td>Normative</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>First</td>
<td>Reprise</td>
<td>Other</td>
<td>Left Subnormal</td>
<td>Normative</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>First</td>
<td>Reprise</td>
<td>Left Subnormal</td>
<td>Other</td>
<td>Normal</td>
<td>Right Subnormal</td>
</tr>
<tr>
<td></td>
<td>First</td>
<td>Reprise</td>
<td>Left Subnormal</td>
<td>Other</td>
<td>Left Subnormal</td>
<td>Normative</td>
</tr>
<tr>
<td></td>
<td>First</td>
<td>Reprise</td>
<td>Other</td>
<td>Left Subnormal</td>
<td>Other</td>
<td>Normal</td>
</tr>
<tr>
<td></td>
<td>First</td>
<td>Reprise</td>
<td>Other</td>
<td>Right Subnormal</td>
<td>Left Subnormal</td>
<td>Normal</td>
</tr>
<tr>
<td></td>
<td>First</td>
<td>Reprise</td>
<td>Left Subnormal</td>
<td>Other</td>
<td>Normal</td>
<td>Right Subnormal</td>
</tr>
</tbody>
</table>

This constraint poses a limitation. Traditionally, the rate of harmonic changes tends to increase with regards to the proximity to a cadence. This often translates into chord sequences in which the rate of harmonic changes is not constant (recall, for instance, the first phrase of the theme in Mozart's *Ah vous dirai-je, Maman*, K. 265/300e, which was shown in Figure 3.10). However, adopting this strategy in the implementation of *Themer* would pose a risk concerning the definition of grouping and metrical structures, according to GTTM's rules, and so the strategy is not implemented.
Chapter 6. The Music Generator

Themer: the generator of themes

6.2.2.4 Generating the theme’s harmonic structure

In order to complete the generation of the theme’s harmonic structure, the previously initiated production rule, Rule 6.1, is completed as follows:

\[
[N\text{-measure theme}] \rightarrow [N/2\text{-measure PB}] + [N/2\text{-measure CS}]
\]

\[
[PB] \rightarrow [\text{random(PB’s template, for } N/2 \text{ measures})]
\]

\[
[CS] \rightarrow [\text{random(CS’s template, for } N/2 \text{ measures})]
\]

(6.2)

where the last two lines represent two stochastic production rules which randomly select PB’s and CS’s harmonic structure from PB’s and CS’s chord-templates (i.e. Tables 6.1 and 6.2 respectively), given a theme of \(N\) measures.

6.2.2.5 An example of the application of Themer’s Step 2

To illustrate the application of Themer’s Step 2, Figure 6.5 shows the application of Rule 6.2 to the theme’s template previously generated in Figure 6.2.
Figure 6.5 consists of an eight-measure phrase (i.e. \(N = 8\)) with the corresponding tree structure defined according to Figure 6.4. The first semi-phrase, consisting of the first four measures, represents PB. The second semi-phrase, consisting of the last four measures, represents CS. The chords in PB and in CS have been stochastically selected from their respective chord-templates. More specifically, PB’s chords have been stochastically selected from Table 6.1 and correspond to the first row in this table for \(N/2\) equal to 4 measures, and CS’s chords have been stochastically selected from Table 6.2 and correspond to the first row in this table for \(N/2\) equal to 4 measures.

6.2.3 | Step 3: generating the chord labels in the theme's first semi-phrase

The third step in the implementation of Themer is to generate the chord labels in PB. To do so, Themer incorporates new production rules to the generative grammar initiated in Rule 6.2. The derivation of these rules is discussed in Sections 6.2.3.1 to 6.2.3.3.

6.2.3.1 | Generating the theme's First Chord

As discussed in Section 6.2.2.2, the possible chords in PB correspond to the first eight elements in Figure 6.4. From these, the first chord label Themer generates is that of First Chord. According to GTTM, the first chord in its basic form + normative structure is a tonic chord. Thus, Themer generates the label of its First Chord according to the
following production rule:

\[
\begin{align*}
\text{[First Chord key is major]} & \rightarrow [I] \\
\text{[First Chord key is minor]} & \rightarrow [i]
\end{align*}
\]

(6.3)

where key refers to the key label Themer was fed as input in Section 6.2.1 and ‘_’ represents an if (i.e. conditional) statement, as in Holtzman (1980).

### 6.2.3.2 Generating the theme's remaining Dominating Chords

The next chord labels Themer generates are those of its remaining Dominating Chords. These include Reprise Chord, Other Chord and Normative Chord. Notice that, as shown in Figure 6.4, these three chords attach to First Chord as right-branching connections. According to GTTM’s stability conditions, which were discussed in Section 3.2.4, right-branching connections are more stable if they ascend along the circle-of-fifths and if their roots are close in the circle. Thus, Themer generates the labels of the Dominating Chords according to the following statistical (i.e. weighted random generation) production rule:

\[
\begin{align*}
\text{[Dominating Chord]} & \overset{p_1}{\rightarrow} [C_1] \\
\text{[Dominating Chord]} & \overset{p_2}{\rightarrow} [C_2] \\
\text{[Dominating Chord]} & \overset{p_3}{\rightarrow} [C_3] \\
\text{[Dominating Chord]} & \overset{p_4}{\rightarrow} [C_4]
\end{align*}
\]

(6.4)

where Dominating Chord may either be Reprise Chord, Other Chord or Normative Chord.

In order to match GTTM’s stability conditions, \( C_i \) are defined as the closest chord labels to First Chord in the ascending (i.e. to the right) diatonic circle-of-fifths. That is to say, \( ^5 \)Hereinafter, dominating chords must not be mistaken with dominant chords.
since First Chord is going to be equal to I or i, depending on the quality of the theme’s key, $C_i$ are defined as:

<table>
<thead>
<tr>
<th>if key is major</th>
<th>if key is minor</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_1 = I$</td>
<td>$C_1 = i$</td>
</tr>
<tr>
<td>$C_2 = V$</td>
<td>$C_2 = V$</td>
</tr>
<tr>
<td>$C_3 = ii$</td>
<td>$C_3 = ii^6$</td>
</tr>
<tr>
<td>$C_4 = vi$</td>
<td>$C_4 = VI$</td>
</tr>
</tbody>
</table>

And the probabilities, $p_i$, depending on whether Dominating Chord is defined as Reprise Chord, Other Chord or Normative Chord, are defined as follows:

<table>
<thead>
<tr>
<th>Reprise</th>
<th>Other</th>
<th>Normative</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_1 = 1/2$</td>
<td>$p_1 = 1/4$</td>
<td>$p_1 = 1/1$</td>
</tr>
<tr>
<td>$p_2 = 0$</td>
<td>$p_2 = 1/4$</td>
<td>$p_2 = 1/2$</td>
</tr>
<tr>
<td>$p_3 = 0$</td>
<td>$p_3 = 1/4$</td>
<td>$p_3 = 1/3$</td>
</tr>
<tr>
<td>$p_4 = 1/2$</td>
<td>$p_4 = 1/4$</td>
<td>$p_4 = 1/4$</td>
</tr>
</tbody>
</table>

Themer first generates its Normative Chord. Notice that, as shown above, the closer a chord label is to First Chord in the diatonic circle-of-fifths, the greater is its Normative probability and vice versa. It could be argued that the tonic chord (i.e. I or i) should not be considered as a possible Normative Chord. That is because this chord takes its name from GTTM’s normative structure and it is intended to represent the main chord onto which First Chord progresses to achieve a rise in tension that will then fall at the final cadence. In this way, if First Chord and Normative Chord are both a tonic chord, such a rise in tension may not exist. However, Themer includes the tonic chord as the most probable Normative Chord because of three reasons: first, it is the chord where the theme’s first semi-phrase will repose and so a tonic chord is a good candidate to convey this sense of closure; second, although a tonic Normative Chord will not translate into a rise in tension from First Chord, such a rise may occur because of Reprise Chord and/or Other Chord, if they exist in PB; and, third, in the worst case scenario, where Normative Chord is a tonic chord immediately preceded by First Chord (i.e. see first row in Table 6.1), the Arranger will later assign a different inversion to Normative Chord so that at least some tension is generated according to MTT’s surface tension, $T_{diss}$.


Themer then generates Reprise Chord and Other Chord, if they exist in PB. The probabilities assigned above to these two chords might however change depending on the newly generated Normative Chord. That is because Themer does not allow Reprise Chord and Other Chord to be the same as Normative Chord. Therefore, if the generated Normative Chord turns out to be, for instance, C₁, the values of \( p_1 \) of both Reprise Chord and Other Chord will be redefined as equal to zero.

As discussed in Section 3.2.4, the reprise in GTTM’s basic form + normative structure consists of a reiteration of the material of a theme’s first tonic chord. Thus, one might think that Reprise Chord should always be a tonic chord. However, considering that Themer does not allow Reprise Chord to be the same as Normative Chord, if the latter is a tonic chord, there will not be a possible chord label to be associated with Reprise Chord. That is why the Reprise probabilities shown above allow the generation of Reprise Chord as either a tonic or sixth-degree chord. Themer includes the latter as it sometimes acts as a tonic function in the tonal context.\(^6\)

Apart from not being the same as Normative Chord, the generation of Other Chord is not subject to any other constraints. That is why the Other probabilities shown above are all equal. Although the generation of Other Chord is presented in Rule 6.4 as being statistical, it may be better described as being stochastic.

6.2.3.3 | Generating the theme’s subordinate chords

The last chord labels Themer generates are those of the left and/or right subordinates that attach to Other Chord and/or Normative Chord (i.e. Left Subother Chord, Right Subother Chord, Left Subnormative Chord and Right Subnormative Chord), if they exist in PB. To generate these, Themer incorporates the following stochastic production rule:

\[
[\text{Subordinate Chord}] \rightarrow [\text{random}(C_1, C_2, \ldots, C_n)]
\]

\(^6\)Sometimes, the third-degree chord also acts as a tonic function. We have decided not to consider this chord as a possible Reprise Chord for the sake of simplicity.
where $C_1, C_2, \ldots, C_n$, with $n \leq 3$, represent a list of valid subordinate chords. But, what makes $C_i$ a valid subordinate?

To decide whether a chord is a valid subordinate, Themer first calculates a list of candidate subordinate chords. Given a Dominating Chord (i.e., either Reprise Chord or Other Chord), the corresponding candidate chords, $C_1$, $C_2$ and $C_3$, will be the three chords that immediately follow the Dominating Chord in the diatonic circle-of-fifths. Notice, however, that the subordinate chords may now either be right- or left-branching subordinates. Thereby, to satisfy GTTM's stability conditions, the list of candidate chords for a right-branching subordinate will be calculated by moving from Dominating Chord to the right in the diatonic circle-of-fifths, whereas the list of candidate chords for a left-branching subordinate will be calculated by moving from Dominating Chord to the left.

For example, let $I$ be the given Dominating Chord. The list of candidate right-branching subordinates will consist of the chords: $C_1=V$, $C_2=ii$ and $C_3=vi$; whereas the list of candidate left-branching subordinates will consist of the chords: $C_1=IV$, $C_2=vii^0$ and $C_3=iii$.

To decide whether the candidate chords, calculated as above, are valid subordinates, Themer makes sure that, if they were to be generated, they would still match the branching connections shown in Figure 6.4. To do so, only those candidate chords that satisfy a collection of constraints will be considered as valid subordinates. These constraints are defined as follows:

- $d_{d\rightarrow c} \leq d_{f\rightarrow c}$,
- $d_{d\rightarrow c} \leq d_{n\rightarrow c}$,
- $d_{d\rightarrow c} \leq d_{r\rightarrow c}$ (if Reprise Chord exists), and
- $d_{d\rightarrow c} \leq d_{o\rightarrow c}$ (if Other Chord exists),

calculated using the TPS distance, $\delta$, as in Rule 3.2:

- $d_{d\rightarrow c} = \delta$(Dominating Chord $\rightarrow$ candidate chord),
- $d_{f\rightarrow c} = \delta$(First Chord $\rightarrow$ candidate chord),
Chapter 6. The Music Generator

Themer: the generator of themes

- \( d_{n \rightarrow c} = \delta(\text{Normative Chord} \rightarrow \text{candidate chord}) \),
- \( d_{r \rightarrow c} = \delta(\text{Reprise Chord} \rightarrow \text{candidate chord}) \), and
- \( d_{o \rightarrow c} = \delta(\text{Other Chord} \rightarrow \text{candidate chord}) \).

Only if a candidate chord satisfies the above constraints it is considered a valid subordinate.

Notice that, by applying the above constraints, Themer only considers a candidate chord as a valid subordinate if its TPS’s distance to the chord that dominates it is equal to or less than any other distance to the rest of the chords in PB. In this way, the branching connections shown in Figure 6.4 will be satisfied.

It may occur that no candidate chord matches the above constraints and so there are no valid subordinate chords. When that is the case, Themer generates the corresponding Subordinate Chord as being equal to its corresponding Dominating Chord, as defined by the following production rule:

\[
[\text{Subordinate Chord}] \rightarrow [\text{Dominating Chord}] \quad (6.6)
\]

6.2.3.4 | An example of the application of Themer’s Step 3

To illustrate the application of Themer’s Step 3, Figure 6.6 shows the application of Rules 6.3, 6.4 and 6.5 to the theme’s template previously generated in Figure 6.5.

Figure 6.6a includes PB’s structure, borrowed from Figure 6.5. Figure 6.6b illustrates the generation of First Chord. Let us imagine that the theme is in the key of C major. First Chord will then be generated as I according to Rule 6.3. Figure 6.6c illustrates the generation of Normative Chord. According to Rule 6.4, Normative Chord may be either I, V, ii or vi. Let us imagine that Themer generates Normative Chord as I, as it will be assigned the highest probability. If that is the case, the output will be that of Figure 6.6c. Figure 6.6d illustrates the generation of Reprise Chord. According to Rule 6.4, Reprise Chord
Figure 6.6: An example of the application of Themer’s Step 3 including: (a) PB’s structure from Figure 6.5, (b) the generation of First Chord, (c) the generation of Normative Chord, (d) the generation of Reprise Chord and (e) the generation of Left Subnormative Chord.

Chord may either be I or vi. However, because of Normative Chord being generated as I, Reprise Chord cannot be I. Thus, it must be vi, as shown in Figure 6.6d.

Finally, Figure 6.6e illustrates the generation of Left Subnormative Chord. According to Rule 6.5, the possible Left Subnormative Chords are IV, vii° and iii. To check whether these are valid subordinates, the following values of TPS distance must be calculated:

- \( d_{d \rightarrow c} = \delta(\text{Normative Chord} \rightarrow \text{candidate chord}) \),
- \( d_{f \rightarrow c} = \delta(\text{First Chord} \rightarrow \text{candidate chord}) \), and
- \( d_{r \rightarrow c} = \delta(\text{Reprise Chord} \rightarrow \text{candidate chord}) \).

<table>
<thead>
<tr>
<th>Candidate = IV</th>
<th>Candidate = vii°</th>
<th>Candidate = iii</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d_{d \rightarrow c} = 5 )</td>
<td>( d_{d \rightarrow c} = 8 )</td>
<td>( d_{d \rightarrow c} = 7 )</td>
</tr>
<tr>
<td>( d_{f \rightarrow c} = 5 )</td>
<td>( d_{f \rightarrow c} = 8 )</td>
<td>( d_{f \rightarrow c} = 7 )</td>
</tr>
<tr>
<td>( d_{r \rightarrow c} = 7 )</td>
<td>( d_{r \rightarrow c} = 8 )</td>
<td>( d_{r \rightarrow c} = 5 )</td>
</tr>
</tbody>
</table>
Notice that, for all three candidates, \( d_{d \rightarrow c} \) is equal to \( d_{f \rightarrow c} \). Notice as well that, only in the case of \( iii \), \( d_{d \rightarrow c} \) is greater than \( d_{r \rightarrow c} \). Therefore, from the three candidates, only IV and vii\(^\circ\) are valid Left Subnormative Chords. Let us imagine that Themer generates Left Subnormative Chord as IV. If that is the case, the output will be that of Figure 6.6e.

6.2.4 | Step 4: generating the chord labels in the theme's second semi-phrase

The fourth step in the implementation of Themer is to generate the chord labels in CS. To do so, Themer incorporates new production rules to the generative grammar consisting of Rules 6.2, 6.3, 6.4 and 6.5. The derivation of these rules is discussed in Sections 6.2.4.1 to 6.2.4.5.

6.2.4.1 | Generating the theme's Last Chord

As discussed in Section 6.2.2.2, the possible chords in CS correspond to the last seven elements in Figure 6.4. From these, the first chord label Themer generates is that of Last Chord. According to GTTM, the last chord in its basic form + normative structure is a tonic chord. Thus, Themer generates the label of the theme's Last Chord following Rule 6.3, which was previously introduced to generate First Chord as a tonic chord.

6.2.4.2 | Generating the theme's Cadential\(^6\) Chord

The next chord label Themer generates is that of Cadential\(^6\) Chord, if it exists in CS. To do so, Rule 6.3 is applied again. In this way, Cadential\(^6\) Chord will also be generated as a tonic chord, but with a dominant function, as proposed in Roca and Molina (2006). The Arranger sub-system will be responsible for generating its inversion.

6.2.4.3 | Generating the theme's Dominant Chord

The next chord label Themer generates is that of Dominant Chord. Although there exist different chords that may act as a dominant function, the penultimate chord in GTTM's
Chapter 6. The Music Generator

Themer: the generator of themes

**basic form + normative structure** is defined as a V chord. Thus, Themer incorporates the following *stochastic* new *production rule* to generate its Dominant Chord:

\[
[\text{Dominant Chord}] \rightarrow [\text{random}(V, V^7)] \tag{6.7}
\]

Notice that the above rule will randomly generate Dominant Chord as either a dominant or a dominant seventh chord.

6.2.4.4 | Generating the theme's Subdominant Chord

The next chord label Themer generates is that of Subdominant Chord. To do so, Themer calculates a list of candidate chords. These include those chords that typically precede the final cadence in GTTM's *basic form + normative structure* acting as subdominant functions (i.e. the chords built upon the fourth-degree, second-degree and sixth-degree in the current diatonic scale) (Lerdahl and Jackendoff, 1983, p.192). Themer stochastically selects Subdominant Chord from the candidate chords, as defined by the following *production rule*:

\[
[\text{Subdominant Chord}] \rightarrow [\text{random}(C_1, C_2, C_3)] \tag{6.8}
\]

where:

<table>
<thead>
<tr>
<th>if key is major</th>
<th>if key is minor</th>
</tr>
</thead>
<tbody>
<tr>
<td>C_1 = IV</td>
<td>C_1 = iv</td>
</tr>
<tr>
<td>C_2 = ii</td>
<td>C_2 = ii\degree</td>
</tr>
<tr>
<td>C_3 = vi</td>
<td>C_3 = VI</td>
</tr>
</tbody>
</table>

6.2.4.5 | Generating the theme's Sub-Subdominant Chord

The last chord labels Themer generates are those of the three left-branching subordinates in CS (i.e. Sub_1, Sub_2 and Sub_3 Subdominant Chords). To generate these, Themer ap-
plies Rule 6.5 again. As discussed in Section 6.2.3.3, in this rule subordinate chords are *stochastically* generated from a list of valid chords.

In order to determine which chords are valid subordinates to Subdominant Chord, *Themer* first calculates a list of candidate chords. Considering that these subordinate chords are all supposed to act as subdominant functions, *Themer* borrows the list of candidate chords from those that apply in Rule 6.8 (i.e. IV, ii and vi, if the key’s theme is major; otherwise, iv, ii° and VI). Likewise, *Themer* also considers the secondary dominants (with and without a diatonic seventh) of Subdominant Chord as candidate chords. That is to say, let X be the Subdominant Chord in a given theme. In addition to the fourth-, second- and sixth-degree chords, V/X and V7/X are also considered candidate subordinate chords to the Subdominant Chord. This is in line with the possible chords considered by GTTM at the subdominant location in the *basic form + normative structure* (Lerdahl and Jackendoff, 1983, pp.191-196).

To decide whether the candidate chords are valid subordinates to Subdominant Chord, *Themer* makes sure that, if they were to be generated, they would still match the branching connections shown in Figure 6.4. To do so, only those candidate chords that satisfy a collection of constraints will be considered as valid subordinates. We have defined these constraints as follows:

- \[ d_{l \rightarrow c} \leq d_{d \rightarrow c} \text{, and} \]
- A candidate chord has not already been generated as a Sub, Subdominant Chord, calculated using the TPS distance, \( \delta \), as in Rule 3.2:

  - \[ d_{d \rightarrow c} = \delta(\text{Dominant Chord} \rightarrow \text{candidate chord}), \] and
  - \[ d_{l \rightarrow c} = \delta(\text{Dominating Chord} \rightarrow \text{candidate chord}). \]

Only if a candidate chord satisfies the above constraints it is considered a valid subordinate.
Notice that, by applying the above constraints, Themer only considers a candidate chord as a valid subordinate if its TPS's distance to the chord that dominates it is equal to or less than the distance to Dominant Chord.

It may be the case that no candidate chord matches the above constraints and so there are no valid subordinate chords. If that is the case, Themer generates the corresponding Sub 1 Subdominant Chord as being equal to Subdominant Chord, as defined by the following production rule:

\[
[Sub_1 \text{ Subdominant Chord}] \rightarrow [\text{Subdominant Chord}]
\] (6.9)

One of the particularities of the generation of Sub 1, Sub 2 and Sub 3 Subdominant Chords is that they are generated from the right to the left with regards to Figure 6.4. That is to say, Sub 1 Subdominant Chord is generated first, and its Dominating Chord will be Subdominant Chord. Sub 2 Subdominant Chord is generated second, and its Dominating Chord will be Sub 1 Subdominant Chord. Sub 3 Subdominant Chord is generated third, and its Dominating Chord will be Sub 2 Subdominant Chord.

### 6.2.4.6 | An example of the application of Themer's Step 4

To illustrate the application of Themer's Step 4, Figure 6.7 shows the application of Rules 6.3, 6.7, 6.8 and 6.5 to the theme's template previously generated in Figure 6.5.

Figure 6.7a includes CS's structure, borrowed from Figure 6.5. Figure 6.7b illustrates the generation of Last Chord. As in Section 6.2.3.4, let us imagine that the theme is in the key of C major. Last Chord will then be generated as I according to Rule 6.3. Figure 6.7c illustrates the generation of Dominant Chord. According to Rule 6.7, Dominant Chord may be either V or V 7. Let us imagine that Themer randomly generates Dominant Chord as V 7. If that is the case, the output will be that of Figure 6.7c. Figure 6.7d illustrates the generation of Subdominant Chord. According to Rule 6.8, Subdominant Chord may be either IV, ii or vi. Let us imagine that Themer randomly
Chapter 6. The Music Generator

Themer: the generator of themes

Figure 6.7: An example of the application of Themer’s Step 4 including: (a) CS’s structure from Figure 6.5, (b) the generation of Last Chord, (c) the generation of Dominant Chord, (d) the generation of Subdominant Chord and (e) the generation of Sub1 Subdominant Chord.

generates Subdominant Chord as ii. If that is the case, the output will be that of Figure 6.7d.

Finally, Figure 6.7e illustrates the generation of Sub1 Subdominant Chord. According to Rule 6.8, the original candidate subordinate chords will be IV, ii or vi, although chords V/ii and V\(^7\)/ii will also be candidate chords (notice that the current Dominating Chord in this case is ii; that is to say, Subdominant Chord). To calculate whether these five chords are valid subordinates, the following values of TPS distance must be calculated:

- \(d_{d\rightarrow c} = \delta(\text{Dominant Chord} \rightarrow \text{candidate chord})\), and
- \(d_{l\rightarrow c} = \delta(\text{Subdominant Chord} \rightarrow \text{candidate chord})\).

<table>
<thead>
<tr>
<th>Candidate = IV</th>
<th>Candidate = ii</th>
<th>Candidate = vi</th>
<th>Candidate = V/ii</th>
<th>Candidate = V(^7)/ii</th>
</tr>
</thead>
<tbody>
<tr>
<td>(d_{d\rightarrow c} = 8)</td>
<td>(d_{d\rightarrow c} = 5)</td>
<td>(d_{d\rightarrow c} = 8)</td>
<td>(d_{d\rightarrow c} = 11)</td>
<td>(d_{d\rightarrow c} = 11)</td>
</tr>
<tr>
<td>(d_{l\rightarrow c} = 7)</td>
<td>(d_{l\rightarrow c} = 0)</td>
<td>(d_{l\rightarrow c} = 5)</td>
<td>(d_{l\rightarrow c} = 8)</td>
<td>(d_{l\rightarrow c} = 9)</td>
</tr>
</tbody>
</table>
Chapter 6. The Music Generator

Themer: the generator of themes

Notice that, for all five candidates, $d_{l_{-c}}$ is less than $d_{d_{-c}}$. Thus, they are all valid Sub$_1$ Subdominant Chords. Let us imagine that Themer randomly generates Sub$_1$ Subdominant Chord as $V/ii$. If that is the case, the output will be that of Figure 6.7e.

6.2.5 | Step 5: generating a melody to fit the theme’s first semi-phrase

The fifth step in the implementation of Themer is to generate a melody to fit PB’s chord labels. To do so, Themer implements a statistical rule-based approach taking into consideration that the generated material must account for GTTM’s rules. In this way, the branching connections between PB’s chords, those on which Section 6.2.3 is based, can be inferred by listeners.

To generate melodic lines against PB’s Dominating Chords (i.e. First Chord, Reprise Chord, Other Chord and Normative Chord), Themer implements two main methods. One method generates new melodic material and the other transforms the newly generated material. The final theme will then consist of a melody and its transformations. This approach allows Themer to have control over the generation of GTTM structures. In this way, the melodic lines against the main chords may be perceived as the main grouping structures and their transformations may be perceived as being similar material, which will lead to symmetrical and parallel GTTM analyses.

The derivation and application of the two methods is discussed in Sections 6.2.5.1 to 6.2.5.4.

6.2.5.1 | How to generate new melodic material

In order to generate new melodic material against a given chord, Themer takes three steps. First, it generates a basic melodic structure. This structure assigns a pitch to every strong beat in the measures the given chord covers. We call this structure the melodic skeleton and its newly generated pitches, for each strong beat, the reference pitches. Second, Themer generates new pitches between the reference pitches in the melodic skeleton. We call this structure, which combines the reference pitches with newly generated
pitches between them, the final melodic line. Third, Themer generates the note values (i.e. the rhythm) of all the pitches in the final melodic line.

Let us take, for instance, the first chord in the first semi-phrase of the theme generated in Figure 6.6. This chord corresponds to the theme’s First Chord and is labelled I in the key of C major. The theme’s time-signature is $\frac{2}{4}$ and the given chord, First Chord, lasts one measure. That is to say, there is only one strong beat in First Chord’s measure. Despite the fact there is only one strong beat in the measure, Themer will generate two reference pitches. In this way, we ensure there is always going to be at least one pair of reference pitches in the melodic skeleton between which new pitches can be generated.

We constrain the first reference pitch in a melodic skeleton to be assigned the class of the root of the given chord. In this way, the first reference pitch, which will coincide with the inception of the first strong beat, may be perceived as the head of the current span, as proposed by MPR 3, TSRPRs 1 and 5, and PRPR 1. Likewise, the root of the given chord will show the closest relation to the local tonic, which is in line with TSRPR 2. We, therefore, assume that the first reference pitch will be the head of the hierarchical connections associated with First Chord.

We constrain the first reference pitch in a melodic skeleton to be assigned the first of the Possible Octaves pre-defined in Themer. By default, these are octaves 4, 5 and 6. Therefore, in the running example, where First Chord (i.e. I/C) is the given chord, the first reference pitch in the melodic skeleton will be c4. This is illustrated in Figure 6.8.

![Figure 6.8: An example of the generation of the first reference pitch in the melodic skeleton against the First Chord, I/C, in Figure 6.6.](image)

In order to assign a class to the remaining reference pitches in the melodic skeleton, Themer stochastically assigns them a class from those in the given chord. This is, again, in line with MPR 3. For instance, in the running example, where I/C is the given chord,
the second reference pitch will be assigned one of the classes in I/C’s triad. That is to say, c, e or g. Let us imagine that *Themer* stochastically assigns the second reference pitch with class g.

To assign an octave to the stochastically assigned reference pitch classes, *Themer* statistically selects (i.e. weighted random selection) one octave from its pre-defined Possible Octaves. In the above example, where g was selected as the pitch class of the second reference pitch, the possible pitches will then be: g₄, g₅ and g₆. This is illustrated in Figures 6.9a, 6.9b and 6.9c, respectively.

![Figure 6.9: An example of the generation of the second reference pitch in the melodic skeleton that was started in Figure 6.8, against the First Chord in Figure 6.6.](image)

In order to statistically select (i.e. weighted random selection) which octave will be assigned to the newly selected reference pitch classes, *Themer* assigns each octave with a weighting that inversely correlates with the number of steps between each pitch and its preceding pitch. For instance, in Figure 6.9, the possible melodic skeletons are ordered from the most probable, at the top (i.e. Figure 6.9a), to the least probable, at the bottom (i.e. Figure 6.9c). Let us imagine that, in the running example, *Themer* generates the most probable as the final melodic skeleton; that is Figure 6.9a.

In the above example, the melodic skeleton consists of one pair of reference pitches. However, a melodic skeleton could consist of more than one pair of reference pitches, if the given chord lasts more than one measure, and so there is more than one strong beats.

In order to complete the generation of new melodic material, *Themer* continues by generating new pitches between pairs of reference pitches in the melodic skeleton. Given a pair of reference pitches, such as that in Figure 6.9a, *Themer* starts by statistically
selecting (i.e. weighted random selection) one new pitch to fit within the pair. To do so, *Themer* weights the pitches in the current diatonic scale. Each pitch in the scale is assigned a weighting, \( w = w_l + w_r \), where \( w_l \) refers to a weighting related to the left reference pitch in the pair and \( w_r \) relates to the right reference pitch in the pair. \( w_l \) and \( w_r \) inversely correlate with the distance from the left and right reference pitches to the pitches in the diatonic scale, respectively. This is illustrated in Figure 6.10. In this figure, the notes in C major's diatonic scale are associated with two weightings. One weighting, shown in Figure 6.10a, corresponds to the values of \( w_l \) when the left reference pitch is c4, as in Figure 6.9a. The other weighting, shown in Figure 6.10b, corresponds to the values of \( w_r \) when the right reference pitch is g4, as in Figure 6.9a. For the sake of conciseness, Figure 6.10 only includes a one-octave range diatonic scale. *Themer*, however, calculates the corresponding weightings associated with the three octaves pre-defined in its Possible Octaves.

Let us imagine that *Themer* statistically selects pitch b4 to fit within the given melodic skeleton (i.e. pair of reference pitches) in Figure 6.9a. This is illustrated in Figure 6.11.

![Figure 6.10: An example of the weightings, (a) \( w_l \) and (b) \( w_r \), associated with the notes in C major's diatonic scale when given the pair of reference pitches in Figure 6.9a.](image)

![Figure 6.11: An example of the generation of one new pitch, b4, to fit the melodic skeleton in Figure 6.9a.](image)

We constrain the generation process according to two additional rules. First, any intervallic jumps greater than a third, between newly generated pitches, will not be followed by jumps greater than a fifth. Second, non-chordal notes (i.e. passing note,
flourishes, suspensions and appoggiaturas) must be correctly treated. We have followed Piston’s (1959) approach to implement the correct treatment of non-chordal notes. According to Piston, intervals between non-chordal pitches must not be greater than a second.

The above two rules are applied by an appropriate manipulation of the weightings $w_l$ and $w_r$. Let us imagine that the next pitch generated in Figure 6.11, considering the two rules, is $a_4$. This is illustrated in Figure 6.12.

Figure 6.12: An example of the generation of another new pitch, $a_4$, to fit Figure 6.11.

As mentioned in the previous paragraph, the above generation process, illustrated in Figure 6.12, is repeated several times until the resulting melodic line is considered a valid final melodic line. In order to determine whether a melodic line is valid, it is checked every time a new pitch has been generated. To do so, we have implemented a collection of rules. First, the number of pitches in the melodic line must be within a specific range. We have defined this range according to the input character. This is shown in Table 6.3. This table includes the possible range of number of pitches expressed in terms of the minimum number of pitches allowed per measure, $r_1$, the maximum number of pitches allowed per measure, $r_2$, the number of beats per measure, $N$, and the number of units of subdivision per beat, $S$ (i.e. $S = 2$ for time-signatures with binary subdivision; $S = 3$ otherwise).

Table 6.3: Allowed ranges of pitches pre-defined in *Themer* according to the input character.

<table>
<thead>
<tr>
<th>Character</th>
<th>$r_1$</th>
<th>$r_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>adagio</td>
<td>1</td>
<td>N</td>
</tr>
<tr>
<td>andante</td>
<td>1</td>
<td>S·N</td>
</tr>
<tr>
<td>allegro</td>
<td>$N$</td>
<td>$2·S·N$</td>
</tr>
<tr>
<td>presto</td>
<td>$S·N$</td>
<td>$2·S·N$</td>
</tr>
</tbody>
</table>

If the number of pitches is less than $r_1$, the generation of new pitches should continue. Otherwise, if the number of pitches is greater than $r_2$, the melodic line will be re-written in its original form (i.e. the melodic skeleton), and the generation will start from scratch.
Chapter 6. The Music Generator

Themer: the generator of themes

Only when the number of pitches in a melodic line is within the range \([r_1, r_2]\), it could be considered valid. In this case, it will be valid based on two conditions. First, if the pitch generated last is equal to the last reference pitch, the melodic line will be considered valid (and the last reference pitch will be removed so that the line does not finish with a repeated pitch). Second, if the number of pitches in the melodic line is equal to \(r_2\) and the interval between the last generated pitch and the last reference pitch is less than or equal to a fifth interval, the melodic line will be considered valid.

For instance, in the melodic line in Figure 6.12 the time-signature is \(\frac{2}{4}\). This time-signature consists of two beats per measure and two units of subdivision per beat. Therefore, according to Table 6.3, \(N = 2\) and \(S = 2\). Since the melodic line is based on the theme previously shown in Figure 6.6, where the input character was allegro, the allowed range of notes in the melodic line, according to Table 6.3 will be: \([r_1, r_2] = [N, 2 \cdot S \cdot N] = [2, 8]\). Notice that the number of pitches in the melodic line in Figure 6.12, 4, is within the range \([2, 8]\). Let us imagine that a new pitch is generated equal to the last reference pitch, \(g_4\). According to the conditions above, the melodic line in Figure 6.12 will be considered valid.

In addition to the above conditions, Themer also checks whether the first pitch in the final melodic line is the highest one. When that is the case, the final melodic line is re-written in its original form (i.e. the melodic skeleton), and the generation must start from scratch, so that the generated final melodic line is in line with TSRPR 3.

The above processes are repeated one hundred times. This results in one hundred possibilities for each melodic skeleton filled with new intermediate pitches. To select the one that will be included in the final melodic line, each of the one hundred skeletons is assigned a weighting, \(w = w_v + w_s + w_j + w_d\). \(w_v\), which concerns pitch variability, is directly proportional to the number of pitches that are different in each line. \(w_s\), which concerns tonal stability, is directly proportional to the sum of the weightings that each pitch in each line is assigned in TPS’s basic space. \(w_j\), which concerns intervallic jumps between pitches, is directly proportional to the number of steps between the pitches in each line. \(w_d\), which concerns pitch density, is equal to one, if the number of pitches in a
line is equal to the number of pitches in the immediately preceding line, zero otherwise.\footnote{In the case of the first line in the final melodic line, $w_d$ is made equal to zero.}

The implementation of the above weightings is motivated by different reasons. $w_v$ rewards pitch variability so as to avoid excessive repetition. $w_s$ rewards tonal stability, with regards to TPS’s basic space, so as to support a stable organisation that matches GTTM’s intuitions. $w_j$ rewards intervallic jumps between pitches so as to avoid an excessive amount of conjunct degrees, which may lack musicality. $w_d$ rewards identical pitch density between consecutive melodies so as to support the generation of parallel GTTM analyses.

Based on the above weightings, the final melodic line is generated by statistically selecting (i.e. weighted random selection) its constituent lines (i.e. filled melodic skeletons).

The newly generated final melodic line is then transformed into new melodic material by assigning its pitches with a note value (i.e. assigning rhythm). We have implemented a preliminary palette of note values in Themer that includes: a sixteenth-note, an eighth-note, a dotted eighth-note, a quarter-note, a dotted quarter-note, a half-note, a dotted half-note, a whole-note, a dotted whole-note and two connected whole-notes. The palette is shown in Figure 6.13.

![Original palette of note values implemented in Themer.](image)

Figure 6.13: Original palette of note values implemented in Themer.

Based on the number of notes in the generated final melodic line, Themer calculates a new palette of possible note values. This palette is constrained by the maximum and minimum note values that a note could have in the new melodic material. We have defined the maximum note value as the duration that a note would have, against the duration of a given chord from the newly generated melodic material, if the rest of the notes in the given chord’s measure were sixteenth-notes, which is the smallest note value that Themer recognises. We have defined the minimum note value as the duration that all notes would have if they all lasted the same. For instance, in the melodic line in
Figure 6.12, if the three last notes were sixteenth notes, the first note's value would be a quarter-note connected to a sixteenth-note. Since this note value does not exist in *Themer*’s original palette of note values, we assign as the maximum note value the closest note value. In Figure 6.13, this is a quarter-note. If all four notes have the same value, they would need to be eighth-notes. Therefore, given the melodic line in Figure 6.12, the maximum and minimum note values calculated by *Themer* will be a quarter-note and an eighth-note, respectively. Thus, the new palette of possible note values from which to assign to the given final melodic line will consist of the note values in Figure 6.13 between the quarter-note and the eighth-note, both inclusive.

*Themer* starts by generating the note value of the first pitch in the final melodic line. This note value is statistically selected (i.e. weighted random selection) from the palette of possible note values calculated above. The weightings are directly proportional to the duration of each note value in seconds. That is to say, the longer a note value’s duration is, the higher is its probability of being selected. Let us imagine that the first pitch in the final melodic line in Figure 6.12 is assigned a dotted eighth-note. This is illustrated in Figure 6.14.

![Figure 6.14: An example of the assignment of the note value to the first pitch in Figure 6.12.](image)

To calculate the note value of the remaining pitches, the palette of possible note values must be calculated again. That is because now that the first pitch has been assigned a note value, the maximum and/or minimum note values may have changed. These are calculated again, in the same way it was discussed above, and so it is the palette of possible note values. By re-defining these parameters, the first note in the new melodic material will have the longest note value in the span. This would help to perceive this note as the head of the span, as intended by *Themer* in line with GTTM’s intuitions.

Finally, *Themer* statistically selects (i.e. weighted random selection), one by one, the
note value of the remaining pitches in the final melodic line. To do so, the pitches are assigned a weighting directly proportional to their degree of stability according to TPS’s basic space. In this way, the more stable a pitch is, the higher it is the probability of it being assigned a long note value, and vice-versa. Likewise, the note values in the palette of possible note values are assigned a weighting directly proportional to how well they would complete the current beat. For instance, in Figure 6.14, pitch b4 is the least stable in this case, would complete the first beat. Therefore, let us imagine that b4 is assigned a sixteenth-note. The remaining pitches, a4 and g4, could have different combinations of note values. The last pitch is more stable than the preceding one, so it should not be shorter. Let us imagine that they are both assigned an eighth-note. This is illustrated in Figure 6.15.

![Figure 6.15: An example of the assignment of the remaining note values in Figure 6.14.](image)

The implementation of the above weightings is motivated by different reasons. The first weighting rewards a direct relation between a note’s value and its tonal stability, so that a more metrically stable rhythm is generated. The second weighting rewards the completeness of beats so that a stable organisation of the beats is achieved.

### 6.2.5.2 | Generating new melodic material against PB’s Dominating Chords

The first melodic line Themer generates is that of First Chord. To do so, Themer applies the method introduced in Section 6.2.5.1 to generate new melodic material against First Chord.

The next melodic line Themer generates is that of Reprise Chord, if it exists in PB. To do so, Themer randomly decides whether Reprise Chord’s melodic line should be new melodic material or a transformation of First Chord’s melodic line. It could be argued that Reprise Chord’s melodic line should not be new melodic material so that it actually behaves as a reprise of the first chord in the theme. However, Themer allows the
generation of new melodic material against Reprise Chord in order to avoid excessive repetition.

The next melodic line Themer generates is that of Other Chord, if it exists in PB. To do so, Themer repeats the method discussed in Section 6.2.5.1, so that a new melodic line is generated.

### 6.2.5.3 Generating transformed melodic material against PB’s Dominating Chords

The next melodic line Themer generates is that of Normative Chord. In order to establish some sense of similarity within PB, this melodic line is generated as a transformation of a reference melody. If Other Chord exists in PB, Themer randomly decides whether the reference melody is that of Other Chord or that of First Chord; otherwise, the reference melody will be that of First Chord.

The melodic lines of Reprise Chord and Normative Chord concern transformations of reference melodic lines. Themer considers four possible transformation techniques, which are illustrated in Figure 6.16. The first phrase of the theme in Mozart’s *Ah vous dirai-je, Maman*, K. 265/300e, is shown in Figure 6.16a as a reference theme. The first transformation technique implemented in Themer is the *transposition* of a reference theme, which involves moving the reference theme up or down by a specific interval. This is shown in Figure 6.16b, where Mozart’s theme is transposed down a fourth. The second transformation technique implemented in Themer is the *inversion* of a reference theme, which involves moving each interval in the reference theme the same number of whole-tones and semitones but in the opposite direction. This is shown in Figure 6.16c, where all intervals in Mozart's theme are inverted. The third transformation technique implemented in Themer is the *retrogression* of a reference theme, which concerns re-writing the reference theme from right to left. This is shown in Figure 6.16d, where Mozart’s theme is re-written from right to left. The fourth and final transformation technique implemented in Themer is the *inverted retrogression*, which concerns re-writing the inverted version of the reference theme from right to left. This is shown in Figure 6.16e, where Figure 6.16c is re-written from right to left.
Figure 6.16: An example of the types of transformations implemented in Themer, including: (a) the first phrase of the theme in Mozart’s *Ah vous dirai-je, Maman*, K. 265/300e, as the reference theme, (b) the transposition of the reference theme down a fourth, (c) the inversion of the reference theme, (d) the retrogression of the reference theme and (e) the inverted retrogression of the reference theme.

From the transformations shown in Figure 6.16, the generation of Reprise Chord’s melodic line applies a transposition. In this way, it will truly be acting as a reprise of the first chord in the theme. Likewise, the generation of Normative Chord’s melodic line corresponds to the inversion of its transposed reference theme (i.e. either the melodic line against Other Chord or First Chord). Themer generates the melodic line against Normative Chord as an inversion in order to give some sense of closure to PB’s semi-phrase; that is to say, if the reference theme is an ascending or a descending melodic line, the melodic line against Normative Chord would be the opposite, a descending or an ascending line, respectively, which may convey some sort of closure to PB.

6.2.5.4 | Generating melodic material against PB’s Subordinate Chords

The next melodic lines Themer generates are those of PB’s Subordinate Chords (i.e. Left Subother Chord, Right Subother Chord, Left Subnormative Chord and Right Subnormative Chord), if they exist in PB.

The Subordinate Chords will be the only chords in the theme to include rests. This
will help listeners perceive the need of Subordinate Chords to attach to a more important event in the theme’s hierarchy, so they can be truly perceived as subordinates.

In left-branching Subordinate Chords, the rest will be generated on the first beat in the span. In this way, the generated melodic line will be perceived as less metrically stable. Against left-branching subordinates, the melodic line will consist of an ascending collection of conjunct notes, in the current diatonic scale, starting at a chordal note. In this way, the generated melodic line will be perceived as not having an actual identity beyond pushing the music forward towards the melodic line of the succeeding Dominating Chord. Likewise, the first note in the melodic line will be assigned the longest note value in the span. In this way, it will be easier to perceive this note as the head of the current span, as proposed by TSRPRs 1 and 6.

In right-branching Subordinate Chords, the rest will be generated on either the last beat or the last two beats in the span. In this way, the generated melodic line will be perceived as a continuation of the melodic line of the corresponding Dominating Chord and not as an independent melodic line. Against right-branching subordinates, the melodic line will consist of a descending collection of conjunct notes, in the current diatonic scale, finishing at a chordal note. In this way, the generated melodic line will be perceived as not having an actual identity beyond resolving the melodic line of the preceding Dominating Chord. Likewise, the last note in the melodic line will be assigned the longest note value in the span. In this way, it will be easier to perceive this note as the head of the current span.

6.2.5.5 | An example of the application of Themer’s Step 5

To illustrate the application of Themer’s Step 5, Figure 6.17 shows the application of the methods discussed in Sections 6.2.5.1 to 6.2.5.4 to the theme’s template previously generated in Figure 6.6.

Figure 6.17a includes PB’s structure, borrowed from Figure 6.5. Figure 6.17b illustrates the generation of a new melodic line against, I, the First Chord. This newly generated melodic line has been generated following the method discussed in Section 6.2.5.1.
Chapter 6. The Music Generator

Themer: the generator of themes

Figure 6.17: An example of the application of Themer’s Step 5 including: (a) PB’s structure from Figure 6.6, (b) the generation of a new melodic line against First Chord, (c) the transposition of First Chord’s melodic line against Reprise Chord, (d) the inversion of First Chord’s melodic line against Normative Chord and (e) the generation of a new melodic line against Left Subnormative Chord.

Figure 6.17c illustrates the transposition of the melodic material generated in Figure 6.17b into vi, the Reprise Chord. Figure 6.17d illustrates the inversion of the melodic material generated in Figure 6.17b against the same chord, I, the Normative Chord. Finally, Figure 6.17e illustrates the generation of one of many possible melodic lines that start with a one-beat rest and lead, from a chordal note, to the first note in Normative Chord’s melodic line.

6.2.6 | Step 6: generating a melody to fit the theme’s second semi-phrase

The sixth step in the implementation of Themer is to generate a melody to fit CS’s chord labels. To do so, Themer implements a statistical rule-based approach taking into consideration that the generated material must account for GTTM’s rules. In this way, the branching connections between CS’s chords, those on which Section 6.2.4 is based, would be easily inferred by the listeners. The derivation of the rules in the approach is discussed
in Sections 6.2.6.1 and 6.2.6.2.

6.2.6.1 | Generating melodic material against CS’s Dominating Chords

The first melodic line Theme creates is that of Last Chord. To do so, Theme stochastically decides whether the melodic line generated against Last Chord will be new melodic material or the repetition of a reference theme. When Theme follows the former approach, it generates a single chordal note\(^9\) that will last Last Chord’s whole span. When Theme follows the latter approach, Last Chord adopts First Chord’s melodic line.

The next melodic line Theme creates is that of Dominant Chord. To do so, Theme transposes First Chord’s melodic line and then generates its inversion. In this way, as with Normative Chord’s melodic line, Theme aims to give some sense of closure to CS’s final cadence. That is to say, if Last Chord’s melody is an ascending or a descending melodic line, the melodic line against Dominant Chord would be the opposite, a descending or an ascending line, respectively, which would convey some sort of closure to CS.

The next melodic line Theme creates is that of Cadential\(^6\), if it exists in CS. To do so, Theme applies the process already discussed in Section 6.2.5.4 concerning the generation of melodic material against left-branching subordinate chords.

6.2.6.2 | Generating melodic material against CS’s Subordinate Chords

The next melodic line Theme creates is that of Subdominant Chord, if it exists in CS. To do so, Theme stochastically selects a reference theme, among those upon First Chord, Reprise Chord and Other Chord, whenever the last two chords exist in PB, and transposes it.

The next melodic lines Theme creates are those of Sub-Subdominant Chords, if they exist in CS. For each existing Sub-Subdominant Chord, Theme stochastically selects a type of transformation, either inversion, retrogression or inverted retrogression, and applies it to a transposed version of Subdominant Chord’s melodic line.

\(^9\)This note will be the first note in First Chord’s melodic line.
6.2.6.3 | An example of the application of Themer’s Step 6

To illustrate the application of Themer’s Step 6, Figure 6.18 shows the application of the methods discussed in Sections 6.2.6.1 and 6.2.6.2 to the theme’s template previously generated in Figure 6.7.

Figure 6.18: An example of the application of Themer’s Step 6 including: (a) CS structure from Figure 6.7, (b) the repetition of First Chord’s melodic line against Last Chord, (c) the inversion of First Chord’s melodic line against Dominant Chord, (d) the transposition of First Chord’s melodic line against Subdominant Chord and (e) the retrogression of Subdominant Chord’s melodic line against Sub1 Subdominant Chord.

Figure 6.18a includes CS structure, borrowed from Figure 6.5. Figure 6.18b illustrates the assignment of First Chord’s melodic line (i.e. that in Figure 6.17b) against Last Chord. Figure 6.18c illustrates the transposed inversion of First Chord’s melodic line against Dominant Chord. Figure 6.18d illustrates transposition of First Chord’s melodic line against Subdominant Chord. Finally, Figure 6.18e illustrates the transposed retrogression of the melodic line against the theme’s Subdominant Chord into its Sub1 Subdominant Chord.
6.2.7 | Step 7: generating the theme’s representations

The seventh and final step in the implementation of Themer is to store representations of the theme. To do so, Themer generates: the theme’s prolongational matrix, which represents the theme’s prolongational reduction; the theme’s melodic data, which include the theme’s key label, the theme’s input tempo in BPM, the theme’s melody as lists of pitches per span, the theme’s melodic rhythm as lists of durations per chord and the theme’s chord, from those in Figure 6.4, per measure; and the theme’s harmonic data, which include the theme’s chord labels by time-span and the contribution to tension of MTT’s surface parameters (i.e. scale degree, sc.dg.; inversion, inv.; and non-harmonic tones, nh.t.).

6.3 | Arranger

The generator of arrangements

The themes generated by Themer consists of a melody and a sequence of chord labels. However, the way the notes in the chords are organised is not defined yet. In order to organise them, autognomus incorporates the Arranger.

There exist many different ways in which the notes in a theme’s chords could be organised. For instance, another melody could be generated, which, while matching the chord labels, interacted with the theme’s original melody in a contrapuntal style. Despite the potential this idea may have, it should be noted that GTTM was designed to only work upon homophonic music, as pointed out by Lerdahl and Jackendoff (1983, p.37):

we are treating all music as essentially homophonic (...); [f]or the more contrapuntal varieties of tonal music, where this condition does not obtain, our theory [i.e. GTTM] is inadequate.

Thus, Arranger must follow a different strategy.

Given a melody and a sequence of chords, an easy way to combine them may be in the form of an accompanied melody (i.e. a melody and its accompaniment); that is
to say, the chords get subordinated to the melody and they just support it. This is the
compositional strategy followed by Arranger.

Although the accompaniment is subordinated to the melody, the chords in an accom-
panied melody can still be arranged in different ways. Common arrangement techniques
include unfolding the chords through arpeggios or splitting the chords into chunks, as in,
for instance, the waltz dance style. For the sake of simplicity, Arranger will arrange the
notes in the theme’s chords in blocks.

Arranger implements the generation of arrangements in two steps. These are dis-
cussed in Sections 6.3.1 and 6.3.2.

6.3.1 | Step 1: generating a theme’s arrangement

The first step in the implementation of Arranger is to generate the arrangement of the
voices of the chords in a given theme in S.A.T.B. form (i.e. Soprano, Alto, Tenor and
Bass). To do so, Arranger implements a rule-based approach. The derivation of the rules
in the approach is discussed in Sections 6.3.1.1 and 6.3.1.2.

6.3.1.1 | Generating the arrangement of the theme’s First Chord

The first arrangement Arranger generates is that of First Chord. To do so, Arranger
incorporates a pre-defined arrangement where the soprano is assigned the fifth of First
Chord in octave 4, the alto is assigned the third of First Chord in octave 4, the tenor
is assigned the fifth of First Chord in octave 3 and the bass is assigned the tonic of
First Chord in octave 3. If the bass happens to be higher than the tenor, the bass is
re-assigned the tonic of First Chord in octave 2.

6.3.1.2 | Generating the arrangement of the remaining chords

Arranger then generates the arrangement of the remaining chords, one by one. To do
so, Arranger takes two steps. First, it generates a bass line for the whole theme. The
main Dominating Chords from GTTM’s basic form (i.e. First Chord, Normative Chord,
Dominant Chord and Last Chord) are generated in root position; Cadential\textsuperscript{6}, if it exists
in the theme, is generated in second inversion; and the rest of the chords are generated in first inversion. In this way, the main Dominating Chords will be perceived as the most stable, whereas the rest of the chords, which include all Subordinate Chords, will be perceived as less stable. In this way, it will be easier to perceive the intended prolongational relations (i.e. hierarchical structure) in the theme.

Second, based on the chords’ triadic notes, Arranger calculates all possible S.A.T.B. permutations that contain the corresponding bass note in the previously generated fixed bass line as candidate arrangements. From these candidates arrangements, Arranger will consider as valid arrangements those that meet the following constraints:

- there are no parallel fifths and octaves,
- there are no distances greater than an octave between consecutive S.A.T.B. voices, and
- S.A.T.B. voices do not cross each other.

If no arrangement meets the above constraints, all candidate arrangements are directly defined as being valid arrangements.

Finally, Arranger generates the final arrangement as that, from the collection of valid arrangements, whose sum of distances within each voice is the shortest. In this way, the arrangements will be in line with common voice-leading conventions (Piston, 1959).

6.3.1.3 | An example of the application of Arranger’s Step 1

To illustrate the application of Arranger’s Step 1, Figure 6.18 shows the application\(^{10}\) of the methods discussed in Sections 6.3.1.1 and 6.3.1.2 to the theme formed by the two semi-phrases previously generated as an example in Figures 6.17 and 6.18.

\(^{10}\) Notice that Arranger has allowed an interval greater than an octave between the alto and the tenor so that the seventh of the dominant chord resolves in a descending manner.
Chapter 6. The Music Generator  

*Developer*: the generator of developments

6.3.2 | Step 2: generating the arrangement’s representation

The second and final step in the implementation of *Arranger* is to store the newly generated arrangement.

6.4 | *Developer*

The generator of developments

In order for the music generated by *Themer* to not be boring, new thematic material needs to be generated, and so *autognomus* incorporates the *Developer*.

The newly generated material must still convey some sense of long-term structure. To do so, given a theme generated with *Themer*, *Developer* generates new harmonic and melodic material which will be incorporated into the theme.

When adding the new material to the original theme, we would like to keep as many hierarchical relations as possible from the theme’s prolongational reduction. In this way, it will be easier to sense long-term structure when listening to the theme followed by its developments. In order to keep these relations, the new chords generated by the *Developer* should be generated as *Subordinate Chords*. If, on the contrary, the new chords were generated as *Dominating Chords*, they would alter the branching relations in the theme’s hierarchical structure.

Figure 6.20 shows an example of one possible way to add new *Subordinate Chords* (shown as dotted lines) to the hierarchical structure previously shown in Figure 6.3.
Notice, however, that the example shown in Figure 6.20 is just one of the many possible ways in which new Subordinate Chords can be added to Figure 6.3.

In order to narrow down the number of possible new trees that can be derived from an original tree representation, we impose the following constraint: each branch in the original tree representation will only include one new Subordinate Chord; that is to say, one new (dotted) branch. For the sake of simplicity, we also constrain the melodic material of the newly generated Subordinate Chords to be generated as transformations (i.e. inversions, retrogressions or inverted retrogressions) of the melodic material of the given theme.

Developer implements the above constraints in two steps. These are discussed in Sections 6.4.1 and 6.4.2.

6.4.1 | Step 1: generating a theme's development

The first step in the implementation of Developer is to generate new Subordinate Chords, each of which will attach to an existing chord in a given theme. To do so, Developer implements a rule-based approach. The implementation of the approach is discussed in Sections 6.4.1.1 and 6.4.1.2.

6.4.1.1 | Generating the development's harmonic structure

As discussed in Section 6.4, given a theme generated by Themer, Developer will generate new material by only generating one Subordinate Chord per branch in the prolongational tree of the original theme. The new Subordinate Chords would attach to
their respective Dominating Chord through either right or left branches. Again, this means that there are lots of possible new trees that can be derived from a single original prolongational tree. For instance, in the case of a six-branch original theme, as that in Figure 6.3, there are 64 possible ways\(^{11}\) in which right and left branches can be attached to the original six branches. In order to have more control over the generation of Subordinate Chords, Developer is constrained by an additional rule: the branches of all Subordinate Chords must follow the same direction, either right or left. That is to say, we have reduced the possible new prolongational trees to two options: one tree where all Subordinate Chords are right-branching subordinates and another one where they are all left-branching subordinates.

The second constraint entails, however, some limitations concerning the first and last chords in the theme (i.e. First Chord and Last Chord). These chords are supposed to be the most stable in the theme, as proposed by GTTM’s basic form and normative structure. Therefore, if First Chord is preceded by a left-branching subordinate, this will result in a less stable structure and so the intended hierarchical representation might not match the one inferred by the listeners. The same would happen if Last Chord is followed by a right-branching subordinate. Therefore, we modify the second constraint as follows: the branches of all Subordinate Chords must follow the same direction, either right or left, except for the First Chord and the Last Chord, whose Subordinate Chords must always be right-branching and left-branching subordinates, respectively.

For instance, let Figure 6.3 be the hierarchical structure of a given theme. Based on the above constraints, the two possible new hierarchical trees will be those shown in Figure 6.21.

To generate the new development, Developer interprets the chord labels from the original theme (i.e. the solid lines in Figure 6.21) as Dominating Chords and generates the chord labels of the new branches in the newly generated hierarchical trees (i.e. the dotted lines in Figure 6.21) as Subordinate Chords. To do so, Developer applies Rule 6.5. This rule was previously introduced in Section 6.2.3.3 as Themer’s rule for the generation

\(^{11}\)Let us calculate the total number of permutations considering repetitions. There are two possible directions, right and left, and six branches. Therefore, the number of permutation of right and left directions considering six branches is equal to \(2^6 = 64\).
Chapter 6. The Music Generator

Developer: the generator of developments

(a) Left-branching subordinates.

(b) Right-branching subordinates.

Figure 6.21: The two possible new hierarchical trees that Developer calculates to generate the development of a theme whose hierarchical structure is that of Figure 6.3. These trees incorporate one new Subordinate Chord (dotted line) per branch, either left or right, except for the first and last chords in the tree, whose branching directions are fixed.

of Subordinate Chords. Recall that, in this rule, given a Dominating Chord, a list of candidate subordinates is generated consisting of the three closest chords to Dominating Chord in the diatonic circle-of-fifths. Recall as well that if Subordinate Chord is right-branched, the closest chords in circle-of-fifths are calculated from Dominating Chord to its right; otherwise, they are calculated to its left. Finally, if the TPS distance from the candidate chords to Dominating Chord is equal to or less than any other distance to the rest of the surrounding dominating chords, the candidate chords are considered as valid subordinates. By applying Rule 6.5, one of the valid subordinates is stochastically selected as the new Subordinate Chord.
6.4.1.2 | Generating the development’s melodic material

There are multiple ways in which the melodic material of the newly generated Subordinate Chords can be generated. For the sake of simplicity, Developer uses the thematic transformations previously introduced in Figure 6.16. Developer applies these to the melodic material of the given theme.

From the possible thematic transformations, Developer only uses the inversion, the retrogression and the inverted retrogression. The transposition is not used in order to avoid excessive repetition.

We have constrained Developer to use the same type of transformation for all the newly generated chords. In this way, the transformation of the original theme’s melodic material will be more coherent and so it will be easier to convey some sense of long-term structure.

To generate the melodic material against the newly generated chord labels, Developer transforms the melodic material of their respective Dominating Chord. And those spans that correspond to the original theme will keep their original melodic material.

For instance, let Figure 6.3 be the theme given as input. Developer will then generate two hierarchical new trees. These will be those in Figure 6.21. When Developer uses, for example, the inversion transformation upon Figure 6.21a, the resulting development will be that shown in Table 6.4.

Notice that, by applying the same type of transformation to all of the newly generated Subordinate Chords, Developer can generate three different developments per newly generated hierarchical tree. For instance, if Figure 6.3 was the input original theme and Figures 6.21a and 6.21b were the two newly generated hierarchical trees, Developer would generate the following developments:

(1) Figure 6.21a where the new melodies are inversions of the original ones,

(2) Figure 6.21a where the new melodies are retrogressions of the original ones,

(3) Figure 6.21a where the new melodies are inverted retrogressions of the original ones,
(4) Figure 6.21b where the new melodies are inversions of the original ones,

(5) Figure 6.21b where the new melodies are retrogressions of the original ones, and

(6) Figure 6.21b where the new melodies are inverted retrogressions of the original ones.

In conclusion, given a theme generated with Themer, Developer will generate six independent developments from it. To generate the arrangement of the newly generated developments, Arranger should be applied once again.

6.4.1.3 | An example of the application of Developer’s Step 1

To illustrate the application of Developer’s Step 1, Figure 6.22 shows the generation of one type of development based on the reference theme whose semi-phrases were previously generated in Figures 6.17 and 6.18. From the six possible types of developments that Developer considers, Figure 6.22 illustrates the generation of the fifth one from those

Table 6.4: Example of the development that Developer will generate when Figure 6.3 is the structure of the input given theme, when Subordinate Chords are left-branching subordinates (i.e. those in Figure 6.21a) and when the transformation applied is the inversion.

<table>
<thead>
<tr>
<th>Event</th>
<th>Chord label</th>
<th>Melodic material</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>original chord label of event 1 in theme</td>
<td>original melody of event 1 in theme</td>
</tr>
<tr>
<td>2</td>
<td>new chord generated using Rule 6.5 (subordinated to event 1)</td>
<td>inversion of the original melody of event 1</td>
</tr>
<tr>
<td>3</td>
<td>new chord generated using Rule 6.5 (subordinated to event 4)</td>
<td>inversion of the original melody of event 4</td>
</tr>
<tr>
<td>4</td>
<td>original chord label of event 4 in theme</td>
<td>original melody of event 4 in theme</td>
</tr>
<tr>
<td>5</td>
<td>new chord generated using Rule 6.5 (subordinated to event 6)</td>
<td>inversion of the original melody of event 6</td>
</tr>
<tr>
<td>6</td>
<td>original chord label of event 6 in theme</td>
<td>original melody of event 6 in theme</td>
</tr>
<tr>
<td>7</td>
<td>new chord generated using Rule 6.5 (subordinated to event 8)</td>
<td>inversion of the original melody of event 8</td>
</tr>
<tr>
<td>8</td>
<td>original chord label of event 8 in theme</td>
<td>original melody of event 8 in theme</td>
</tr>
<tr>
<td>9</td>
<td>new chord generated using Rule 6.5 (subordinated to event 10)</td>
<td>inversion of the original melody of event 10</td>
</tr>
<tr>
<td>10</td>
<td>original chord label of event 10 in theme</td>
<td>original melody of event 10 in theme</td>
</tr>
<tr>
<td>11</td>
<td>new chord generated using Rule 6.5 (subordinated to event 12)</td>
<td>inversion of the original melody of event 12</td>
</tr>
<tr>
<td>12</td>
<td>original chord label of event 12 in theme</td>
<td>original melody of event 12 in theme</td>
</tr>
</tbody>
</table>
introduced in Section 6.4.1.2; that is to say, the generation consists of mostly new right-branching subordinates and retrogressions of the theme’s melodic line.

Figure 6.22: An example of the application of Developer’s Step 1 including: (a) PB’s chords from Figure 6.17, (b) the generation of new right-branching Subordinate Chords (bold) according to Rule 6.5 (c) the generation of melodic lines against the newly generated Subordinate Chords as retrogressions of those in (b), (d) CS’s chords from Figure 6.18, (e) the generation of new right-branching Subordinate Chords (bold) according to Rule 6.5 and (f) the generation of melodic lines against the newly generated Subordinate Chords as retrogressions of those in (c).

Figure 6.22a includes PB’s chords, borrowed from Figure 6.17. For each chord in PB, Figure 6.22b generates a new right-branching Subordinate Chord, shown in bold, which is placed to the right of its respective Dominating Chord. The melodic line of each Subordinate Chord is generated in Figure 6.22c by applying the retrogression technique to the melodic line of their respective Dominating Chord. Figure 6.22d includes CS’s chords, borrowed from Figure 6.18. For each chord in CS, except for Last Chord, Figure 6.22e generates a new right-branching Subordinate Chord, shown in bold, which is placed to the right of its respective Dominating Chord. Last Chord’s new subordinate is placed to its left. The melodic line of each Subordinate Chord is generated in Figure 6.22f by applying the retrogression technique to the melodic line of their respective
Dominating Chord.

Figure 6.22b illustrates the generation of new Subordinate Chords in PB. For each chord in Figure 6.22a, three new chords will be candidate subordinates according to Rule 6.5. To decide whether these candidate chords are valid subordinates, the following values of TPS distance must be calculated:

- \( d_{d\rightarrow c} = \delta(\text{Dominating Chord} \rightarrow \text{candidate chord}) \),
- \( d_{f\rightarrow c} = \delta(\text{First Chord} \rightarrow \text{candidate chord}) \),
- \( d_{n\rightarrow c} = \delta(\text{Normative Chord} \rightarrow \text{candidate chord}) \), and
- \( d_{r\rightarrow c} = \delta(\text{Reprise Chord} \rightarrow \text{candidate chord}) \).

Both First Chord and Normative Chord are equal to I in Figure 6.22a. According to Rule 6.5, when these chords are the Dominating Chord, the calculations will be as follows:

<table>
<thead>
<tr>
<th>Candidate = V</th>
<th>Candidate = ii</th>
<th>Candidate = vi</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d_{d\rightarrow c} = 5 )</td>
<td>( d_{d\rightarrow c} = 8 )</td>
<td>( d_{d\rightarrow c} = 7 )</td>
</tr>
<tr>
<td>( d_{f\rightarrow c} = 5 )</td>
<td>( d_{f\rightarrow c} = 8 )</td>
<td>( d_{f\rightarrow c} = 7 )</td>
</tr>
<tr>
<td>( d_{n\rightarrow c} = 5 )</td>
<td>( d_{n\rightarrow c} = 8 )</td>
<td>( d_{n\rightarrow c} = 7 )</td>
</tr>
<tr>
<td>( d_{r\rightarrow c} = 8 )</td>
<td>( d_{r\rightarrow c} = 5 )</td>
<td>( d_{r\rightarrow c} = 0 )</td>
</tr>
</tbody>
</table>

Reprise Chord is equal to vi in Figure 6.22a. According to Rule 6.5, when this chord is the Dominating Chord, the calculations will be as follows:

<table>
<thead>
<tr>
<th>Candidate = iii</th>
<th>Candidate = vii(^0)</th>
<th>Candidate = IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d_{d\rightarrow c} = 5 )</td>
<td>( d_{d\rightarrow c} = 8 )</td>
<td>( d_{d\rightarrow c} = 7 )</td>
</tr>
<tr>
<td>( d_{f\rightarrow c} = 7 )</td>
<td>( d_{f\rightarrow c} = 8 )</td>
<td>( d_{f\rightarrow c} = 5 )</td>
</tr>
<tr>
<td>( d_{n\rightarrow c} = 7 )</td>
<td>( d_{n\rightarrow c} = 8 )</td>
<td>( d_{n\rightarrow c} = 5 )</td>
</tr>
<tr>
<td>( d_{r\rightarrow c} = 5 )</td>
<td>( d_{r\rightarrow c} = 8 )</td>
<td>( d_{r\rightarrow c} = 7 )</td>
</tr>
</tbody>
</table>
Left Subnormative Chord is equal to IV in Figure 6.22a. According to Rule 6.5, when this chord is the Dominating Chord, the calculations will be as follows:

<table>
<thead>
<tr>
<th>Candidate = I</th>
<th>Candidate = V</th>
<th>Candidate = ii</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_{d\rightarrow c} = 5$</td>
<td>$d_{d\rightarrow c} = 8$</td>
<td>$d_{d\rightarrow c} = 7$</td>
</tr>
<tr>
<td>$d_{f\rightarrow c} = 5$</td>
<td>$d_{f\rightarrow c} = 5$</td>
<td>$d_{f\rightarrow c} = 8$</td>
</tr>
<tr>
<td>$d_{n\rightarrow c} = 5$</td>
<td>$d_{n\rightarrow c} = 5$</td>
<td>$d_{n\rightarrow c} = 8$</td>
</tr>
<tr>
<td>$d_{r\rightarrow c} = 7$</td>
<td>$d_{r\rightarrow c} = 8$</td>
<td>$d_{r\rightarrow c} = 5$</td>
</tr>
</tbody>
</table>

For all Dominating Chords shown above, the closest chord in the diatonic circle-of-fifths in each case (i.e. the candidate in the first column in each case) always satisfies that $d_{d\rightarrow c}$ is equal to or less than the rest of the TPS distances. For the sake of simplicity, it is this chord, in each case, that is generated as new Subordinate Chord in Figure 6.22b.

Figure 6.22e illustrates the generation of new Subordinate Chords in CS. For each chord in Figure 6.22d, three new chords may be candidate subordinates according to Rule 6.5. To decide whether these candidate chords are valid subordinates, the following values of TPS distance must be calculated:

- $d_{d\rightarrow c} = \delta($Dominant Chord $\rightarrow$ candidate chord), and
- $d_{l\rightarrow c} = \delta($Dominating Chord $\rightarrow$ candidate chord).

Sub$_1$ Subdominant Chord is equal to V/ii in Figure 6.22d. According to Rule 6.5, when this chord is the Dominating Chord, the calculations will be as follows:

<table>
<thead>
<tr>
<th>Candidate = ii/ii</th>
<th>Candidate = vi/ii</th>
<th>Candidate = iii/ii</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_{d\rightarrow c} = 7$</td>
<td>$d_{d\rightarrow c} = 10$</td>
<td>$d_{d\rightarrow c} = 10$</td>
</tr>
<tr>
<td>$d_{l\rightarrow c} = 5$</td>
<td>$d_{l\rightarrow c} = 8$</td>
<td>$d_{l\rightarrow c} = 7$</td>
</tr>
</tbody>
</table>

Subdominant Chord is equal to ii in Figure 6.22d. According to Rule 6.5, when this chord is the Dominating Chord, the calculations will be as follows:

<table>
<thead>
<tr>
<th>Candidate = vi</th>
<th>Candidate = iii</th>
<th>Candidate = vii$^\circ$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_{d\rightarrow c} = 8$</td>
<td>$d_{d\rightarrow c} = 7$</td>
<td>$d_{d\rightarrow c} = 7$</td>
</tr>
<tr>
<td>$d_{l\rightarrow c} = 5$</td>
<td>$d_{l\rightarrow c} = 8$</td>
<td>$d_{l\rightarrow c} = 7$</td>
</tr>
</tbody>
</table>
Dominant Chord is equal to V\(^7\) in Figure 6.22d. According to Rule 6.5, when this chord is the Dominating Chord, the calculations will be as follows:

<table>
<thead>
<tr>
<th>Candidate</th>
<th>(d_{d\rightarrow c})</th>
<th>(d_{l\rightarrow c})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(ii)</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>(vi)</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>(iii)</td>
<td>7</td>
<td>7</td>
</tr>
</tbody>
</table>

Last Chord is equal to I in Figure 6.22d. According to Rule 6.5, when this chord is the Dominating Chord, the calculations will be as follows:

<table>
<thead>
<tr>
<th>Candidate</th>
<th>(d_{d\rightarrow c})</th>
<th>(d_{l\rightarrow c})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(IV)</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>(vii^6)</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>(iii)</td>
<td>7</td>
<td>7</td>
</tr>
</tbody>
</table>

For all Dominating Chords shown above, the closest chord in the diatonic circle-of-fifths in each case (i.e. the candidate in the first column in each case) always satisfies that \(d_{l\rightarrow c}\) is equal to or less than \(d_{d\rightarrow c}\). For the sake of simplicity, it is this chord, in each case, that is generated as new Subordinate Chord in Figure 6.22e.

The final development generated by Developer will then consist of the combination of two semi-phrases, those that are shown in Figure 6.22c and Figure 6.22f.

6.4.2 | Step 2: generating the developments’ representations

The second and final step in the implementation of the Developer is to generate the representations of the data generated concerning the theme’s developments. To do so, Developer stores the same data that Themer stored in Section 6.2.7. That is to say, it stores the developments’ prolongational matrices, their melodic data and their harmonic data.

6.5 | Morpher

The morphing system

In order for the theme and developments generated in the previous sections to match input tension profiles in real time, while still focusing on long-term structure, autognomus
incorporates the *Morpher*.

As discussed in Section 6.1.3, given a theme, generated with *Themer*, and its six developments, generated with *Developer*, *Morpher* will loop them. The theme will be played twice, followed by one of its developments, and this process will repeat ad infinitum. While the theme and its developments are being played, *Morpher* is fed input degrees of tension in real time. For instance, in the video game example in Section 6.1.3, the distance from the player to the monsters will be calculated and transformed into tension values in real time. These are the values fed into *Morpher*.

In order to contribute to the perception of long-term structure, *Morpher* will play the original material in the theme and its developments as long as the input degree of tension is below a tension threshold. Otherwise, *Morpher* will generate new material to match the input degree of tension.

*Morpher* adopts a scale of tension that relates to the possible values of TPS distance within a key. These values range between 0 and 25, as shown in Bigand et al. (1996),\(^1\) and so these are the minimum and maximum values of tension considered by *Morpher*, respectively. Considering this scale of tension, we have defined *Morpher*’s tension threshold as equal to 8. That is because the greatest value of TPS distance among diatonic triads is equal to 8. In this way, if an input degree of tension is equal to or less than 8, diatonic material is expected to be played, which recall is the context in which the theme and its developments were generated.

Despite the fact that the input degrees of tension may change in real time, *Morpher* does not always react in real time. Given a theme, *Morpher* starts playing it chord by chord. As discussed in Section 6.2.2.3, chords may last one, two or three measures in the theme, depending on how PB and CS are defined according to Tables 6.1 and 6.2, respectively. Every time a chord has been played in full, *Morpher* reads the current input degree of tension. If this input degree of tension is below the tension threshold, *Morpher* will play the next chord in the theme. Otherwise, *Morpher* will generate new melodic material, against one newly generated chord, to match the input degree of tension.

---

\(^1\)It should be noted that the values of TPS distance included in Bigand et al. (1996) were calculated using TPS outdated version (i.e. Lerdahl (1988)). In this dissertation, we have re-calculated these values using the updated version (i.e. Lerdahl and Krumhansl (2007)).
Why does *Morpher* not start generating new material as soon as the input degree of tension is greater than the tension threshold, even if this happens in the middle of a measure? If *Morpher* did so, the intended hierarchical relations may not be matched. Therefore, long-term structure may not be appropriately conveyed. The implications of this decision are discussed in more detail in Chapter 7.

*Morpher* implements the generation of new material to match input tension profiles in five steps. These are discussed in Sections 6.5.1 to 6.5.5.

**6.5.1 | Step 1: reading the input data**

The first step in the implementation of *Morpher* is to read the necessary input data. *Morpher* needs a theme's harmonic, melodic and prolongational data, taken from *Themer*’s output, as in Section 6.2.7; as well as the theme's arrangement data, taken from *Arranger*’s output, as in Section 6.3.2. If the theme has been previously developed, *Morpher* will also need the developments' harmonic, melodic, prolongational and arrangement data, taken from *Developer*’s output, as in Section 6.4.2.

**6.5.2 | Step 2: generating new harmonic material that matches input degrees of tension**

Whenever *Morpher* generates new music, that is when the input degree of tension is greater than the pre-defined tension threshold, the second step in its implementation is to generate a new chord to fit the current time-span so that it matches the input degree of tension. We have implemented three different methods to generate this chord depending on whether the input tension is increasing, remains unchanged or is decreasing. These are introduced below in Sections 6.5.2.1, 6.5.2.2 and 6.5.2.3, respectively.

**6.5.2.1 | Harmonic generation when tension is increasing**

If the input degree of tension is greater than the tension threshold and is greater than the previous degree of tension, *Morpher* generates a new chord to match the increasing
input tension.

Inspired by Prechtl (2016), we have defined a total of thirty-six different chords that Morpher could generate. These include the twelve major triads, the twelve minor triads and the twelve diatonic seventh chords\(^{13}\) than can be built upon the twelve tones in the chromatic scale.

As discussed in Section 6.4, the chords that Morpher will generate are considered Subordinate Chords (recall that, in this way, the intended hierarchical relations can be better ensured). To do so, Morpher starts by defining the chord in the previous time-span (i.e. the last span played by Morpher) as the current Dominating Chord. In this way, the chord to be generated will be a right-branching Subordinate Chord, since right branches represent increasing tension in GTTM’s prolongational reduction.

Morpher then fills a vector with the values of TPS distance from Dominating Chord to the thirty-six chords pre-defined in Morpher. As discussed in Section 3.3, each chord can have many different values of TPS distance depending on the label it is assigned. As discussed in Section 3.3.3, according to the TPS’s “principle of the shortest path”, the preferred distance between two events is the smallest one (Lerdahl, 2004, p.74). Therefore, the distance-vector includes the shortest value associated with each transition.

Morpher continues by calculating \(\Delta\) as the absolute value of the difference between the input degree of tension and the values in the distance-vector. Morpher then generates a list of candidate chord labels that consists of those whose value of \(\Delta\) is the shortest. Notice that the shorter the value of \(\Delta\), the closer the corresponding chord transitions are to the input degree of tension. Finally, Morpher stochastically selects the label of the new Subordinate Chord from the list of candidate chord labels.

6.5.2.2 | Harmonic generation when tension remains unchanged

If the input degree of tension is greater than the tension threshold and remains unchanged, Morpher generates a new Subordinate Chord to match the flat input tension.

\(^{13}\)These include the seventh chords that can be built by adding a diatonic seventh to the major and minor diatonic triads that exist in a given key.
We constraint the Subordinate Chord to also be a right-branching subordinate, as in the increasing tension case. That is because, according to Lerdahl and Krumhansl (2007, p.342), “unless schematic intuitions are strong, listeners tend to construe events in a right-branching manner”. Thus, the new Subordinate Chord is generated as in Section 6.5.2.1.

There is, however, a difference in the generation of chords in this case. Instead of matching the unchanged input degree of tension, the degree to be matched is made equal to zero. During some pilot studies of preliminary versions of Morpher we observed that, when keeping the input degree of tension as the one to be matched, listeners still experienced a rise in tension. By making this degree equal to zero, listeners’ judgements tended to correlate better with the input degrees of tension. In other words, the candidate chords are directly selected from the distance-vector, since $\Delta$ is equal to zero.

6.5.2.3 | Harmonic generation when tension is decreasing

If the input degree of tension is greater than the tension threshold and is less than the previous degree of tension, Morpher generates a new chord to match the decreasing input tension. The chord to be generated will be a left-branching Subordinate Chord, since left branches represent a tension decrease in GTTM’s prolongational reduction.

The generation of left-branching subordinates poses an implementation challenge. That is because the current Dominating Chord is not in the previous time-span but will be in upcoming spans. Morpher has no other alternative but to estimate what chord would be acting as Dominating Chord. To do so, Morpher first calculates the tension’s rate of decrease as the difference between the previous degree of input tension and the current degree of input tension. Morpher assumes that the decreasing rate, in the input tension profile, will remain the same and so predicts the degree of tension that will be associated with the upcoming spans. Morpher then identifies the first chord’s time-span where the predicted tension will be below the tension threshold. Morpher defines the chord in this span, in the original theme, as the current Dominating Chord.

Morpher continues by applying the steps described in Section 6.5.2.1 to generate
Subordinate Chords from the estimated Dominating Chord. Starting at the estimated Dominating Chord, Morpher generates a left-branching Subordinate Chord associated with the time-span preceding the Dominating Chord’s span. The newly generated Subordinate Chord is then identified as the new Dominating Chord and so a new left-branching Subordinate Chord is generated associated with the preceding span. This process is repeated until a left-branching Subordinate Chord is generated for the current span. This will result in a sequence of left-branching chords in which each chord will be dominated by its succeeding chord.

Notice that the above method predicts the best sequence of left-branching Subordinate Chords in case the tension’s rate of decrease does not change. However, the method entails a limitation. If the rate of decrease does change, the degree of MTT’s hierarchical tension, $T_{hier}$, associated with the generated Subordinate Chord may, in the end, not exactly match the input degree of tension. We believe there is no other alternative for the implementation but to perform an estimation, since we are dealing with left-branching subordinates.

6.5.2.4 | An example of the application of Morpher’s Step 2

To illustrate the application of Morpher’s Step 2, Figure 6.23 shows the generation of new Subordinate Chords to match the input tension profile in Figure 6.24.

Figure 6.23a includes the harmonic sequence of the theme previously generated in Figure 6.19. Let us imagine that this theme is fed as input into Morpher. Figure 6.24 shows a tension profile that follows a rise and fall behaviour. Let us imagine that this is the input tension profile that the music generated by Morpher has to match.

Recall that only if the input tension is greater than the tension threshold will Morpher generate new music; otherwise, it will play the music from the input theme. In Figure 6.24, measures 1 to 8 are associated with the following degrees of tension: 0, 10, 20, 16, 12, 8, 4 and 0, respectively. Since the tension threshold is equal to 8, Morpher will only generate new material to fill measures 2, 3, 4 and 5. This is shown in Figure 6.23b, where measures 1, 6, 7 and 8 have been assigned the theme’s original material in these
Figure 6.23: An example of the application of *Morpher*’s Step 2 including: (a) the theme previously generated in Figure 6.19 and (b) the newly generated chords to match the tension profile in Figure 6.24.

measures, respectively.

Figure 6.23a includes the hierarchical tree of the input theme. Since the input tension increases in measures 2 and 3 and decreases in measures 4 and 5, the branches associated with these measures must adapt to the new tension behaviour. This is shown in Figure 6.23b, where measures 2 and 3 now follow a sequence of right branches, whereas measures 4 and 5 now follow a sequence of left branches.

In measures 2 and 3, the input tension is increasing and so *Morpher* will apply the method discussed in Section 6.5.2.1. To generate a chord in measure 2, *Morpher* defines
the preceding chord (i.e. the chord in measure 1, I/C) as the current Dominating Chord. The degree of tension associated with measure 2 is equal to 10. Morpher then calculates the chord transitions, from the current Dominating Chord, whose TPS distance is closer to 10. In this case, the closer transitions will go from Dominating Chord to either D major or B♭ major. Let us imagine that Morpher stochastically selects the latter as the chord to fit measure 2. To generate a chord in measure 3, Morpher re-defines the Dominating Chord as the newly generated B♭ major chord in measure 2. Morpher then calculates the chord transitions, from the re-defined Dominating Chord, whose TPS distance is closer to 20. In this case, the closer transition will go from Dominating Chord to B major, and so let this be the chord generated by Morpher to fit measure 3.

In measures 4 and 5, the input tension is decreasing and so Morpher will apply the method discussed in Section 6.5.2.3. To generate a chord label in measure 4, Morpher calculates the rate of decrease from measure 3 to measure 4. Since the degree of tension in the former is equal to 20 and the degree of tension in the latter is equal to 16, the rate of decrease is equal to 4. Morpher assumes that this rate of decrease would be kept in upcoming measures. In this way, it predicts that the degrees of tension in the upcoming measures would be equal to 12 and 8. The last predicted tension degree, 8, which would
correspond to measure 6, is equal to the tension threshold. Therefore, Morpher defines the chord in measure 6 in the original theme (i.e. ii/I) as the current Dominating Chord. From this chord, Morpher calculates the chord transitions whose TPS distance is closer to the predicted degree of tension in measure 5 (i.e. 12). In this case, the closer transitions will go from Dominating Chord to either E♭ major, B major, C minor, B♭ or C♯ diminished. Let us imagine that Morpher stochastically selects the last one as the chord that would fit measure 5 in the current prediction. To generate the chord in measure 4, Morpher calculates the chord transitions, from the current Dominating Chord (i.e. C♯ diminished), whose TPS distance is closer to 16. In this case, the closer transitions will go from Dominating Chord to either F major or F minor. Let us imagine that Morpher stochastically selects the latter to fit measure 5. If the rate of decrease from measures 4 to 5 were different from the rate of decrease from measures 5 to 6, Morpher would repeat the above steps to generate a chord to fit measure 5. It would then predict a new Dominating Chord below the tension threshold and it would predict new sequences of subordinates from the right to the left. In this case, since the rate of decrease in Figure 6.24 remains unchanged, let us imagine that the chord generated to fit measure 5 is the one that Morpher already predicted to fit this measure (i.e. C♯ diminished) when generating a chord to fit measure 4.

6.5.3 | Step 3: generating new melodic material that matches input degrees of tension

Whenever Morpher generates new music, that is when the input degree of tension is greater than the pre-defined tension threshold, the third step in its implementation is to generate a new melodic line to fit the current chord so that it matches the input degree of tension. We have implemented two different methods to generate new melodic material depending on whether the input tension is increasing or remains unchanged, or is decreasing. These are introduced below in Sections 6.5.3.1 and 6.5.3.2, respectively.
6.5.3.1 | Melodic generation when tension is increasing or remains unchanged

If the input degree of tension is greater than the tension threshold and is either greater than the previous degree of tension or remains unchanged, Morpher generates new melodic material to match the increasing or flat input tension, respectively.

Given a Subordinate Chord, newly generated as in either Section 6.5.2.1 or Section 6.5.2.2, in order to generate its new melodic material, Morpher starts by defining the number of notes it should consist of. To do so, the number of notes is made equal to the number of notes in the melodic line of the Dominating Chord that dominates upon the current Subordinate Chord. We have made this decision so that the melodic lines against both chords have the same note density. In this way, it would be possible to perceive both melodic lines as having similar grouping analyses, as proposed by GPR 6.

Morpher generates the first note in the new melodic line as the chordal note, from the newly generated Subordinate Chord, that is higher in pitch and the closest to the last melodic note in the preceding melodic line. In this way, we ensure that the transition from the melodic material against Dominating Chord to that against Subordinate Chord is as smooth as possible while still associating the first beat in the Subordinate Chord with a stable note. The remaining notes are generated from a pre-defined template. Finally, Morpher assigns the newly generated melodic material against Subordinate Chord with the rhythm of the melodic material against Dominating Chord. This, again, supports the analysis of similar grouping structures between consecutive melodic materials, as proposed by GPR 6.

The melody-templates mentioned above are shown in Figure 6.25, which consists of pre-composed melodic lines. In this figure, each row represents a sequence of pitches. The number in bold represents the number of notes in the current melodic line. The rest of the values represent the indexes in the current diatonic scale when the first note in Subordinate Chord has been generated as its tonic (Figure 6.25a), its third (Figure 6.25b) or its fifth (Figure 6.25c), the three of which are represented in each figure by index 0, respectively.

We have composed the melodic lines in Figure 6.25 in such a way that they consist of
arpeggios or sequences of conjunct notes that fit in one measure. In this way, they will be perceived as less important structures in GTTM’s hierarchies. Since the shortest note value considered by autognomus is the sixteenth-note, the maximum number of notes in a melodic line will be sixteen, if all notes are sixteenth-notes and the time-signature is quaternary (i.e. $\frac{4}{4}$ or $\frac{12}{8}$). The minimum number of notes in a melodic line will be one, if there is only one note in the current measure. Therefore, Figure 6.25 includes the templates for melodic lines that include from one to sixteenth notes. Likewise, notice that they are all ascending melodic lines. In this way, if tension is increasing, the contours of the melodic lines will also match the input tension trend. Recall that it was observed in Farbood (2012) and Lerdahl and Krumhansl (2007) that, sometimes, melodic contour is essential to modelling musical tension.

In Figure 6.25, we have only included one template melodic line associated with each value of number of notes in the line. There are, however, lots of possible template melodic lines we could have considered. In previous versions of Morpher, we had included many more possible templates. However, we observed that, if tension was increasing for a long time, and so lots of new melodic lines were generated to match the prolonged tension increase, at some point the contour of the newly generated melodic lines stopped matching the increasing tension. That occurred because the notes in the newly generated melodic line reached the farthest note considered in Morpher’s possible range of notes. At this point, the following notes were transposed an octave down so that they could be played by Morpher. When the tension increase lasted for a while, the contour of the melody ended up following a saw-tooth form, because of the transposition strategy. We believe this type of contour can affect perceived musical tension. We designed the template melodic lines in Figure 6.25 so that this could be avoided, insofar as possible.

For example, Figure 6.26 illustrates the application of the template in Figure 6.25a when a Subordinate Chord has been generated as I/C and its first note has been defined as the tonic of this chord.
Chapter 6. The Music Generator

Morpher: the morphing system

Figure 6.25: Diatonic index-based templates of the possible new melodic lines, per measure, pre-defined in *Morpher*.

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(a) Tonic template.

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(b) Third template.

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(c) Fifth template.
Figure 6.26: Example of the application of Figure 6.25a to I/C. Melodic lines a to p refer to the templates 1 to 16 in Figure 6.25a, respectively.
### 6.5.3.2 | Melodic generation when tension is decreasing

If the input degree of tension is greater than the tension threshold and is less than the previous degree of tension, *Morpher* generates new melodic material to match the decreasing input tension, respectively. To do so, *Morpher* applies the method described in Section 6.5.3.1 in a slightly different way. The first note in the melody is no longer generated as the chordal note that is higher in pitch and the closest to the last melodic note in the preceding melodic line. Instead, it is generated as the chordal note that is lower in pitch and the farthest, within one octave, to the last melodic note in the preceding melodic line. A new melodic line is then generated from the templates in Figure 6.25, as in Section 6.5.3.1. Notice that this melodic line will be an ascending line, and so its contour will not match the input tension. Therefore, the melodic line is retrogressed; that is to say, it is re-written from the right to the left. That is why the first note is generated as the farthest chordal note, since it will no longer be the first note in the newly generated melodic line but the last.

### 6.5.3.3 | An example of the application of *Morpher*’s Step 3

To illustrate the application of *Morpher*’s Step 3, Figure 6.27 shows the generation of new melodic material against the Subordinate Chords previously generated by *Morpher* in Figure 6.23b. Again, the new melodic material is trying to match the input tension previously shown in Figure 6.24.

Figure 6.27a includes the melodic line of the theme previously generated in Figure 6.19. As in Section 6.5.2.4, *Morpher* will only generate new material to fill measures 2, 3, 4 and 5. The remaining measures, whose degree of tension according to Figure 6.24 is below the tension threshold, will keep the material of the original theme. This is shown in Figure 6.27b, where the empty measures indicate what should be generated to match the input tension profile.

In measure 2, the input tension is increasing and so *Morpher* will apply the method discussed in Section 6.5.3.1. To generate new melodic material in measure 2, *Morpher* first finds the chordal note in the newly generated Subordinate Chord associated with
Figure 6.27: An example of the application of Morpher’s Step 3 including: (a) the new harmonic sequence in Figure 6.23b, (b) the measures where new melodic material needs to be generated so that the input tension profile in Figure 6.24 is matched and (c) the newly generated melodic material.

this measure (i.e. B♭ major) that is higher in pitch and the closest to the last melodic note in the preceding measure. The last melodic note in measure 1 is g4. The chordal notes in B♭ major consist of b♭, d and f. From the versions of these notes that are higher in pitch than g4, b♭4 is the closest to g4. Therefore, Morpher generates b♭4 as the first note of the new melodic material in measure 2. Since this note is the tonic of the current Subordinate Chord, the new melodic material will be defined by the template lines in Figure 6.25a. From this template, Morpher will use the template line in row four. That is because the number of notes in measure 1 is four. In this row, the template melodic line is defined by the sequence of diatonic indexes 0, 1, 2 and 4. Index 0 corresponds to the already generated tonic note, b♭4. In the diatonic scale oriented at the current chord, B♭ major, index 1 corresponds to the note one step apart from b♭4; that is to say, c5. Similarly, index 2 corresponds to note d5 and index 4 corresponds to the chord’s fifth, f5. Therefore, the new melodic line in measure 2 will consist of notes b♭4, c5, d5 and f5. This melodic line is then assigned the melodic rhythm of the preceding line, that in measure 1, and so it becomes the melodic material shown in measure 2 in Figure 6.27c.

In measure 3, the input tension is also increasing and so Morpher will repeat the steps described in the above paragraph. In this case, the newly generated Subordinate Chord in measure 3 is B major. In this chord, the closest chordal note to the last note
in the newly generated melody in measure 2 is f♯5. This will then be the first note in
the melodic line in measure 3. Since this note is the fifth of the current chord, the new
melodic line will follow the template line in Figure 6.25c’s fourth row. By assigning to
this new melodic line the rhythm in measure 2, the output will be that of measure 3 in
Figure 6.27c.

In measure 4, the input tension is decreasing and so Morpher will apply the method
discussed in Section 6.5.3.2. To generate new melodic material in measure 4, Morpher
first finds the chordal note in the newly generated Subordinate Chord associated with
this measure (i.e. F minor) that is lower in pitch and the farthest, within one octave,
to the last melodic note in the previous measure. The last melodic note in the newly
generated measure 3 is d♯6. The triadic notes in F minor consist of f, a♭ and c. From
the versions of these notes that are lower in pitch than d♯6, f5 is the farthest from d♯6.
Therefore, Morpher generates f5 as the provisional first note of the new melodic material
in measure 4. Since this note is the tonic of the current Subordinate Chord, the new
melodic material will be defined by the template lines in Figure 6.25a’s fourth row. The
provisional new melodic line will then consists of notes f5, g5, a♭5 and c6. In order
to make this a descending melodic line, it is retrogressed, and so the final new melodic
line in measure 4 will consist of notes c6, a♭5, g5 and f5. As above, this melodic line is
assigned the melodic rhythm of the preceding line, that in measure 3, and so it becomes
the melodic material shown in measure 4 in Figure 6.27c.

In measure 5, the input tension is also decreasing and so Morpher will repeat the
steps described in the above paragraph. In this case, the newly generated Subordinate
Chord in measure 5 is C♯ diminished. In this chord, the chordal note farthest to the last
note in the newly generated melody in measure 4 is g4. This will then be the last note in
the melodic line in measure 5. Since this note is the fifth of the current chord, the new
melodic line will follow the template line in Figure 6.25c’s fourth row, but retrogressed.
By assigning to this new melodic line the rhythm in measure 4, the output will be that of
measure 5 in Figure 6.27c.
6.5.4 | Step 4: generating a new arrangement that matches the input tension

Whenever **Morpher** generates new music, that is when the input degree of tension is greater than the pre-defined tension threshold, the fourth step in its implementation is to generate a new arrangement of the newly generated **Subordinate Chord** so that it matches the input degree of tension. We have implemented two different methods to generate the new arrangements depending on whether the input tension is increasing or remains unchanged, or is decreasing. These are introduced below in Sections 6.5.4.1 and 6.5.4.2, respectively.

6.5.4.1 | Arrangement generation when tension is increasing or remains unchanged

If the input degree of tension is greater than the tension threshold and either is greater than the previous degree of tension or remains unchanged, **Morpher** generates a new arrangement of the newly generated **Subordinate Chord** to match the increasing or flat input tension, respectively.

Given a newly generated **Subordinate Chord**, **Morpher** starts by generating its bass note as the closest chordal note to the bass note of the preceding chord (i.e. the **Dominating Chord**). Similarly, the soprano note is generated as the closest chordal note to the soprano note of the preceding chord. Both the newly generated bass and soprano notes must have a higher pitch than their respective preceding notes. In this way, the newly generated pitch height of the bass and soprano voices will also match the input tension, as proposed by Farbood (2012), in case it is increasing.

To generate the notes of the remaining voices, **Morpher** generates all possible S.A.T.B. voicings, based on the **Subordinate Chord**'s chordal notes, that have as bass and soprano notes those that have already been generated. From all the possible voicings, **Morpher** selects the one whose sum of distances from the previous alto note to the current alto note and the distance from the previous tenor note to the current tenor note is the shortest. This ensures that the intermediate voices will satisfy common voice-leading conventions (see Piston (1959, pp.20-23) for more detail).
6.5.4.2 | Arrangement generation when tension is decreasing

If the input degree of tension is greater than the tension threshold and is less than the previous degree of tension, *Morpher* generates new melodic to match the decreasing input tension. To do so, *Morpher* applies the method described in Section 6.5.4.1 in a slightly different way. In this case, the heights of the bass and *soprano* notes are generated as being lower than those in the preceding chord.

6.5.4.3 | An example of the application of *Morpher*’s Step 4

To illustrate the application of *Morpher*’s Step 4, Figure 6.28 shows the generation of the arrangement of the Subordinate Chords previously generated in Figure 6.23b, together with the melodic material previously generated in Figure 6.27c.

Figure 6.28a includes the input theme, previously generated in Figure 6.19. Figure 6.28b illustrates the measures (i.e. 2, 3, 4 and 5) that need to be filled with material so that the input tension profile in Figure 6.24 is matched.

In measure 2, the input tension is increasing and so *Morpher* will apply the method discussed in Section 6.5.4.1. To generate a new arrangement in measure 2, *Morpher* first finds the chordal notes in the newly generated Subordinate Chord associated with this measure (i.e. B♭ major) that are higher in pitch and the closest to the bass and *soprano* notes in the preceding chord. In this case, these are notes d3 and b♭4, respectively. *Morpher* then calculates the possible combinations of S.A.T.B. voices, using the chordal notes of B♭ major, whose bass and *soprano* notes are d3 and b♭4, respectively. From these, *Morpher* selects the voicing whose distance from the *tenor* and *alto* notes in measure 1 is the smallest to the *tenor* and *alto* notes in measure 2. In Figure 6.28c these notes correspond to f3 and f4.

In measure 3, the input tension is also increasing and so the steps discussed in the above paragraph are taken again. This will result in a S.A.T.B. form consisting of: b4, d♯4, b3 and f♯3. This is shown in Figure 6.28c.

In measure 4, the input tension is decreasing and so *Morpher* will apply the method discussed in Section 6.5.4.2. To generate a new arrangement in measure 4, *Morpher* first
Morpher: the morphing system

Figure 6.28: An example of the application of Morpher's Step 4 given Figure 6.24 as the input tension profile to be matched, including: (a) the theme in Figure 6.19 as input theme, (b) the measures that need to be filled with new material to match the input tension profile and (c) the newly generated material according to the harmonic sequence previously generated in Figure 6.23b and the new melodic lines previously generated in Figure 6.27c.

finds the chordal notes in the newly generated Subordinate Chord associated with this measure (i.e. F minor) that are lower in pitch and the closest to the bass and soprano notes in the preceding chord. In this case, these are notes f3 and ab4, respectively. Morpher then calculates the possible combinations of S.A.T.B. voices, using the chordal
notes of $B^\flat$ major, whose bass and soprano notes are $f_3$ and $a^\flat 4$, respectively. From these, *Morpher* selects the voicing whose distance from the tenor and alto notes in measure 3 is the smallest to the tenor and alto notes in measure 4. In Figure 6.28c these notes correspond to $a^\flat 3$ and $c_4$.

In measure 5, the input tension is also decreasing and so the steps discussed in the above paragraph are taken again. This will result in a S.A.T.B. form consisting of: $g_4$, $c^\# 4$, $g_3$ and $e_3$. This is shown in Figure 6.28c.

### 6.5.5 | Step 5: playing the generated music in real time

The fifth step in the implementation of *Morpher* is to play the generated music in real time. To do so, *Morpher* transforms the content in each time-span into objects in MIDIUTil (Wirt, 2018) so that the generated music can be played in MIDI format (Moog, 1986) in real time.

*Morpher* offers the possibility to transform loudness and tempo, in real time, so that these features also match the input degree of tension. These two features have been shown to be strong predictors of perceived arousal (Farbood, 2012; Ilie and Thompson, 2006).

A simple but effective strategy to make loudness and tempo match an input tension profile was implemented in *Escape Point* (Prechtl, 2016). This music generation system was given minimum and maximum degrees of loudness and tempo within which these features were interpolated to match the real-time changing input tension. Inspired by this approach, we have defined the minimum loudness degree as equal to 5 and the maximum as equal to 120, since MIDIUTil’s volume range is $0 – 127$, as per the MIDI standard.\textsuperscript{14}

And we have defined the minimum tempo degree as the input tempo used to generate the original theme and the maximum as twice as much the minimum. By means of a linear interpolation, *Morpher* can calculate new degrees of loudness and tempo, per time-span, to adapt, in real time, to the input degrees of tension. It should be noted, however, that we will not alter tempo and loudness when evaluating *autognomus*. We made this

\textsuperscript{14}In order for the melody to stand out, and so to be perceived as such, its loudness is assigned 7 more units than that assigned to the harmonic voices.
decision so that the results will only concern the rhythmic and harmonic characteristics of autognomus. We will come back to this idea in Chapter 7.

6.6 | Summary and conclusions

In order to seek answers to our Research Question, Chapter 6 has designed and implemented a system capable of generating tonal music with long-term structure that matches input tension profiles in real time.

Section 6.1 has introduced autognomus, our publicly available music generation system. To design autognomus, Section 6.1 has identified the vertical mixing technique as the compositional strategy best suited to our goals. Inspired by the four main tasks this technique consists of, autognomus was designed as a system consisting of four sub-systems, Themer, Arranger, Developer and Morpher, each of which tackles one of the tasks.

Section 6.2 has introduced the Themer, autognomus’s first sub-system, which is capable of automatically generating musical themes. The themes are generated in such a way that their structure is well-defined, according to Lerdahl and Jackendoff’s Generative Theory of Tonal Music (GTTM) rules, and so can be used as reference. To do so, Themer applies a collection of rules to give control over the generation of themes. These rules force the themes to form specific grouping structures, metrical structures and melodic and harmonic relations. In this way, the themes follow an intended hierarchical structure and so it will be possible to apply Lerdahl’s Model of Tonal Tension (MTT) to them.

Section 6.3 has introduced the Arranger, autognomus’s second sub-system, which is capable of automatically arranging the notes in the chords of a given theme. To do so, Arranger applies a collection of rules to give control over the generation of accompaniments. These rules determine the inversions of the chords in an accompaniment so that their stability matches the intended hierarchical structure. Likewise, the rules force the chords to satisfy some common voice-leading conventions.

Section 6.4 has introduced the Developer, autognomus’s third sub-system, which is capable of automatically generating new musical material that relates to a given theme.
The new material is generated in such a way that it conveys some sense of long-term structure. To do so, Developer applies a collection of rules to give control over the generation of the new material. These rules force the new material to be generated by adding chords to the original theme and new melodies defined by the transformations of the melodies in the original theme.

Finally, Section 6.5 has introduced the Morpher, autognomus's fourth sub-system, which is capable of automatically transforming a given theme, and its developments, into new musical material. The new material is generated in such a way that it matches input tension profiles in real time. To do so, Morpher applies a collection of rules to give control over the generation of the new material. These rules force the chords in the new material to be selected from those that may best suit the input degree of tension according to Lerdahl's MTT. They also force the contours of the newly generated melodies and the accompaniments to match the input tension profiles.
Chapter 7

Evaluating the Music Generator

*Methodology, computational tests and empirical studies*

In this chapter, we evaluate the performance and applicability of the Music Generator. To do so, we will:

1. define the methodology for evaluating the performance and applicability of the Music Generator,
2. design a collection of computational tests and empirical studies according to the defined evaluation methodology,
3. present the procedures and the results of the designed computational tests and empirical studies, and
4. analyse the results of the designed computational tests and empirical studies and discuss their impact with regards to our Research Question.

The first step is addressed in Section 7.1, the second and third steps are addressed in Sections 7.2, 7.3 and 7.4, which cover a computational test and two empirical studies, respectively, and the fourth step is addressed in Section 7.5.
Chapter 7. Evaluation

7.1 | Evaluating the Music Generator

To evaluate the capabilities and limitations of autognomus we need to evaluate the extent to which the generated music meets our goals; that is to say, evaluate the extent to which the music generated by autognomus is generated in real time, has long-term structure and matches input tension profiles. Sections 7.1.1, 7.1.2 and 7.1.3 define the evaluation methodologies with regards to each goal, respectively.

7.1.1 | To what extent is the Music Generator applicable in real time?

As discussed in Chapter 6, the generation of music in autognomus takes place chord by chord. As a consequence, even if the input degrees of tension change while in the middle of a chord, autognomus will not generate new material to match the new degree of tension until the current chord has been played in full. We made this decision in Section 6.5 so that the intended hierarchical relations are matched and so some sense of long-term structure can be sensed. However, because of this decision, autognomus may not be able to adapt, in real time, to sudden tension changes. Therefore, the real-time capabilities and limitations of autognomus will depend on the extent to which the degrees of tension considered by autognomus, once every chord is played in full, correlate with the continuous flow of tension of the input tension profile.

To illustrate the above issue, let us imagine that autognomus is fed as input the tension profile that corresponds to the solid line in Figure 7.1. This curve corresponds to the judgements of tension recorded by the participants in Lerdahl and Krumhansl (2007) when they listened to a reduction of Chopin’s Prelude No. 9 in E major, Opus 28 (see Figure 7.2). These data were provided by C. Krumhansl (personal communication, October, 2019).

Let us also imagine that the tempo, the time-signature and the number of measures given as input to autognomus are defined as in Figure 7.2. autognomus will not consider all the data points in the input tension profile (i.e. the solid line in Figure 7.1), but only
Chapter 7. Evaluation

Evaluating the Music Generator

Figure 7.1: Judgements of tension provided by the participants in Lerdahl and Krumhansl (2007) when they listened to Figure 7.2 (solid line) and the data points read by autognomus at the beginning of each measure if the solid line was fed as input into autognomus (dashed line).

Figure 7.2: Reduction of Chopin's Prelude No. 9 in E major, Opus 28, used in the empirical study in Lerdahl and Krumhansl (2007).
those that correspond to the beginning of each time-span. If the time-spans lasted one measure, the data points considered by autognomus at the beginning of each measure will be those connected by a dashed line in Figure 7.1. Despite the fact that the tension data points considered by autognomus do not contain all the data points in the original input profile, the correlation between both profiles is strong.\footnote{According to Lerdahl and Krumhansl’s (2007) evaluation methodology, which was previously introduced in Section 3.5.3: $R^2 = .98, df = (1, 46), R_{adj}^2 = .98$ and $p < .0001$.} In Figure 7.1 we refer to the input degrees of tension considered by autognomus, those that correspond to the degrees at the beginning of each span, as the tension read by autognomus. Hereinafter, we will use this terminology.

autognomus’s accuracy, concerning its capability of generating music in real time, will depend on the amount of tension data points it misses from the input profile. As discussed in Section 6.2.2.3, a chord’s time-spans will consist of either one, two or three measures, as defined in Tables 6.1 and 6.2. The minimum mismatching between an input degree of tension and that read by autognomus will be equal to zero, when the input tension changes right before a new chord is being generated, whereas the maximum mismatching will be equal to the total duration of a chord. To illustrate the latter, Figure 7.3 shows the duration of one-measure time-spans, in seconds, for different tempo markings, in beats per minute (BPM).

According to Figure 7.3, the mismatchings between the changes of the input degrees of tension and those read by autognomus could be of a couple of seconds. In those situations, the music generated by autognomus will not match input tension profiles in real time because the mismatchings are too long. However, there are two phenomena that must be taken into consideration. First, in autognomus, tempo can proportionally change with regards to the input degree of tension, as discussed in Section 6.5.5. In this way, the tenser the input tension profile gets, the faster the tempo the generated chord’s time-span will be assigned, and so, according to Figure 7.3, the shorter the real-time mismatching will be. Second, the tension curves described by most pieces of music usually show some sort of smooth behaviour, meaning they do not tend to be erratic (i.e. they do not include many sudden changes of tension in short periods of time). See for
Figure 7.3: Duration of a measure with regards to tempo, in BPM, for time-signatures with binary division, 2/4 and 6/8 (solid line); with ternary division, 3/4 and 9/8 (dashed line); and with quaternary division, 4/4 and 12/8 (dotted line).

instance the solid line in Figure 7.1. In this way, even when autognomus’s mismatches prevent the system from considering all data points in the input tension profile, the points that are read may still strongly correlate with the input profile.

To evaluate the extent to which autognomus is applicable in real time, we have carried out a computational test where we fed different tension profiles, extracted from real pieces of music, into autognomus to be matched by the automatically generated music. In this way, it is possible to analyse the correlation between the degrees of tension read by autognomus against the degrees of tension of the pieces’ profiles. This test is discussed in more detail in Section 7.2.
7.1.2 | To what extent does the generated music have long-term structure?

The next step in the evaluation is to analyse whether the music generated by *autognomus* has long-term structure. As concluded in Chapter 2, MTT can be used to test the accuracy of generated GTTM structures, which are the structures the generated music is based on. Therefore, we can use MTT’s evaluation methodology to meet this next step.

To evaluate the extent to which *autognomus* generates music that has long-term structure, we have carried out an empirical study where a group of participants recorded the degrees of tension they perceived while listening to themes generated by *autognomus*’s *Themer*. In this way, it is possible to analyse the correlation of the degrees of tension estimated by *autognomus*, based on the intended GTTM’s hierarchical structures, against the participants’ judgements of tension. This study is discussed in more detail in Section 7.3.

7.1.3 | To what extent does the generated music match input tension profiles?

The last step in the evaluation is to analyse whether the music generated by *autognomus* matches input tension profiles. As observed in Chapter 2, from the existing methodologies, the most used, and probably most developed, is that where human judgements of tension are recorded. Well-grounded methodologies for designing this type of study and evaluating its results have been developed in Farbood (2012); Krumhansl (1996); Lerdahl and Krumhansl (2007) (recall these have already been introduced in Chapters 2 and 3).

To evaluate the extent to which *autognomus* generates music that matches input tension profiles, we have carried out an empirical study where a group of participants recorded the degrees of tension they perceived while listening to themes morphed by *autognomus*’s *Morpher*. In this way, it is possible to analyse the correlation between the degrees of tension fed as input into *autognomus* against the participants’ judgements of tension. This study is discussed in more detail in Section 7.4.
Chapter 7. Evaluation

Computational test

7.2 | Computational test

As discussed in Section 7.1.1, we have carried out a computational test to evaluate the extent to which autognomus is applicable in real time. The test’s materials, procedure and results are presented in Sections 7.2.1, 7.2.2 and 7.2.3, respectively.

7.2.1 | Materials

As proposed in Section 7.1.1, in order to carry out the computational test, it is preferable to use tension profiles from real pieces of music. In this way, the capabilities and limitations derived from the test’s results will be more realistic.

The tension profiles used in the computational test consist of those of the first one hundred pieces in the Interactive GTTM Analyser (IGA) dataset.\(^2\) AuToTen was applied to these pieces to extract their degrees of hierarchical tension, \(T_{\text{hier}}\). In order to get the degrees of tension as accurate as possible, we manually calculated the key labels, chord labels and degrees of TPS’s surface parameters of the one hundred pieces, instead of using AuToTen’s automatic extraction features.

Tempo and loudness remained unchanged throughout the pieces to make sure the results of the study concern autognomus’s main melodic and harmonic components.

7.2.2 | Procedure

To carry out the computational test, we used Themer to generate a template theme. We then fed Morpher with the tension profiles from the one hundred IGA pieces to transform the template theme so that the tension profiles were matched. In order to investigate the differences of accuracy with regards to tempo, we fed the nine different tempo markings shown in Figure 7.3 (i.e. 40, 60, 80, 100, 120, 140, 160, 180 and 200 BPM) into Morpher. Likewise, in order to investigate the differences of accuracy with regards to time-signature, we used each tempo marking with a binary time-signature (2/4), a ternary time-signature (3/4) and a quaternary time-signature (4/4). That is to say, Morpher was

\(^2\)http://gttm.jp/
applied a total of two thousand seven hundred times: for each of the one hundred tension profiles, each of the nine tempo markings was combined with each of the three time-signatures \((i.e. 100 \cdot 9 \cdot 3 = 2700)\).

7.2.3 | Results

To calculate the degrees of correlation between the tension profiles of IGAs one hundred pieces and the degrees of tension read by autognomus, for each of the two thousand seven hundred generations, we calculated a linear least-squares regression. Figure 7.4 shows the degrees of correlation, for each tempo marking, concerning a binary time-signature \((2/4)\), a ternary time-signature \((3/4)\), and a quaternary time-signature \((4/4)\), respectively. In this figure, the boxes show the quartiles of the correlation coefficients, \(r\), of the one hundred pieces for each tempo marking, while the whiskers extend to show the rest of the distribution.

Table 7.1 summarises the results of the linear regressions. Each cell in the table refers to one hundred regressions, in the same way the boxes in Figure 7.4 do. The table includes an average degree of correlation, \(r_{av}\), and the amount of significant regressions \((i.e. p \leq .05)\), \(\%_{sig}\), for each tempo marking.

Table 7.1: A summary of the average correlation coefficients, \(r_{av}\), and the percentage of significant regressions, \(\%_{sig}\), between the tension profiles of IGAs one hundred pieces and the degrees of tension read by autognomus in real time.

<table>
<thead>
<tr>
<th>Tempo (BPM)</th>
<th>Time-signature</th>
<th>Variables</th>
<th>40</th>
<th>60</th>
<th>80</th>
<th>100</th>
<th>120</th>
<th>140</th>
<th>160</th>
<th>180</th>
<th>200</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2/4</td>
<td>(r_{av})</td>
<td>.71</td>
<td>.81</td>
<td>.86</td>
<td>.88</td>
<td>.89</td>
<td>.91</td>
<td>.92</td>
<td>.92</td>
<td>.93</td>
</tr>
<tr>
<td></td>
<td>%_{sig}</td>
<td></td>
<td>93</td>
<td>98</td>
<td>98</td>
<td>98</td>
<td>97</td>
<td>99</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>3/4</td>
<td>(r_{av})</td>
<td>.72</td>
<td>.81</td>
<td>.86</td>
<td>.89</td>
<td>.88</td>
<td>.91</td>
<td>.92</td>
<td>.92</td>
<td>.94</td>
</tr>
<tr>
<td></td>
<td>%_{sig}</td>
<td></td>
<td>94</td>
<td>98</td>
<td>98</td>
<td>98</td>
<td>97</td>
<td>99</td>
<td>100</td>
<td>99</td>
<td>99</td>
</tr>
<tr>
<td></td>
<td>4/4</td>
<td>(r_{av})</td>
<td>.72</td>
<td>.81</td>
<td>.86</td>
<td>.89</td>
<td>.88</td>
<td>.9</td>
<td>.9</td>
<td>.93</td>
<td>.92</td>
</tr>
<tr>
<td></td>
<td>%_{sig}</td>
<td></td>
<td>94</td>
<td>98</td>
<td>98</td>
<td>98</td>
<td>97</td>
<td>98</td>
<td>99</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

From the results in Figure 7.4 and in Table 7.1, three observations can be made: first, as previously shown in Figure 7.3, the faster the input tempo marking, the more accurate autognomus’s real-time performance is; second, in spite of the differences across time-signatures shown in Figure 7.3, the differences in the degrees of correlation are barely
Chapter 7. Evaluation

Computational test

Figure 7.4: Degrees of correlation, $r$, between the tension profiles of IGA’s one hundred pieces and the degrees of tension read by autognomus.
noticeable across time-signatures; and third, in spite of the long mismatchings proposed in Figure 7.3 for the slower tempo markings, autognomus's performance seems to be fairly accurate in most cases (more than 90% of the regressions are significant and the average correlation coefficients are greater than .7).

Section 7.5.1 will discuss the impact that the results we have presented in this section have with regards to our Research Question, concerning the real-time generation of music.

7.3 | Empirical study I

As discussed in Section 7.1.2, we have carried out an empirical study to evaluate the extent to which autognomus generates music that has long-term structure. Its participants, materials, procedure and results are presented in Sections 7.3.1, 7.3.2, 7.3.3 and 7.3.4, respectively.

7.3.1 | Participants

We recruited a total of eighty-five participants online to participate in the empirical study. Table 7.2 groups the participants by gender and by years of musical training.

Table 7.2: A summary of the participants that took part in the empirical study grouped by gender and years of musical training.

<table>
<thead>
<tr>
<th>Number of participants</th>
<th>Gender</th>
<th>≥ 10 years of training</th>
<th>&lt; 10 years of training</th>
<th>No training</th>
<th>All participants</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>female</td>
<td>11</td>
<td>7</td>
<td>15</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>male</td>
<td>31</td>
<td>7</td>
<td>11</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>other</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>not said</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>TOTAL</td>
<td>43</td>
<td>15</td>
<td>27</td>
<td>85</td>
</tr>
<tr>
<td>Average years of training</td>
<td>21.3</td>
<td>4.3</td>
<td>0</td>
<td>16.4</td>
<td></td>
</tr>
</tbody>
</table>
7.3.2 | Materials

We generated a total of four themes using autognomus’s Themer. To do so, we fed Themer with the following inputs: number of measures equal to 8, tempo equal to 60 BPM, time-signature equal to 2/4 and character equal to allegro. Two themes were generated in the key of C major and the other two in the key of C minor. In this way, as both keys share the same tonal centre, when changing from one theme to another in the study, participants would not perceive a great difference and so their perception of tension would not be greatly affected when listening to a new theme.

The generated four themes are shown in Figure A.1 in Section A.1 in Appendix A. At first, we saved the generated four themes in MIDI format (Moog, 1986) using MIDIUTil (Wirt, 2018). We generated the scores of the MIDI files, those in Section A.1, using Musescore 2.3.2 (Watson, 2018). We then used Musescore to convert the MIDI files into WAV format (Whibley et al., 2016).

The music generated by autognomus consists of five voices: a melodic voice and four harmonic voices. Considering how autognomus is implemented, the range of the melodic voice shares part of the ranges of the upper harmonic voices. That is to say, the melodic voice could cross the upper harmonic voices or it could even duplicate some of their notes. Therefore, the melodic and harmonic voices must not be played with the same instrument so that participants can easily identify the main melody.

In Escape point (Prechtl, 2016), a synthesizer is used to play the automatically generated music. One of the benefits of using a synthesizer is that it allows us to control the decay of generated notes, compared to the decay they may experience if they were played with another instrument such as the piano. In this way, a continuous flow of tension could be perceived more easily. Inspired by Prechtl’s use of synthetic sounds, we assigned the melodic voice in the four generated themes to Musescore’s sine synthesizer and assigned the four harmonic voices to Musescore’s string synthesizer.

We did not alter the degrees of loudness and tempo in the four generated themes. Therefore, we will not consider these features in our analysis.

---

3https://musescore.org/
7.3.3 | Procedure

Participants were sent an online invitation to participate in the study. The invitation directed them to an online platform where the study was carried out. Participants were asked to read some information about the study as well as a consent form. They were also given a questionnaire from which the data shown in Table 7.2 was taken. Finally, they were asked to listen to the four themes introduced in Section 7.3.2 and to record the degree of tension they perceived, in real time, using a virtual slider. Our definition of tension was not discussed with participants. The implications that this decision may have for the results are discussed in more detail in Section 8.4 in Chapter 8. For more details about the experimental interface materials, see Section A.2 in Appendix A.

The virtual slider used in the study is shown in Figure 7.5. The slider values were collected every 100 milliseconds.

![Virtual slider](image)

**Figure 7.5:** Virtual slider used in autognomus’s empirical studies.

Because of the COVID-19 pandemic, we were forced to carry out the study in an online format. This poses some disadvantages and some advantages. On one hand, concerning the former, there may not be total control of the data recording process. On the other hand, concerning the latter, it is easier to recruit a bigger group of participants and so the results of the study may be more robust. For instance, in other face-to-face tension-related studies that follow a similar experimental methodology, such as those in Farbood (2012); Farbood and Upham (2013); Krumhansl (1996); Lehne et al. (2014); Lerdahl and Krumhansl (2007), the number of participants ranges between fifteen and twenty-five.
Since we carried out the study in an online format, we reached a greater group of participants. We designed the experimental methodology so as to be as simple as possible, in such a way that the chances of not having control of the data recording process were minimised.

The study followed a *continuous-tension* task rather than a *stop-tension* task (i.e. in the latter, participants judge each event one by one, with the music stopping at each event of study, whereas, in the former, participants judge the whole pieces of music using a slider in real time), which recall were introduced in Chapter 2, since the former is less time-consuming and so may be more appealing to participants. Since Lerdahl and Krumhansl (2007) concluded that both tasks produce similar results, the *continuous-tension* task has been preferred over the *stop-tension* task. See, for instance, Farbood (2012); Farbood and Upham (2013); Lehne et al. (2014).

Despite the fact all the pieces of music used in the study start at the tonic chord, and so their degree of tension will be close to zero, we set the initial position of the slider in the online interface at the slider’s middle value. We made this decision based on comments given by participants in previous pilot studies. Although this was taken into consideration, according to Hackworth and Fredrickson (2010), the initial position of a continuous slider does not seem to make a difference in the end.

### 7.3.4 | Results

The results of the empirical study to investigate whether the generated music could convey some sense of long-term structure are presented in Section 7.3.4.4. The evaluation methodology used in this section takes into account the pre-processing we applied to the raw data for the degrees of inter-subject agreement and for the time lags in the responses. Sections 7.3.4.1, 7.3.4.2 and 7.3.4.3 discuss how we dealt with these issues, respectively.

#### 7.3.4.1 | Data pre-processing

Before calculating any experimental results, we pre-processed the participants’ judgements following Farbood’s (2012) methodology. First, we re-sampled the judgements
to a data point every 500 milliseconds. This value corresponds to the duration of an eighth-note in the pieces used in the study. In this way, each data point is assigned to an eighth-note in the corresponding piece. Second, we standardised the judgements so that they were all on the same scale.

### 7.3.4.2 Inter-subject agreement

To decide how the participants’ judgements should be analysed, we first calculated the degree of inter-subject agreement across all pieces used in the study. In Krumhansl (1996) and in Farbood (2012), to calculate the degrees of inter-subject agreement, they calculate Pearson’s correlation coefficient for every pair of judgements provided by the participants. Then, for each piece of music, they calculate the average value of Pearson’s correlation coefficient as an indicator of inter-subject agreement.

According to Koo and Li (2016, p.156):

> Pearson correlation coefficient is only a measure of correlation, and hence, nonideal measure of reliability. A more desirable measure of reliability should reflect both degree of correlation and agreement between measurements.

Therefore, we need another measure of correlation that accounts for agreement between measurements.

According to Booth and Narayanan (2020), approaches to calculating inter-subject agreement using continuous data include Cronbach’s $\alpha$ and the Intra-class Correlation Coefficient (ICC), and so we have used these measures here instead of Pearson’s correlation coefficient. We have used both methods to make sure the calculated degrees of agreement are robust.

There exist at least ten different types of ICCs and calculating one or another depends on the characteristics of the study. According to Koo and Li (2016); Perinetti (2018), in order to determine the ICC that suits the characteristics of our study, we need to identify the model we are using, the type of measurement we are interested in and the goal we want to meet with the analysis. Table 7.3 summarises the criteria to be used in the analysis.
Table 7.3: Criteria used to determine the type of Intra-class Correlation Coefficient (ICC) one needs according to Koo and Li (2016); Perinetti (2018).

<table>
<thead>
<tr>
<th>Criterium</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td></td>
</tr>
<tr>
<td>One-way random-effect</td>
<td>Different groups of participants, randomly chosen from a population, provide judgements of different samples.</td>
</tr>
<tr>
<td>Two-way random-effect</td>
<td>A unique group of participants, randomly chosen from a population, provides judgements of the same samples.</td>
</tr>
<tr>
<td>Two-way mixed-effect</td>
<td>A unique group of participants, not randomly chosen form a population, provides judgements of the same samples.</td>
</tr>
<tr>
<td>Type of measurement</td>
<td></td>
</tr>
<tr>
<td>Single rater</td>
<td>The basis of the analysis are the judgements from a single rater.</td>
</tr>
<tr>
<td>Mean of multiple raters</td>
<td>The basis of the analysis is the mean of the judgements from a set of raters.</td>
</tr>
<tr>
<td>Goal</td>
<td></td>
</tr>
<tr>
<td>Absolute agreement</td>
<td>Are the judgements of different raters the same?</td>
</tr>
<tr>
<td>Consistency</td>
<td>Do the judgements of different raters correlate in an additive manner?</td>
</tr>
</tbody>
</table>

Let us determine the ICC that suits the characteristics of our study. First, we have used a unique group of randomly selected participants, each of which provided judgements of tension of the same collection of musical stimuli. Therefore, according to Table 7.3, our model is a two way random-effect model. Second, we want to focus on the average judgements of tension, as in Farbood (2012); Krumhansl (1996); Lehne et al. (2014); Lerdahl and Krumhansl (2007), so that the conclusions may extend to the whole population. Therefore, according to Table 7.3, our type of measurement is the mean of multiple raters. Third, we would like to analyse the agreement between participants as a matter of absolute agreement rather than correlation between their judgements, since their judgements of tension have already been standardised in Section 7.3.4.1. Therefore, according to Table 7.3, our goal is absolute agreement.

In summary, our study is a two-way random-effect model based on multiple raters and absolute agreement. In Shrout and Fleiss's (1979) convention, this type of study corresponds to the coefficient ICC $(2, k)$.

Table 7.4 shows the degrees of inter-subject agreement for all eighty-five participants across the four themes used in the study according to Cronbach’s $\alpha$ and ICC $(2, k)$. We
calculated Cronbach’s $\alpha$ using the TCI statistic library\(^4\) and calculated ICC ($2, k$) using the statistical package Pingouin\(^5\). Notice the degrees of ICC include a confidence interval (95%).

Table 7.4: Degrees of inter-subject agreement for all eighty-five participants across the four themes used in the study.

<table>
<thead>
<tr>
<th>Profile themes</th>
<th>Piece</th>
<th>Cronbach’s $\alpha$</th>
<th>Intra-class Correlation Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>ICC ($2, k$)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Confidence interval</td>
</tr>
<tr>
<td>1</td>
<td>.94</td>
<td>.94</td>
<td>.91 – .97</td>
</tr>
<tr>
<td>2</td>
<td>.85</td>
<td>.85</td>
<td>.78 – .91</td>
</tr>
<tr>
<td>3</td>
<td>.82</td>
<td>.82</td>
<td>.75 – .89</td>
</tr>
<tr>
<td>4</td>
<td>.95</td>
<td>.95</td>
<td>.93 – .97</td>
</tr>
</tbody>
</table>

In Table 7.4, the degrees of agreement calculated using Cronbach’s $\alpha$ and ICC ($2, k$) are identical and so we consider them to be reliable indicators of inter-subject agreement.

According to Cicchetti and Sparrow (1981), an ICC smaller than .5 should be considered as an indicator of poor agreement, an ICC between .5 and .75 should be considered as an indicator of a fair agreement, an ICC between .75 and .9 should be considered as an indicator of a good agreement and an ICC greater than .9 should be considered as an indicator of an excellent agreement. Following this convention, two out of the four themes used in the study show excellent agreement in Table 7.4 and the other two themes show good agreement. Because of this, we decided to use the average of the judgements across all eighty-five participants, for each theme, for all subsequent analysis. For more details about the judgements of tension provided by the participants, see Section A.3 in Appendix A.

### 7.3.4.3 | Time lag

Before calculating any results, it should be noted that there may be a delay in the participants’ judgements of perceived tension. That is because participants had to listen to the music, reflect on the degree of tension they were perceiving, move the slider and then

\(^4\)https://github.com/anthropedia/tci-stats
\(^5\)https://pingouin-stats.org/
their responses had to be recorded. These four tasks take some time to be performed. To compensate for this delay, we used a temporal lag of 2 seconds to interpret the participants’ judgements. This lag is the same as the one used by Farbood and Upham (2013), it is equivalent to the lag of two beats proposed by Krumhansl (1996) and it is in line with Schubert’s (2004) conclusions, who has observed that real-time judgements of arousal in continuous signals often present a temporal lag between 1 and 3 seconds.

7.3.4.4 | Results to investigate whether long-term structure could be sensed

Our goal is to investigate the extent to which the music generated by autognomus has long-term structure. According to Aspromallis and Gold (2016); Carnovalini and Rodà (2020); Lerdahl and Jackendoff (1983), long-term structure often takes the form of recurring patterns. autognomus was designed based on this assumption. That is why Themer generates musical themes with well-defined patterns which will be used by Developer to generate new material based on these patterns. In this way, the combination of the themes and their developments could convey some sense of long-term structure.

Evaluating whether the themes and their developments convey long-term structure is a challenging task. According to Hall and Pearce (2021, p.220):

[l]arge-scale structure – the global organisation of a work’s material – encompasses several different concepts. First, a distinction can be made between thematic and tonal structures, the first concerning the structuring of repeated musical material, the second the hierarchical organisation of harmonies relating to key.

The sonata form is a good example to illustrate what thematic and tonal structures entail and how their interaction helps to acquire long-term structure. A classical sonata usually consists of three sections: the exposition, the development and the recapitulation. In the exposition, two contrasting themes are introduced and are played twice. To Hall and Pearce (2021, p.220), “[r]epetition seems very likely to play an important role in the

---

6 In Sections 8.4.1 and 8.4.2, in Chapter 8, we discuss the concept of long-term structure and its implications in the present research.
perception of large-scale”. The first theme is played in the piece’s global key. The second theme is either played in the key of the dominant or in the key of the major relative, depending on whether the piece’s key is major or minor, respectively. The transition between the themes usually ends on a semicadence of the second theme’s key. And the second theme usually ends on an expanded cadence. In the development, the harmony modulates through different keys. Its melodic musical ideas often resemble those in the exposition’s themes. Finally, in the recapitulation, both themes are played again, the second of which is transposed to the key of the first theme.

In the Classical sonata form, the themes are usually based on classical eight-measure phrases and the texture is often an accompanied melody. The main thematic structures in a sonata are its two themes. The schematic harmonic structure in the exposition and in the recapitulation, as well as the modulation phase in the development, are the tonal structures. The combination of both structures is what gives the sonata its particular form, a form with long-term structure.

Notice that autognomus’s generated themes and their developments are based on the repetition of patterns, as discussed in Sections 6.2 and 6.4. We believe that our implementation matches Hall and Pearce’s idea of thematic structures, since it is based on repeated patterns. The question is now whether their idea of tonal structures is also matched. In order to investigate this question, we ask ourselves whether the patterns generated by autognomus in the generated themes are perceived as such by human listeners. If so, we will assume that autognomus matches Hall and Pearce’s above interpretation, and so we will conclude that the music generated with autognomus can convey some sense of long-term structure.

The patterns generated by autognomus are based on GTTM’s grouping, metrical and prolongational relations. Therefore, these patterns will be perceived as such by human listeners if the musical representations inferred by the listeners match those predicted by GTTM. As concluded in Section 3.5.4, MTT can be used to test the accuracy of generated GTTM structures, as supported by Lerdahl (2011); Lerdahl and Krumhansl (2007). Therefore, to evaluate the extent to which the patterns intended by autognomus are perceived as such by human listeners, we must follow a methodology similar to MTT’s
evaluation methodology (i.e. that in Lerdahl and Krumhansl (2007)), which recall was introduced in Section 3.5. Following Lerdahl and Krumhansl' methodology, we have carried out an empirical study where a group of participants recorded the degrees of tension they perceived while listening to the four themes generated by autognomus’s Themer. In this way, it is possible to analyse the correlation of the degrees of tension estimated by autognomus, based on the intended GTTM’s hierarchical structures, against the participants’ judgements of tension. High degrees of correlation between both tension curves would indicate that the patterns generated by autognomus are likely perceived as such by the listeners. If that is the case, we will assume that Hall and Pearce’s (2021) above ideas are matched in full and so we will conclude that the music generated with autognomus can convey some sense of long-term structure.

According to Lerdahl and Krumhansl (2007), calculating a multiple regression using theoretical and judged tension profiles will allow us to analyse the degree of fitting between both profiles as well as the independent contribution of each musical feature to the regression’s predictions. We, therefore, will calculate a collection of multiple regressions using the generated themes introduced in Section 7.3.2. In order to apply MTT to the analysis, we must always consider MTT’s hierarchical tension, $T_{\text{hier}}$, and attraction, $\alpha$, as independent variables in our multiple regressions.

As discussed in Sections 3.5 and 2.4.2, it was observed by Lerdahl and Krumhansl (2007) and by Farbood (2012), respectively, that, sometimes, the contour of the main melody, $\text{cont}_{\text{melody}}$, the contour of the bass voice, $\text{cont}_{\text{bass}}$, and the contour of the soprano voice, $\text{cont}_{\text{soprano}}$, are also essential to model perceived musical tension. Therefore, we might also consider these variables as independent variables in our multiple regressions. In order to determine which features should be considered as independent variables, Farbood focused on the most salient features in each musical stimulus. As shown in Section 3.5, Lerdahl and Krumhansl determined the salience of $\text{cont}_{\text{melody}}$ by calculating the degree of fitting between their collected judgements of tension and $\text{cont}_{\text{melody}}$. Whenever the correlation between the two variables was statistically significant (i.e. $p < .05$), it was concluded that $\text{cont}_{\text{melody}}$ was salient.

We have followed Lerdahl and Krumhansl’s methodology to determine whether $\text{cont}_{\text{melody}}$,
cont\textsubscript{bass}, and/or cont\textsubscript{soprano} are salient for each of the four generated themes. Table 7.5 includes the degrees of fitting, $R^2$, and the $p$ values of the corresponding regressions.

Table 7.5: Results of the independent regressions of the participants' judgements of tension, of the four generated themes, against cont\textsubscript{melody}, cont\textsubscript{bass}, and cont\textsubscript{soprano}.

<table>
<thead>
<tr>
<th>Theme</th>
<th>cont\textsubscript{melody}</th>
<th>cont\textsubscript{bass}</th>
<th>cont\textsubscript{soprano}</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$R^2$</td>
<td>$p$</td>
<td>$R^2$</td>
</tr>
<tr>
<td>1</td>
<td>.01</td>
<td>.585</td>
<td>.02</td>
</tr>
<tr>
<td>2</td>
<td>.01</td>
<td>.599</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>.46</td>
<td>&lt; .0001</td>
<td>.02</td>
</tr>
<tr>
<td>4</td>
<td>.33</td>
<td>&lt; .0001</td>
<td>.31</td>
</tr>
</tbody>
</table>

In Table 7.5, cont\textsubscript{soprano} is the only statistically significant variable in Themes’ 1 and 2 regressions; cont\textsubscript{soprano} and cont\textsubscript{melody} are the only statistically significant variables in Theme’s 3 regression; and, cont\textsubscript{soprano}, cont\textsubscript{melody} and cont\textsubscript{bass} are all statistically significant in Theme’s 4 regression. Therefore, the predictions of tension, $P$, corresponding to the multiple regression we will calculate for each theme, should take the following forms:

**Theme 1:**

$$P = \beta_t \cdot T_{hier} + \beta_a \cdot \alpha + \beta_s \cdot \text{cont\textsubscript{soprano}}$$

**Theme 2:**

$$P = \beta_t \cdot T_{hier} + \beta_a \cdot \alpha + \beta_s \cdot \text{cont\textsubscript{soprano}}$$

**Theme 3:**

$$P = \beta_t \cdot T_{hier} + \beta_a \cdot \alpha + \beta_m \cdot \text{cont\textsubscript{melody}} + \beta_s \cdot \text{cont\textsubscript{soprano}}$$

**Theme 4:**

$$P = \beta_t \cdot T_{hier} + \beta_a \cdot \alpha + \beta_m \cdot \text{cont\textsubscript{melody}} + \beta_b \cdot \text{cont\textsubscript{bass}} + \beta_s \cdot \text{cont\textsubscript{soprano}}$$

since the above expressions always include $T_{hier}$ and $\alpha$, as concluded above, as well as those variables whose regression in Table 7.5, respectively, is statistically significant.

The results of the above multiple regressions are shown below in Table 7.6. As in Section 3.5, Table 7.6 includes a value of $R^2$ as a goodness-of-fit measure (i.e. proportion of variation) of MTT’s predictions; the degrees of freedom, $df$, associated with $R^2$, the first of which indicates the number of predictor (i.e. independent) variables there are in the regression and the second of which corresponds to the data points in the regression minus the first degree of freedom minus one; an adjusted value of $R^2$, $R^2_{adj}$, so that
it could be compared with other models for similar data that have different degrees of freedom; the coefficients in the linear model, $\beta$; and the regression’s $p$ values, which, by convention, indicate that $R^2$ may be significant when $p < .05$.

Table 7.6: Results of the multiple regressions, that include the corresponding salient features, of the four generated themes.

<table>
<thead>
<tr>
<th>Theme</th>
<th>$R^2$</th>
<th>$R^2_{adj}$</th>
<th>df</th>
<th>variable</th>
<th>$\beta$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.77</td>
<td>.75</td>
<td>(3, 30)</td>
<td>$T_{\text{hier}}$</td>
<td>.18</td>
<td>.0022</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\alpha$</td>
<td>.12</td>
<td>.0456</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\text{cont}_{\text{soprano}}$</td>
<td>.31</td>
<td>$&lt; .0001$</td>
</tr>
<tr>
<td>2</td>
<td>.7</td>
<td>.67</td>
<td>(3, 30)</td>
<td>$T_{\text{hier}}$</td>
<td>-.05</td>
<td>.0103</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\alpha$</td>
<td>.22</td>
<td>$&lt; .0001$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\text{cont}_{\text{soprano}}$</td>
<td>.13</td>
<td>.0013</td>
</tr>
<tr>
<td>3</td>
<td>.76</td>
<td>.73</td>
<td>(4, 29)</td>
<td>$T_{\text{hier}}$</td>
<td>-.1</td>
<td>.0075</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\alpha$</td>
<td>.11</td>
<td>.0268</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\text{cont}_{\text{melody}}$</td>
<td>.13</td>
<td>.0046</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\text{cont}_{\text{soprano}}$</td>
<td>.09</td>
<td>.0101</td>
</tr>
<tr>
<td>4</td>
<td>.98</td>
<td>.97</td>
<td>(5, 28)</td>
<td>$T_{\text{hier}}$</td>
<td>-.48</td>
<td>$&lt; .0001$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\alpha$</td>
<td>.7</td>
<td>$&lt; .0001$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\text{cont}_{\text{melody}}$</td>
<td>.09</td>
<td>.0001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\text{cont}_{\text{bass}}$</td>
<td>.25</td>
<td>$&lt; .0001$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\text{cont}_{\text{soprano}}$</td>
<td>.79</td>
<td>$&lt; .0001$</td>
</tr>
</tbody>
</table>

Hereinafter, we adopt Ratner’s (2009) approach to interpret the results of linear regressions. That is to say, when analysing the degrees of linear relationship, we will refer to weak correlation when $0.0 \leq R^2_{adj} < .3$, moderate correlation when $.3 \leq R^2_{adj} < .7$ and strong correlation when $.7 \leq R^2_{adj} \leq 1$.

As seen in Table 7.6, the multiple regression of Theme 2 shows a moderate correlation, whereas the multiple regressions of the remaining themes show a strong correlation.

Figure 7.6 shows the graphical representations of the predictions of each multiple regression in Table 7.6 against the corresponding participants’ judgements.

Section 7.5.2 will discuss the impact that the results we have presented in this section have with regards to our Research Question, concerning long-term structure.
Figure 7.6: Graphical representation of the predictions of musical tension of each multiple regression in Table 7.6 against the corresponding participants’ judgements.
7.4 | Empirical study II

As discussed in Section 7.1.3, we have carried out an empirical study to evaluate the extent to which *autognomus* generates music that matches input tension profiles. Its participants, materials, procedure and results are presented in Sections 7.4.1, 7.4.2, 7.4.3 and 7.4.4, respectively.

7.4.1 | Participants

The same eighty-five participants we mentioned in Section 7.4.1 participated in the second empirical study. Table 7.2 is shown again below as Table 7.7. This table groups, again, participants by gender and by years of musical training.

Table 7.7: A summary of the participants that took part in the empirical study grouped by gender and years of musical training (repeated table).

<table>
<thead>
<tr>
<th>Gender</th>
<th>≥ 10 years of training</th>
<th>&lt; 10 years of training</th>
<th>No training</th>
<th>All participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>female</td>
<td>11</td>
<td>7</td>
<td>15</td>
<td>33</td>
</tr>
<tr>
<td>male</td>
<td>31</td>
<td>7</td>
<td>11</td>
<td>49</td>
</tr>
<tr>
<td>other</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>not said</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>TOTAL</td>
<td>43</td>
<td>15</td>
<td>27</td>
<td>85</td>
</tr>
<tr>
<td>Average years of training</td>
<td>21.3</td>
<td>4.3</td>
<td>0</td>
<td>16.4</td>
</tr>
</tbody>
</table>

7.4.2 | Materials

We generated a total of sixteen themes using *autognomus's Theme*. To do so, we fed *Theme* with the following inputs: number of measures equal to 8, tempo equal to 60 BPM, time-signature equal to 2/4 and character equal to allegro. Half of the themes were generated in the key of C major and the other half in the key of C minor. As in Section 7.3.2, since both keys share the same tonal centre, when changing from one theme to another in the study, participants would not perceive a great difference and so their perception of tension would not be greatly affected when listening to the new theme.
We then transformed the generated sixteen themes, using autognomus's Morpher, according to a collection of input tension profiles.

Inspired by the most common tension profiles in narrative structures, Lopes et al. (2016) propose seven main tension trends in the context of the generation of content for interactive experiences:

- an *escalating* trend, where tension increases,
- a *decreasing* trend, where tension decreases,
- a *surprising* trend, where tension reaches the highest peak in the profile,
- a *resting-point* trend, where tension reaches the lowest valley in the profile,
- a *cliffhanger* trend, where tension increases and its last degree is the highest one in the profile,
- a *denouement* trend, where tension reaches a high peak close to the end of the profile, but not exactly at the end, and
- an *unresolved* trend, where tension remains unchanged (i.e. the tension profile is flat, like a plateau).

Lopes et al. identified four tension profiles, designed from combinations of the above trends, that provided the most varied outputs when automatically generating music for the interactive video game Sonancia (Lopes et al., 2015). They named these tension profiles with regards to the most prominent trend they were based on, namely cliffhanger, decreasing, surprise and unresolved. We have adopted these tension profiles in our study and have adapted them to Morpher's tension range. The profiles are shown in Figure 7.7.
Figure 7.7: Input tension profiles Morpher is fed with.
There are two issues to be taken into consideration concerning the tension profiles in Figure 7.7. First, notice that they all start at tension zero. That is because the first chord in all themes generated by Themer will be a tonic chord. In this way, all profiles must either increase or remain flat at the very beginning, even in the decreasing tension profile. Second, considering how Morpher operates, whenever the degrees of tension are less than Morpher’s tension threshold, which recall is defined by default as equal to 8, Morpher will play the corresponding material of the original theme and will not consider the degrees of tension in the input profile.

The sixteen themes generated to study tension were split into four groups, each group containing two themes in a major key and two themes in a minor key. Each group was assigned with a tension profile, from those in Figure 7.7. We then used autognomus’s Morpher to transform the themes so that they matched the profile associated with their respective group. The new compositions for the cliffhanger, decreasing, surprise and unresolved tension profiles are shown in Figures A.2, A.3, A.4 and A.5, respectively, in Section A.1 in Appendix A.

The audio format of the sixteen morphed themes has the same characteristics introduced in Section 7.3.2. That is to say: at first, we saved the morphed themes in MIDI format (Moog, 1986) using MIDIUTil (Wirt, 2018); we then generated the scores of the MIDI files using Musescore 2.3.2 (Watson, 2018);\(^7\) finally, we used Musescore to convert the MIDI files into WAV format (Whibley et al., 2016), where we assigned the melodic voice in the morphed themes with Musescore’s sine synthesizer and assigned the four harmonic voices with Musescore’s string synthesizer.

As in Section 7.3.2, we did not alter the degrees of loudness and tempo in the sixteen morphed themes. Therefore, we will not consider these features in our analysis.

7.4.3 | Procedure

The procedure of the second study was the same as that previously described in Section 7.3.3. That is to say: the study was carried out online; participants were asked to listen to the
sixteen morphed themes introduced in Section 7.4.2 and to record the degree of tension they perceived, in real time, using a virtual slider; we did not discuss with participants our definition of tension; and the slider values were collected every 100 milliseconds. For more details about the experimental interface materials, see Section A.2 in Appendix A.

7.4.4 | Results

The results of the empirical study to investigate whether the generated music matches input tension profiles are presented in Section 7.4.4.4. The evaluation methodology used in this section is based on the pre-processing we applied to the raw data for the degrees of inter-subject agreement and for the time lags in the responses. Sections 7.4.4.1, 7.4.4.2 and 7.4.4.3 discuss how we dealt with these issues, respectively.

7.4.4.1 | Data pre-processing

In the second study, we used the same pre-processing strategy previously described in Section 7.3.4.1. That is to say, we re-sampled the participants’ judgements to a data point every 500 milliseconds, which we then standardised.

7.4.4.2 | Inter-subject agreement

As in Section 7.3.4.2, we calculated the degrees of inter-subject agreement according to Cronbach’s α and ICC (2, k). Table 7.8 shows the degrees of inter-subject agreement for all eighty-five participants across the sixteenth morphed themes used in the study.

As in Table 7.4, in Table 7.8 the degrees of agreement calculated using Cronbach’s α and ICC (2, k) are almost identical and so, again, we consider them to be reliable indicators of inter-subject agreement.

According to Cicchetti and Sparrow’s (1981) convention, fifteen pieces used in the study show excellent agreement in Table 7.8 and the remaining piece shows good agreement. Because of this, we decided to use the average of the judgements across all eighty-five participants, for each piece, for all subsequent analysis. For more details about the judgements of tension provided by the participants, see Section A.3 in Appendix A.
Table 7.8: Degrees of inter-subject agreement for all eighty-five participants across the sixteen pieces used in the study.

<table>
<thead>
<tr>
<th>Profile</th>
<th>Piece</th>
<th>Cronbach’s α</th>
<th>Intra-class Correlation Coefficient Confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>ICC (2, k)</td>
</tr>
<tr>
<td>cliffhanger</td>
<td>1</td>
<td>.99</td>
<td>.99</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.99</td>
<td>.99</td>
</tr>
<tr>
<td>decreasing</td>
<td>1</td>
<td>.96</td>
<td>.96</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.95</td>
<td>.95</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.95</td>
<td>.95</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.94</td>
<td>.94</td>
</tr>
<tr>
<td>surprise</td>
<td>1</td>
<td>.99</td>
<td>.99</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.99</td>
<td>.99</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.99</td>
<td>.99</td>
</tr>
<tr>
<td>unresolved</td>
<td>1</td>
<td>.96</td>
<td>.96</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.97</td>
<td>.97</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.9</td>
<td>.89</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.96</td>
<td>.96</td>
</tr>
</tbody>
</table>

7.4.4.3 | Time lag

As in Section 7.3.4.3, we used a temporal lag of 2 seconds to interpret the participants’ judgements, which recall is line with Farbood and Upham’s (2013), Krumhansl’s (1996) and Schubert’s (2004) conclusions.

7.4.4.4 | Results to investigate whether input tension profiles are matched

Our goal is to investigate the extent to which the music generated by autognomus matches input tension profiles. Once again, the above problem is similar to one already investigated by Lerdahl and Krumhansl (2007). In one of their studies, Lerdahl and Krumhansl recorded participants’ judgements of tension using both the stop-tension and continuous-tension tasks (i.e. in the former, participants judged each event one by one, with the music stopping at each event of study, whereas, in the latter, participants judged the whole pieces of music using a slider in real time). To determine whether the continuous
judgements fit the discrete ones they calculated a linear regression.

In our work, the input degrees of tension input into Morpher, those in Figure 7.7, consist of a collection of discrete tension data points, similar to Lerdahl and Krumhansl's stop-tension data, whereas the average of the participants' judgements consists of a continuous profile, similar to Lerdahl and Krumhansl's continuous-tension data. Therefore, to determine whether the latter fit the former, we have calculated a linear regression for each piece.

As discussed in Section 7.4.2, the four input profiles include fragments where the input degrees of tension are smaller than Morpher's tension threshold. Thus, in the sixteen pieces transformed by Morpher, only when the input degrees of tension are above the threshold has there been a transformation to make the music match input tension profiles. Therefore, only the data concerning the degrees of tension that are greater than Morpher's tension threshold are relevant when calculating the linear regressions against participants' judgements.

Table 7.9 includes the results of the calculated sixteen linear regressions, one per piece. The regressions' predictions and the participants' judgements are graphically represented in Figures 7.8, 7.9, 7.9 and 7.11.

As can be seen in Table 7.9, all four regressions concerning the cliffhanger, the decreasing and the surprise pieces, respectively, show a strong correlation. However, in the case of the unresolved pieces, the degrees of correlation are the poorest of them all. Two of the pieces show a moderate correlation (pieces 1 and 2) and the other two show a weak correlation (pieces 3 and 4). Notice that the regressions of the last two pieces are the only non-statistically-significant ones (i.e. \( p > .05 \)).
Table 7.9: Results of the linear regressions of the average participants’ judgements against the input tension profiles.

<table>
<thead>
<tr>
<th>Profile</th>
<th>Piece</th>
<th>$R^2$</th>
<th>$R^2_{adj}$</th>
<th>$df$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>cliffhanger</td>
<td>1</td>
<td>.99</td>
<td>.99</td>
<td>(1,20)</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.99</td>
<td>.99</td>
<td>(1,20)</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.99</td>
<td>.99</td>
<td>(1,20)</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.96</td>
<td>.96</td>
<td>(1,20)</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>decreasing</td>
<td>1</td>
<td>.94</td>
<td>.94</td>
<td>(1,15)</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.71</td>
<td>.69</td>
<td>(1,15)</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.77</td>
<td>.76</td>
<td>(1,15)</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.86</td>
<td>.85</td>
<td>(1,15)</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>surprise</td>
<td>1</td>
<td>.95</td>
<td>.95</td>
<td>(1,16)</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.94</td>
<td>.94</td>
<td>(1,16)</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.96</td>
<td>.96</td>
<td>(1,16)</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.97</td>
<td>.97</td>
<td>(1,16)</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>unresolved</td>
<td>1</td>
<td>.58</td>
<td>.56</td>
<td>(1,20)</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.54</td>
<td>.52</td>
<td>(1,20)</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.0008</td>
<td>-.05</td>
<td>(1,20)</td>
<td>.899</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.15</td>
<td>.11</td>
<td>(1,20)</td>
<td>.072</td>
</tr>
</tbody>
</table>
Figure 7.8: Graphical representation of the cliffhanger profiles regressions' predictions against participants' judgements.
Figure 7.9: Graphical representation of the decreasing profiles regressions' predictions against participants' judgements.
Figure 7.10: Graphical representation of the surprise profiles regressions' predictions against participants' judgements.
Figure 7.11: Graphical representation of the unresolved profiles regressions' predictions against participants' judgements.
Section 7.5.3 will discuss the impact that the results we have presented in this section have with regards to our Research Question, concerning tension.

### 7.5 | Discussion

Section 7.1 introduced three questions which have guided our evaluation. These questions concern the extent to which the Music Generator is applicable in real time and the extent to which the generated music has long-term structure and matches an input tension profile. Sections 7.5.1, 7.5.2 and 7.5.3 answer these questions based on the results previously presented in Sections 7.2.3, 7.3.4 and 7.4.4, respectively.

#### 7.5.1 | The extent to which the Music Generator is applicable in real time

Section 7.2 has introduced the computational test that we carried out to evaluate the extent to which autognomus is applicable in real time. The test consisted in calculating the degrees of correlation between the tension profiles of the first one hundred pieces in IGA’s dataset and the degrees of tension read by autognomus in real time. Results, which were summarised in Table 7.1 and graphically represented in Figure 7.4, show that the degrees of correlation are strong in all studied cases and that more than 90% of the calculated regressions are significant.

Some of the non-statistically-significant regressions that were recurrent in the computational test concern the tension profiles of the following pieces: L. van Beethoven’s Bagatelle In A Minor Für Elise, WoO 59; A. Ellmenreich’s Musikalische Genrebilder number 4, Op.14; and H. Necke’s Csikós Post. We analysed the pieces’ tension profiles and observed that the three of them include many peaks of tension in short periods of time, unlike the other pieces. We believe this issue made autognomus’s tension-reading process less effective and so their regressions were not statistically significant.

Based on the results of the computational test and on the analysis of the three above pieces, we conclude that autognomus is capable of adapting to real-time changing sce-
narios, as long as the input tension profiles do not contain too many steep transitions.

7.5.2 | The extent to which the generated music has long-term structure

Section 7.3 has introduced the empirical study that we carried out to evaluate the extent to which autognomus generates music that has long-term structure. A group of participants were asked to record, using a slider in real time, the tension they perceived while listening to four themes generated with autognomus’s Themer. Results were summarised in Table 7.6 and graphically represented in Figure 7.6.

In Table 7.6, the average of the judgements provided by the participants strongly fit the degrees of tension of the four themes used in the study, with values of $R^2_{adj}$ equal to or greater than .67. One may think that, in order for the regressions to show that the predicted patterns in the themes are perceived as such by the listeners, the values of $\beta_t$ must be positive. However, that is not the case. Recall that $T_{hier}$ and $\alpha$ may be complementary. For example, as discussed in Section 2.1.1, a perfect cadence (i.e. $V \rightarrow I$) is associated with a large degree of $\alpha$, which relates to the expectation for resolution that arises after the dominant chord, but is also associated with a small degree of $T_{hier}$, which relates to the resolution to the tonic chord. Notice how the values of $\beta_t$ and $\beta_a$ in Table 7.6 can be preceded by opposite signs; that is to say, when one is positive the other is negative and vice versa. Likewise, these values show that these features strongly contribute to predicting tension in the themes (i.e. the respective contributions of $T_{hier}$ and $\alpha$ to the total predicted tension, according to the regressions’ $\beta_t$ and $\beta_a$, correspond to: 30% and 20%, for the first theme; 10% and 60%, for the second theme; and 20% and 30% for the third and fourth themes). Therefore, in line with Lerdahl and Krumhansl’s (2007) conclusions, the hierarchical structures, and so the musical patterns, intended by autognomus in the generated four themes have likely been perceived as such by the participants.

Going back to Hall and Pearce’s (2021) ideas, previously introduced in Section 7.3.4.4, we have now shown that autognomus matches the tonal structures involved in their inter-
pretation of long-term structure. Recall that, based on autognomus’s design, we argued that it also matches the thematic structures. We then conclude that autognomus can convey some sense of long-term structure.

Determining the extent to which long-term structure is conveyed is a challenging task. Hall and Pearce (2021, p.220) summarise the effects of long-term structure of a piece of music by the term coherence, which they conceive as “the extent to which all elements of a piece can be considered to form a unified whole”. We believe that there does not exist yet a fully developed methodology to empirically determine whether the combination of thematic and tonal structures is coherent. For instance, Huang et al. (2019, p.9) compared different generation algorithms, which all focus on long-term structure, by asking human listeners to rate, using a Likert scale, which of the algorithms’ outputs “is more musical”. We, however, believe that this type of evaluation methodology does not shed enough light into the problem at hand. Developing an appropriate evaluation would need a thorough review of the approaches to defining long-term structure and a comprehensive framework that includes what to ask participants. Such is the complexity of this problem that it could be the sole focus of a whole new dissertation and so exploring this approach is beyond the scope of this chapter. Therefore, in this dissertation, we conclude that autognomus can generate music with long-term structure, but additional experimentation is needed to determine the extent to which this structure is coherent.

7.5.3 | The extent to which the generated music matches input tension profiles

Section 7.4 has introduced the empirical study that we carried out to evaluate the extent to which autognomus generates music that matches input tension profiles. A group of participants were asked to record, using a slider in real time, the tension they perceived while listening to sixteen pieces generated with autognomus’s Morpher. The pieces were generated so that they matched the input tension profiles shown in Figure 7.7. Results were summarised in Table 7.9 and graphically represented in Figures 7.8, 7.9, 7.10 and 7.11.
In Table 7.9, it was shown that the average of the judgements provided by the participants strongly fit the degrees of tension of the twelve pieces transformed to match the cliffhanger, the decreasing and the surprise profiles, with values of $R^2_{adj}$ equal to or greater than .76. Thus, we conclude that autognomus is capable of generating music that matches these types of input profiles.

The degrees of correlation of the four pieces transformed to match the unresolved profile are very weak and, what is more, two of the regressions are not statistically significant. In Farbood’s (2012) online study, it was observed that listeners only associated a flat tension profile with pieces of music whose pitches remained unchanged. Recall, however, that Morpher’s implementation constraints do not allow the generation of repeated chords to avoid boredom. Based on the poor degrees of correlation, this decision has proven not to be effective and so Morpher’s flat-tension algorithm may need to be refined. One way to improve the algorithm could be to incorporate time in the calculation of Morpher’s tension threshold; that is to say, if the input degree of tension remains unchanged for a certain amount of time, instead of repeating the same chords and melodies ad infinitum, Morpher could start playing a new theme, previously generated by Themer, whose relation to the original theme matches the unchanged degree of tension.

### 7.6 | Summary and Conclusions

In order to seek answers to our Research Question, Chapter 7 has evaluated the performance and applicability of autognomus.

Section 7.1 has argued that to evaluate autognomus three questions must be answered: to what extent is the Music Generator applicable in real time? To what extent does the generated music have long-term structure? And, to what extent does the generated music match input tension profiles?

Section 7.2 has approached the first question by means of a computational test. In the test, the degrees of tension of the first one hundred pieces in the Interactive GTTM Analyser (IGA) dataset were fed as input into autognomus and they were compared to the degrees of tension read by autognomus in real time. Results have shown that, in most
cases, the degrees of correlation are strong.

Sections 7.3 and 7.4 have approached the remaining two questions by means of two empirical studies. In the studies, eighty-five participants listened to four and sixteen pieces of music, respectively, generated with *autognomus*. They were asked to record, in real time using a virtual slider, the degrees of tension they perceived. Participants’ data were processed and analysed based on a combination of the methodologies previously developed by Lerdahl and Krumhansl (2007) and by Farbood (2012). Results of our both empirical studies have shown strong degrees of correlation between judged and predicted/input tension profiles.

Section 7.5 has discussed the implications of the conducted experiments with regards to our Research Question. Concerning the computational test, it was concluded that *autognomus* is applicable to real-time scenarios as long as the input tension profiles do not contain too many steep transitions. Concerning the first empirical study, it was concluded that *autognomus* can generate music with long-term structure. It was, however, observed that additional experimentation is needed to determine whether the achieved long-term structure is coherent. Concerning the second empirical study, it was concluded that *autognomus* can generate music that matches input tension profiles, as long as the tension in the profiles does not remain unchanged.
Conclusions

Dissertation review, research insights and contributions, limitations and future directions

In this chapter, we draw conclusions from the conducted research in relation to our goals. To do so, we will:

(1) sum up the conducted research,

(2) draw the main conclusions acquired from our research,

(3) point out the main original contributions, and

(4) discuss the limitations of the conducted research and propose directions for future work.

The first step is addressed in Section 8.1, the second step is addressed in Section 8.2, the third step is addressed in Section 8.3 and the fourth step is addressed in Sections 8.4 and 8.5, which cover the limitations and future work, respectively.
8.1 | Dissertation review

In Chapter 1, our work was motivated in terms of the difficulty of introducing long-term structure and tension into automatically generated music, and arguing that there was value in improving those aspects of automatically composed music. In order to take a step forward towards filling this gap, Chapter 1 defined our research approach. The approach focused on meeting three goals: generating music that matches input tension profiles, that has long-term structure and that is generated in real time. Chapter 1 also pointed out that, in order to meet our goals, three challenges must be taken into consideration: first, that the concept of musical tension had to be thoroughly dissected; second, that the generation approach must be based on a balanced focus between musical tension and long-term structure; and, third, that for theoretical models of tension and computational generation methods to suit our goals they must be applicable in real time. We phrased the pursuit of our goals, while considering the challenges, in the form a Research Question: *How can tonal music be generated in real time so that it has long-term structure and matches a given tension profile?*

In Chapter 2, we identified Lerdahl's Model of Tonal Tension (MTT) as the model best suited to our goals. To arrive at this conclusion, we framed the scope of and defined the concept of musical tension as the listeners’ reactions to musical expectations. We then reviewed the main approaches to modelling the concept of musical tension from the point of view of the adopted methodologies, which include theoretical accounts, empirical studies and generation models of musical tension. Finally, we identified the requirements that a model of musical tension should meet to suit our goals. According to these requirements, for a model to be suitable it must: be consistent with our definition of tension; provide a method to calculate quantitative values of tension; have been empirically tested and have shown strong correlations against the degrees of musical tension perceived by human listeners; consist of rhythmic, melodic and harmonic components; consider both pre- and post-outcome responses; take into account a piece’s structure; and, be applicable in real time.

In Chapter 3, we reviewed Lerdahl’s work. First, we contextualised his work, intro-
duced the components of his model of tension and discussed how they were derived. We then thoroughly reviewed the components: Generative Theory of Tonal Music (GTMM), Tonal Pitch Space (TPS) and Model of Tonal Tension (MTT). In order to illustrate their application, we provided a collection of examples based on the first phrase of the theme in Mozart’s *Ah vous dirai-je, Maman*, K. 265/300e, to which we applied the components. Finally, we reviewed MTT’s evaluation.

In Chapter 4, we fully automated Lerdahl’s MTT. To do so, we first identified the main issues that posed a challenge in previous attempts at the automation of the model, which include the automatic calculation of TPS’s parameters $i$ and $j$, the computational representation of a piece’s hierarchical structure according to GTTM’s rules and the calculation of distances between chords in different tonal regions. In order to overcome these challenges, we designed and implemented two systems, which we have made publicly available. The first system, *AutoTPS*, is capable of automatically calculating distances between chords within TPS. The second system, *AutoTen*, is capable of automatically calculating tonal tension according to MTT’s rules for a given piece of music. Finally, we evaluated the accuracy of these systems using a total of one hundred test cases, all of which were successfully passed.

In Chapter 5, we determined that a music generation system, where statistical methods are combined with rule-based methods and generative grammars, was an effective computational strategy to meet our goals. To arrive at this conclusion, we defined the concept and framed the scope of Automatic Music Generation as the application of algorithmic composition generation techniques by computational means. We then developed a taxonomy of the methods in the field of Automatic Music Generation and briefly reviewed the characteristics of each method. From this review, we determined that statistical methods, rule-based methods and generative grammars were suited to meet our goals. Finally, we discussed the advantages of hybrid generation systems, concluding that an appropriate form of hybridisation to address our goals is that of generative grammars combined with statistical methods, as well as rule-based methods combined with statistical methods.

In Chapter 6, we covered the design and implementation of *autognomus*, our Music
Generator, in order to seek answers to our Research Question. *autognomus* is a system capable of generating tonal music that has long-term structure and that matches input tension profiles in real time, which we have made publicly available. *autognomus’s* design is inspired by the vertical mixing technique that is often used in video games’ soundtracks when a single theme is transformed in different ways to adapt to a narrative in real time.

We identified four tasks related to the vertical mixing technique: generating a main theme, arranging the voices in the theme’s harmony, developing new material based on the theme’s structure and morphing the theme and its developments so that they match changes in the given narrative. Based on these tasks, we designed *autognomus* as a music generation system that consists of four sub-systems: *Themer*, the first sub-system, responsible for the generation of a musical theme with a well-defined structure; *Arranger*, the second sub-system, responsible for the generation of an accompaniment, for the original theme, that matches common voice-leading conventions; *Developer*, the third sub-system, responsible for the generation of new musical material based on the structure and hierarchical relations of the original theme; and *Morpher*, the fourth sub-system, responsible for the transformation of generated materials, according to GTTM’s and MTT’s rules, so that they match input tension profiles in real time.

In Chapter 7, we evaluated *autognomus*. To do so, we defined a methodology for evaluating its performance and applicability with regards to our goals. The evaluation methodology consisted of a computational test, to evaluate the extent to which *autognomus* is applicable in real time, and two empirical studies, to evaluate the extent to which the automatically generated music has long-term structure and matches input tension profiles, respectively. The results of the computational test have shown that *autognomus* is applicable in real time as long as the input tension profiles do not contain too many steep transitions. Results of the first empirical study have shown that *autognomus* can generate music with long-term structure, although additional experimentation is needed to determine whether the achieved long-term structure is coherent. Results of the second study have shown that *autognomus* can generate music that matches input tension profiles, as long as the tension in the profiles does not remain unchanged.
8.2 | Research insights

We present below the main lessons we have learnt while seeking to answer our Research Question.

In Chapter 2, we have demonstrated the vagueness and ambiguity that surrounds the concept of musical tension. In that chapter, we have pointed out how important it is to find the appropriate definition of the concept of musical tension and have opened the door to a discussion about the nature and meaning of the concept.

In Chapter 4, we have shown how crucial it can be to apply the last version of Lerdahl's MTT when calculating the distance between distant chords within TPS. We identified in Chapter 3 how some scholars are still using outdated versions of MTT which could sometimes lead to results that do not accurately match listeners' perceptions.

In Chapter 7, we have proven that Lerdahl and Krumhansl's (2007) evaluation methodology can be expanded in line with that of Farbood (2012). To so so, contour-related features of the harmonic voices must be considered together with MTT's components when analysing the degrees of correlation in continuous-tension-related empirical studies. This evaluation methodology opens the door to replicating Lerdahl and Krumhansl's experiments to investigate whether using the expanded methodology leads to new conclusions.

Finally, in Chapters 6 and 7, we have shown that GTTM-based hierarchical structure and MTT-based tension can be used to automatically generate music that has long-term structure and that matches input tension profiles in real time. Our work has not, however, fully exploited the capabilities of Lerdahl's theories; it has just taken the first step. Using GTTM-based hierarchical structure and/or MTT-based tension seems to be a good way to develop new approaches to be followed in the field of Automatic Music Generation.

8.3 | Research contributions

Below are presented this dissertation’s main contributions. These concern insights learnt from the present research that may have an impact on the research community and could either lead to new research projects or suggest new directions of work in existing projects.
Chapter 2 has presented a novel definition of musical tension that encompasses the main existing interpretations of the concept that exist in the literature. This definition may come into conflict with some views of the concept from the perspective of musicology and those of the field of music cognition; however, it could be of use in human-centred investigations (i.e. whenever a tension-related investigation involves human listeners). Chapter 2 has also presented a novel taxonomy of the approaches to modelling musical tension from the point of view of the adopted methodology.

Chapter 3 has presented a thorough review of Lerdahl's work, including his Generative Theory of Tonal Music (GTTM), Tonal Pitch Space (TPS) and Model of Tonal Tension (MTT), as well as a collection of illustrative examples of the application of these theories. This review is probably the first of its kind. Most research projects based on Lerdahl’s work usually include a brief summary of the main ideas and, unless the readers are already familiar with these ideas, they would have to spend a considerable amount of time going through Lerdahl’s theories to fully understand them. Chapter 3 can be used by non-experts to acquire a reasonable understanding of Lerdahl’s work.

Chapter 4 has presented AutoTPS and AuToTen as computational automations of Lerdahl’s TPS and MTT, respectively, both of which we have made publicly available. These have demonstrated how to “tighten [Lerdahl’s] entire theory to the point that it is implemented computationally”, as Lerdahl (2009, p.193) proposed as a future prospect more than a decade ago. Likewise, the architecture of both AutoTPS and AuToTen is designed in such a way their components can be independently used or further developed by the research community.

Chapter 5 has presented a definition of the field of Automatic Music Generation as well as a taxonomy of the main methods in the field. The critical review of the methods with regards to their suitability concerning our goals (i.e. generation that focuses on musical tension, long-term structure and real-time applicability) adds an additional point of view to the way the methods are usually reviewed in the field, in taxonomies such as that of Nierhaus (2009) or Fernández and Vico (2013).

Chapter 6 has presented autognomus, a system capable of generating tonal music that has long-term structure and that matches input tension profiles in real time, which
we have made publicly available. autognomus has been shown to benefit from the main advantages of the field of Procedural Content Generation, which include generating original and endless material, providing inspiration and helping in reducing efforts and costs, among others. Likewise, autognomus could be used for many applications, such as assisting novice composers with their compositional process, automatically generating music in interactive environments, such as video games, or serving as a tool in tension-related research projects.

Chapter 7 has presented the first stages of a methodology for evaluating whether a piece of music has long-term structure. This methodology can be the starting point of new research that aims to develop a fully-formed long-term structure automatic analyser.

8.4 | Limitations

Some of the limitations of the present research have already been introduced in previous chapters. Most of these limitations concern the decisions we made while designing and implementing autognomus, the Music Generator, and the methodology we designed to evaluate autognomus’s capabilities. Sections 8.4.1 and 8.4.2 summarise the main of these limitations, respectively. However, there are other limitations that have not been discussed in detail so far. These have to do with the quality of the music generated by autognomus and the implications of the scope of the present research. Sections 8.4.3 and 8.4.4 address these limitations, respectively.

8.4.1 | The design and implementation of the Music Generator

In Chapter 6, many of the design decisions were made for the sake of simplicity with the idea that, once the simplest version of autognomus proved useful, more interesting and complicated features would be implemented. Because of this, the design and implementation of autognomus has some limitations.

In general, most of the limitations we are focusing on in this section are the result of the approach we followed to design and implement autognomus. Our approach is
rather deterministic, mostly because of the generation methods (i.e. statistical methods, rule-based methods and generative grammars) and the theories (i.e. Lerdahl's GTTM, TPS and MTT) the approach is based on. Some examples of the implementation of autognomus's sub-systems that pose a limitation include: concerning Themer, its pre-defined structures for the themes and the rules that control the repetition and transformation of musical ideas; concerning Arranger, the fact that it only generates block-chords (extending its functionality should not be a difficult task, although evaluating the outcome of all possible extensions may be time-consuming); concerning Developer, its rules that restrict the possible ways to develop the musical ideas in previously generated themes; and concerning Morph, the fact that it does not generate new musical material until the previous measure has been played in full, even if the input degree of tension has changed, or the fixed pre-composed melodic lines.

As a result of the above limitations, two key limitations of autognomus's capabilities were covered in Chapter 7: it was observed that autognomus may fail to adapt to some real-time scenarios as well as fail to match some input tension profiles, such as those where tension remains unchanged.

Another key limitation has to do with the concept of long-term structure. In Chapter 7, inspired by Hall and Pearce (2021), we restricted our view of the concept to musical ideas that repeat throughout a piece of music and are coherently organised within a hierarchy. To some scholars, the concept of long-term structure is also understood in terms of musical form. For instance, Aspromallitis and Gold (2016, p.1) identify long-term structure with “musical forms that might be traditionally recognised e.g. blues, pop songs, binary, tertiary etc.”. We believe that autognomus also satisfies this condition considering the way it organises the generated music: let A be a theme generated by Themer and B, C, D... its corresponding developments generated by Developer; as discussed in Chapter 7, the resulting piece of music generated by Morph would have the form ABACAD..., which shares some similarities with the rondo form. It was shown, however, in Chapter 7, that we were not able to test our form-related assumption with our evaluation methodology. We give more detail about this limitation in the next section.

Finally, the limitation which may be the most crucial of them all may be the extent
to which *autognomus* is creative. Despite the fact *autognomus* can be used to generate endless pieces of music, the truth is that, after some time, the generated pieces would start to be too similar to each other. What is more, *autognomus* is not capable of learning yet. Other existing music generation systems incorporate some way to use feedback to learn how to improve future generations. Because of this issue, we believe that the most important contribution made by this dissertation is the methodology used to design and to implement *autognomus* rather than the system itself.

### 8.4.2 | The evaluation methodology

Our evaluation methodology, the one we used in Chapter 7, has shown some key limitations. In the first empirical study, we only used four different themes. In the second empirical study, we only used four different input tension profiles. Likewise, the pieces of music we used in both studies are just a few seconds long. We made these decisions for the sake of effectiveness and feasibility. Using more themes, more input tension profiles and/or longer pieces of music would translate into longer experimental sessions. This could have had an impact on the participants’ responses, and so could have affected our results and our conclusions. We believe that, despite these limitations, the results we gathered from our empirical studies provide enough information to evaluate the key capabilities of *autognomus*, with regards to our Research Question.

In Chapter 7, from the results of the first empirical study, we concluded that *autognomus* can generate music that consists of well-identified patterns that are repeated and transformed within a well-defined hierarchical structure. We, however, also pointed out that a fully developed methodology to evaluate long-term structure in music is much needed, but does not exist yet. That is why, considering the points we made in the previous section about our understanding of the concept of long-term structure, we concluded that *autognomus* can generate music with long-term structure although additional experimentation is needed. The additional experimentation should focus on evaluating whether listeners believe that the music generated by *autognomus* organises its patterns coherently and whether an ABACAD form is perceived.
Another limitation of our evaluation methodology concerns its lack of attention to the effects loudness and tempo could have on perceived tension with regards to the music generated with autognomus. We did not investigate these features in our empirical studies so that we could make sure the results concerned the melodic, rhythmic and harmonic components in Lerdahl’s theories, on which autognomus is based.

Finally, it is important to address the fact that we did not discuss our definition of tension with participants in the listening studies. In the past, in tension-based listening studies, some scholars have discussed a definition of the concept of tension with the participants at the beginning of their studies. Examples of this approach include the studies carried out in Bigand et al. (1996, p.132) or Farbood and Price (2017, p.422). The former did it so to “familiarize the participants with the notion of musical tension”. The latter came up with a definition of musical tension whose wording was designed to “be broad without resorting to circular definitions (e.g., “more tension” described as “feeling more tense”) as well as avoiding terminology that evoked specific auditory percepts such as loudness (e.g., “intensity”).

We decided not to discuss our definition of the concept of tension with participants in our studies for three reasons. First, according to Lehne et al. (2014, p.1517), discussing a definition of the concept with participants can bias their understanding of the concept “towards one specific aspect of the tension experience (e.g. consonance/dissonance, stability/instability, local vs non-local implicative relationships)”. Second, lots of others scholars, such as Krumhansl (1996; 1997), Lerdahl and Krumhansl (2007) or Farbood (2012), have found high degrees of inter-subject agreement in their tension-related studies without discussing a definition of the concept of musical tension with participants. Third, and most importantly, this approach is the one that best matches the motivations we presented in Chapter 1. With this dissertation we aim to improve the use of automatically generated music in interactive experiences, such as video games. These experiences would include tension-changing scenarios, but they would rarely provide a definition of the concept. The automatically generated music would have to match tension regardless of the listeners understanding of the concept.

The discussion about whether a definition of the concept of tension should be pro-
vided in tension-related empirical studies is far from being solved. New lines of research are needed to investigate the issue further. Mainly, we believe the new research should focus on: determining the extent to which providing or not a definition of the concept of musical tension could bias listeners’ perception (e.g. by carrying out a comparative study); analysing whether different definitions of the concept of musical tension (e.g. relating the concept to expectation, surprise, suspense, stress or energy) result in different tension responses for the same pieces of music; and investigating whether the provision of a particular definition of tension could make listeners differentiate between the tension they would assume the music is supposed to express and the actual tension they experience (recall the discussion about this issue in footnote 6 on page 21).

8.4.3 | The quality of the automatically generated music

In Chapter 2, we identified Lerdahl’s MTT as the model of tension best suited to our goals. One of the main reasons to do so was that, by incorporating GTTM, the model accounts for the hierarchical structure of a piece of music. We believe this is a critical requirement because of two reasons: it provides us with a way of having control over long-term structure, which is one of our main goals, and it accounts for the hierarchical organisation of music, which has proven to be an essential element in describing listeners’ perception of musical tension (Bigand and Parncutt, 1999; Bigand et al., 1996; Farbood, 2012; Lerdahl and Krumphantsl, 2007).

Choosing Lerdahl’s MTT as our main theoretical reference carried some limitations. Most notably, it restricted the generation to homophonic music and to the tonal context. We also restricted the implementation of autognomus so that its outputs would be simple accompanied melodies. By simple we mean that: the generated themes are purely diatonic; the chords are played in blocks; the generated metrical patterns are regular; and the melodic contours are not intricate. We made this decision because of the complexity of Lerdahl’s theories, because of the ambiguity of some of its rules and because this is the first time his three main theories have been used to automatically generate music. Most notably, we also made this decision in order to make sure we were able to model tension
as accurately as possible. As a result, the style of the music generated with autognomus may not match the music styles commonly used in tension-driven experiences. And, what is more, the music generated with autognomus sometimes violates some common rules of the tonal tradition so that it matches the input tension profile while obeying the imposed computational rules (see, for instance, the pieces in Section A.1 in Appendix A, which were used in the studies in Chapter 7).

In order to overcome the above limitations, a different theoretical model might have been more productive in generating music with a different and more appropriate style. For instance, the use of neo-Riemannian theory in game and film music is quite extensive (Lehman, 2014). Neo-Riemannian theory characterises the relations between chords and harmonic progressions, as does TPS, but gives particular attention to chromatic music (Cohn, 2011).

What would have been the advantages and disadvantages of having used neo-Riemannian theory in the context of this dissertation? On one hand, as discussed above, the style of the generated music would have probably been more appropriate for the application of interactive experiences. Likewise, since neo-Riemannian theory is based on small-scale transformations, its implementation would have probably posed less restrictions. On the other hand, neo-Riemannian theory does not explicitly address the concept of musical tension. Lerdahl and Krumhansl (2007) developed a promising method to estimate tonal tension sequentially using neo-Riemannian theory, in the context of chromatic music. They observed, however, that this method is likely to be problematic when dealing with modulations and when calculating (tension-related) attractions. Also, neo-Riemannian theory does not address long-term structure.

8.4.4 | The scope of the present research

To conclude the limitations section, we would like to address the issue of whether the music automatically generated with autognomus is ideologically neutral. Our system is based on GTTM, TPS and MTT principles, and so it is constrained to the Western tonal musical tradition. This tradition is based on theories that not only theorise about music
but also imply things about society. This is because of the impact that culture, society, philosophy and racial supremacy have had on the development of these theories and, therefore, on the Western tonal musical tradition (Cook, 2007; Ewell, 2021). Because of this, we would like to acknowledge the fact that autognomus is limited to the framework in which it has been implemented; that is, within some conventions of the Western tonal musical tradition. And, most importantly, it should be noted that, when generating music, autognomus’s choices of what may be appropriate are therefore limited to whatever falls within this framework. We believe the reader should keep this in mind when using autognomus.

8.5 | Future work

Based on the limitations introduced in the previous section, two directions for future work are proposed below for each of our original goals: musical tension, long-term structure and real-time applicability.

Directions of future work concerning musical tension include:

- carrying out a comparative study of the different interpretations of the concept of musical tension and expectation and comparing their respective degrees of correlation against the degrees of tension perceived by human listeners, and

- building a dataset that contains tension profiles of real pieces of music (this type of dataset can be used to improve autognomus’s learning capabilities or to implement other generation systems according to autognomus’s theoretical principles but based on machine learning techniques, among other applications).

Directions of future work concerning long-term structure include:

- developing a human-centred methodology (i.e. that involves human listeners) for evaluating whether a piece of music has a coherent long-term structure, and

- testing Developer’s capabilities and limitations using the newly developed evaluation methodology.
Directions of future work concerning real-time applicability include:

- implementing autognomus within an interactive scenario, such as for instance as the music engine in a video game, and
- evaluating whether the music automatically generated with autognomus improves people’s immersion in real-time interactive experiences.

8.6 | Closing remarks

In the last two decades, automatically generated music has proven immensely useful for generating endless musical ideas, for providing inspiration to composers and for reducing efforts and costs. Despite these benefits, it is still not used in our day to day lives. This dissertation has investigated how to take a step forward towards filling this gap. In doing so, we have demonstrated how the application of high-level cognitive features, such as musical tension and long-term structure, can be the point from which to start building the next generation of automatic music generation systems. If a single message had to be taken away from the present research, that would be the one.

The problem of automatically generating music, of quality and efficiently, is far from being solved, but it is our hope that this dissertation takes us closer.
Appendix A

Experimental materials and additional data

This appendix includes the materials and some additional data concerning the empirical studies discussed in Chapter 7. Section A.1 includes the music materials used in the empirical studies. Section A.2 includes the information provided to the participants in the empirical studies. Section A.3 includes the graphical representations of the judgements of tension provided by the participants in the empirical studies.

A.1 | Experimental music materials

The scores of the four pieces of music used in the first empirical study are shown in Figure A.1. The score of the sixteen pieces of music used in the second empirical study are shown in Figures A.2, A.3, A.4 and A.5, each of which refers to the four pieces that were generated concerning the cliffhanger, the decreasing, the surprise and the unresolved tension profiles, respectively.

For the sake of simplicity, both studies were carried out in one go. The pieces were presented to the participants in the following order:
(1) cliffhanger piece 1 (Figure A.2a)

(2) surprise piece 2 (Figure A.4b)

(3) theme 3 (Figure A.1c)

(4) theme 2 (Figure A.1b)

(5) decreasing piece 2 (Figure A.3b)

(6) unresolved piece 4 (Figure A.5d)

(7) surprise piece 4 (Figure A.4d)

(8) theme 1 (Figure A.1a)

(9) decreasing piece 1 (Figure A.3a)

(10) unresolved piece 2 (Figure A.5b)

(11) cliffhanger piece 3 (Figure A.2c)

(12) decreasing piece 4 (Figure A.3d)

(13) cliffhanger piece 2 (Figure A.2b)

(14) theme 4 (Figure A.1d)

(15) surprise piece 3 (Figure A.4c)

(16) unresolved piece 3 (Figure A.5c)

(17) surprise piece 1 (Figure A.4a)

(18) decreasing piece 3 (Figure A.3c)

(19) unresolved piece 1 (Figure A.5a)

(20) cliffhanger piece 4 (Figure A.2d)
Figure A.1: Score of the four themes generated with Themer that were used in the first empirical study.
Figure A.2: Score of the four pieces generated with Morpher, based on the cliffhanger profile, that were used in the second empirical study.
Figure A.3: Score of the four pieces generated with Morpher, based on the decreasing profile, that were used in the second empirical study.
Figure A.4: Score of the four pieces generated with Morpher, based on the surprise profile, that were used in the second empirical study.
Figure A.5: Score of the four pieces generated with Morpher, based on the unresolved profile, that were used in the second empirical study.
A.2  |  Experimental interface materials

At the beginning of the empirical studies, participants were presented with Figure A.6, which includes the relevant information about the studies. Next, participants were asked to read the consent form in Figure A.7 and to agree with its terms if they wanted to take part in the studies. The questionnaire in Figure A.8 was then presented to the participants. Finally, participants were asked to listen to the twenty pieces introduced in Section A.1 and to record the tension they perceived in real-time using a slider, as shown in Figure A.9.
Appendix A. Experimental materials and additional data

Interface materials

Introduction

You are being invited to take part in a research study. Before you decide whether or not to take part, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully.

What is the purpose of the study?
This research project aims to study the degree of tension people perceive while listening to computer-generated music.

Who can take the study?
Anyone older than 18 years old with no hearing impairments can be a participant.

What will I be asked to do if I agree to take part?
You will be played a collection of musical cues, each of which will have a duration of 20 seconds. You will be asked to record the degree of tension you perceive, in real time, while listening to the music.

What device should I use to take the study?
You should use a laptop or PC with a mouse or a similar device.

Am I being tested?
No, you are not. In this study, you will just record how tense you perceive some music, so there are no right or wrong answers.

How long will the study take?
The study will take around 15 minutes.

Who is running the study?
A research team from the school of Computing and Communications at the Open University (UK).

How can I contact them?
You can reach us at: german.ruiz-marcos@open.ac.uk

Many thanks for taking the time to read this information!

Last update: 20/01/2021

Figure A.6: Information about the empirical study participants were first presented with.
Appendix A. Experimental materials and additional data

Interface materials

Informed consent

By taking this survey, you will be agreeing with the following terms:

■ You understand that taking part in the study involves listening to several pieces of music and recording the degree of tension you perceive.
■ You consent voluntarily to be a participant in this study.
■ You understand that you are still free to withdraw at any time during the study and without giving a reason.
■ You understand that the study has been approved by the OU Human Research Ethics Committee.
■ You understand that any data you provide will be anonymised so that it will not be possible to link it back to you.
■ You understand that the information you provide will be used as part of the lead researcher’s thesis, as well as in technical reports and academic publications, but will always be anonymised.

This research project has been reviewed by, and received a favourable opinion, from the OU Human Research Ethics Committee.
HREC reference number: HREC/3529/Ruiz-Marcos
http://www.open.ac.uk/research/ethics/

The Open University is the Data Controller for the personal data that you provide. The lawful reason for processing your data will be that conducting academic research is part of the Open University’s public task. If you are concerned about the way we have processed your personal information, you can contact the Information Commissioner’s Office (ICO). Please visit the ICO’s website for further details.

Do you agree to the terms and conditions shown above?

☐ Yes, I do.

Figure A.7: Consent form participants were presented with and must agree with to proceed.
Appendix A. Experimental materials and additional data

Interface materials

What do we need to know about you?

1. How old are you? ____
2. What is your gender?
   - Female
   - Male
   - Other
   - Prefer not to say
3. Are you musically trained (either at music school/conservatory or self-taught)?
   - Yes
   - No
4. If so, how many years have you spent studying music? ____

Figure A.8: Questionnaire participants were presented with to collect data about them.

How do you feel tension behaves in the following piece of music?

Start by doing click on the file, as shown by the animation. Then click on the PLAY button at the bottom left corner. Once the file has started to play, please scroll the slider to the right when you feel tension is increasing or to the left when you feel tension is decreasing.

Figure A.9: Interface participants were presented with to record their judgements of tension of the twenty pieces of music used in the empirical study.
A.3 | Graphical representations

The graphical representations of the degrees of tension provided by the participants when they listened to the pieces shown in Section A.1 are shown in Figures A.10, A.12, A.14, A.16 and A.18, which represent the averages of all eighty-five participants together, and in Figures A.11, A.13, A.15, A.17 and A.19, which represent the averages with regards to two different groups into which where participants are split depending on their years of musical training. In these figures, the shaded areas represent the confidence intervals.
Figure A.10: Average judgements of tension, of all 85 participants, corresponding to the four themes used in the first study.
Figure A.11: Average judgements of tension, grouped by training, corresponding to the four themes used in the first study.
Figure A.12: Average judgements of tension, of all 85 participants, corresponding to the cliffhanger pieces used in the second study.
Figure A.13: Average judgements of tension, grouped by training, corresponding to the cliffhanger pieces used in the second study.
Figure A.14: Average judgements of tension, of all 85 participants, corresponding to the decreasing pieces used in the second study.
Figure A.15: Average judgements of tension, grouped by training, corresponding to the decreasing pieces used in the second study.
Figure A.16: Average judgements of tension, of all 85 participants, corresponding to the surprise pieces used in the second study.
Figure A.17: Average judgements of tension, grouped by training, corresponding to the surprise pieces used in the second study.
Figure A.18: Average judgements of tension, of all 85 participants, corresponding to the unresolved pieces used in the second study.
Figure A.19: Average judgements of tension, grouped by training, corresponding to the unresolved pieces used in the second study.
Glossary of musical terms

agogics “those aspects of performance related to duration, and by extension tempo, in the way that dynamics are related to loudness; thus, the use of rubato or other departures from strictly notated durations.” (Randel, 2003, entry: Agogic, definition: 2). 26

appoggiatura “[a]n ornamental note falling on the beat, that is, one that replaces the main note at the moment of its attack, then resolves to the pitch of that note” (Randel, 2003, entry: Appoggiatura [It.], definition: 2). 160

cadence “[a] melodic or harmonic configuration that creates a sense of repose or resolution. Cadences thus most often mark the end of a phrase, period, or complete composition (...). The cadences of Western tonal music are usually classified through the harmonic analysis of their constituent elements and, to a lesser extent, according to the voice leading of the highest and lowest parts. The names that have been assigned to these cadences are for the most part an accumulation of historical accidents. Cadences are the principal means by which tonal music projects the sense of one pitch as a central or tonic pitch in a passage or work. The strongest cadence in tonal music is the progression from the dominant harmony to the tonic harmony, V–I, and is termed an authentic cadence (...). The progression IV–I is termed a plagal cadence (...). A deceptive cadence (also termed interrupted) is one in which the dominant is followed by a harmony other than the tonic, most often VI, but sometimes IV or some other harmony instead (...). A half cadence (also termed a

chord “[t]hree or more pitches sounded simultaneously or functioning as if sounded simultaneously.” (Randel, 2003, entry: Chord [Fr. accord; Ger. Akkord, Zusammen- klang; It. accordo; Sp. acorde], definition: 1). xvii, 4, 10–13, 15, 16, 18, 22, 26, 28, 30, 34, 36, 45, 46, 59–61, 66–76, 78, 81, 89–95, 97–99, 105–109, 125, 126, 133, 134, 139–142, 144–146, 148, 149, 151–154, 156–158, 164, 166, 168, 169, 171–173, 175, 180–191, 199, 200, 204, 206, 208, 211, 230, 242, 247, 249

chord degrees in this dissertation, we use this term to refer to chords built upon a specific scale degree (see scale degrees). 74, 95

chord function “[i]n tonality, chords relate to each other and to a central harmony, the tonic triad; they are identified by the root upon which they are built and by the intervallic relationships between the root and the other pitches. The identification of a chord by the scale degree (indicated with a roman numeral) that is its root is called the function of the chord. (...) [In the key of C major] the C-major triad is labeled with the key designation C and the roman numeral I, indicating that it is formed on the first scale degree in C major and is thus the tonic triad in C; it has tonic function in C. (In the key of F major, this same triad would have dominant function, V; in G major, it would have subdominant function, IV.)” (Randel, 2003, entry: Harmony [fr. Gr., Lat. harmonia; Fr., Ger. Harmonie; It. armonia; Sp. armonía], definition: 1, II. Chord function). 282, 283, 290, 292–294

chord label see chord function, harmonic analysis and scale degrees. xii, xvi, 66, 67, 70, 77, 98, 101, 105, 107, 108, 128, 135, 144–147, 151, 152, 156, 168, 171, 176, 178, 186, 190

chromatic “[h]armony or melody that employs some if not all of the pitches of the chromatic scale in addition to those of the diatonic scale of some particular key, whether or not the harmony or melody in question can be understood within the context of

chromatic scale “[t]he scale that includes all of the 12 pitches (and thus all of the 12 semitones) contained in an octave, as distinct from the diatonic scale.” (Randel, 2003, entry: Chromatic [Gr., colored], definition: 1). 28, 68, 76, 186


diatonic scale “[a] scale with seven different pitches (heptatonic) that are adjacent to one another on the circle of fifths; thus, one in which each letter name represents only a single pitch and which is made up of whole tones and semitones arranged in the pattern embodied in the white keys of the piano keyboard.” (Randel, 2003, entry: Diatonic, definition: 1). xviii, 68, 76, 152, 159, 167, 192, 197

dissonance “[a] means of classifying the interval between two simultaneous notes. Very generally, consonant intervals are regarded as primary and stable, whereas dissonant intervals are regarded as secondary and unstable. Theorists use this distinction to explain the sense of motion that occurs within pieces of music.” (Randel, 2003, entry: Consonance and dissonance, definition: 1). 26, 34, 46, 60, 76, 77, 79

division see metre. xx, 157, 209, 288

dominant chord “[t]he triad and the seventh chord built on this degree [the fifth scale degree of the major or minor scale] as root are the dominant triad and dominant seventh, respectively.” (Randel, 2003, entry: Dominant, definition: 1). 11–13, 15, 36, 67, 140, 145, 173, 240, 283

dominant function see chord function, functional harmony and tonal function. 22, 108, 151

dominant seventh chord see dominant chord. 12, 13, 95, 152
eighth-note  [also known as quaver] see note value. 53, 163, 164, 218

enharmonic “[i]n modern theory, pitches that are one and the same even though named or ‘spelled’ differently, e.g., G♯ and A♭ or E and F♭.” (Randel, 2003, entry: Enharmonic, definition: 2) [embedded quote marks are originally written in Randel (2003) as "", but in this definition are written as ‘’ for the sake of clarity]. 90–92, 95–98, 108

flourish [also known as neighboring tone or auxiliary tone] “a tone a step above (...) or a step below (...) a consonant tone. (...) [N]eighboring tones may be either strong or weak metrically.” (Randel, 2003, entry: Counterpoint [Lat. contrapunctus, fr. contra punctum, against note; contrepunt; Ger. Kontrapunkt; It. contrappunto; Sp. contrapunto], definition: II. Dissonance treatment in tonal counterpoint, 2. Neighboring tone or auxiliary tone). 160

functional harmony “[a] theory of tonal harmony developed by Hugo Riemann according to which all harmonies can be analyzed as having one of three functions: tonic, dominant, and subdominant (designated T, D, and S, respectively, in analyses of this type). Scale degrees II, III, and VI are often interpreted as the relative minors of IV, V, and I, respectively, and thus as having the functions S, D, and T. III can also function as the ‘upper relative’ of V and thus have the function D, as does VII. The letters T, D, and S are added to in various ways to indicate chromatic alteration and the addition of dissonant tones. The term functional harmony is sometimes loosely applied to tonal harmony in general as it is understood in prevailing methods of harmonic analysis, which regard each of the seven diatonic scale degrees as having a separate function.” (Randel, 2003, entry: Functional harmony, definition: 1) [embedded quote marks are originally written in Randel (2003) as "", but in this definition are written as ‘’ for the sake of clarity]. 283, 290, 292, 294

half-note [also known as minim] see note value. 53, 162

harmonic analysis “[a]nalysis of harmonic functions and their relationship to the larger dimensions of a musical work. Analysis of Western tonal harmony, principally of
the 18th and 19th centuries, includes elucidation of chord types and their functional basis (expressed by roman-numeral notation); evaluation of their relative harmonic strength and their patterns within the phrase; the relationship of these to the key and to changes of key; and ultimately the larger and smaller relationships of individual and assembled harmonies to the overall structure of the work.” (Randel, 2003, entry: Harmonic analysis, definition: 1) . 282, 290, 293

**interval** “[t]he relationship between two pitches (...). For purposes of Western tonal music, intervals are named according to (1) the number of diatonic scale degrees included, as represented in the letter names of the two pitches, and (2) the number of semitones (the smallest interval in the Western system) between the two pitches. The former is expressed as a number, determined by counting the letters of the alphabet beginning with that of the lower pitch and including that of the higher (remembering that only the first seven letters are used and then repeated). Thus, c–c is a prime or unison, c–d a second, c–e a third, c–f a fourth, c–g a fifth, c–a a sixth, c–b a seventh, c–c’ an octave. Intervals larger than an octave can be named similarly (ninth, tenth, eleventh, etc.), though they are also known as compound intervals, since they can be thought of as consisting of an octave plus a smaller interval (e.g., a tenth is the same as an octave plus a third). For most purposes, compound intervals function as do their corresponding simple intervals (e.g., a tenth functions much as does a third, both being consonant (...)). The number of semitones between the two pitches is indicated by a qualifying adjective (perfect, major, minor, diminished, or augmented).” (Randel, 2003, entry: Interval [Fr. intervalle; Ger. Intervall; It. intervallo; Sp. intervalo], definition: 1). xv, 4, 23, 26, 50, 51, 62, 71, 76, 81, 95, 123, 126, 161, 162, 165, 173

**inversion** “[t]wo chords are related by inversion if both contain the same pitch classes and have the same root, but have different pitch classes in the bass or lowest-sounding position. In the case of the triad, a root-position chord has the root in the bass with the third and fifth above. If, however, the third is the lowest-sounding pitch, the chord is in first inversion, and if the fifth is the lowest, the
chord is in second inversion. Thus, the root-position triad c–e–g becomes e–g–c’ in first inversion and g–c’–e’ in second inversion. For seventh chords, if the seventh is in the bass, the chord is in third inversion (...). The nomenclature for chord inversions thus depends solely on the lowest-sounding pitch and is not affected by the particular disposition of the remaining pitches. Because the intervals formed above its lowest-sounding pitch are a sixth and a fourth, the second-inversion triad is termed a six-four chord; similarly, a first inversion triad is termed a six-three chord, or simply a six or sixth chord.” (Randel, 2003, entry: Inversion, definition: II. Chords). 60, 61, 76, 77, 146, 151, 173

**key** “[i]n tonal music, the pitch relationships that establish a single pitch class as a tonal center or tonic (or key note), with respect to which the remaining pitches have subordinate functions. There are two types or modes of keys, major and minor, and any of the twelve pitch classes can serve as a tonic. There are thus in principle 24 different keys. Because a pitch class may have more than one name, however (e.g., C♯ and D♭, which are said to be enharmonic equivalents), the nomenclature of keys includes more than 24. The key of a composition or passage is described in terms of its tonic and its mode (e.g., C major, D minor), and a work or passage is said to be ‘in’ a certain key. The key of a work is defined in terms of the particular major or minor scale from which its principal pitches are drawn. This is indicated in the first instance by a key signature—an arrangement of sharps or flats (or the absence of both) at the beginning of each staff that specifies the principal pitches. (Other pitches may be used as well, producing chromaticism.) The notion of scale embodies not only the selection of seven pitch classes from the available twelve, however, but also the organization of the seven in a hierarchy around the one that serves as tonic. When the pitches of a scale are arranged as a scale, the tonic is placed first. Furthermore, each key signature represents one major and one minor key that share the same basic pitch collection, but have different tonics. For example, the pitches of the white keys of the piano (represented by a key signature of no sharps and no flats) can be arranged in a scale with C first and thus as tonic
(yielding C major) or with A first as tonic (yielding A minor). These two ‘modes’ of presenting the same pitches differ in the arrangement of tones and semitones on either side of the tonic (...). Thus, in order to be in a given key, a composition must not only give prominence to the seven pitch classes of the appropriate scale, but it must also treat the tonic as the single pitch class of greatest stability and toward which all tonal movement ultimately tends. A piece in a given key will virtually always conclude with the tonic and will most often include a number of prominent cadences on the tonic.” (Randel, 2003, entry: Key, definition: 1) [embedded quote marks are originally written in Randel (2003) as “”, but in this definition are written as ‘ ’ for the sake of clarity]. xvi, 4, 16, 28, 36, 68–70, 72–75, 80, 90–92, 94–99, 101, 105, 106, 108, 109, 137, 145, 146, 149, 153, 154, 157, 171, 184, 186, 211, 215, 221, 222, 227, 230

**key signature** “[i]n tonal music, an arrangement of sharps or flats (or the absence of both) at the beginning of each staff that defines the principal pitches employed in the composition in question. Each sharp or flat indicates, respectively, a raising or lowering by a semitone of all pitches (in whatever octave) with the letter name of the line or space on which it is placed. This may be countermanded in individual cases by means of a natural sign or other accidental.” (Randel, 2003, entry: Key signature, definition: 1). 135, 137

**leading tone** “[t]he seventh degree of the major and harmonic or ascending melodic minor scales, which lies a semitone below the tonic and in tonal music often leads or resolves to the tonic.” (Randel, 2003, entry: Leading tone, note [Fr. (note) sensible; Ger. Leitton; It. (nota) sensibile; Sp. (nota) sensible], definition: 1). 12, 13

**measure** [also known as bar] “[a] unit of musical time consisting of a fixed number of note-values of a given type, as determined by the prevailing meter, and delimited in musical notation by two bar [(measure)] lines (...). The absolute duration of a measure is a function of tempo, i.e., the rate at which any note-value is performed. Informally, a measure may be said to consist of a given number of beats, with

**metre** [also known as meter] “[t]he pattern in which a steady succession of rhythmic pulses is organized; also termed time. Most works of Western tonal music are characterized by the regular recurrence of such patterns. One complete pattern or its equivalent in length is termed a measure or bar and in musical notation is enclosed between two bar [(measure)] lines. The meter of a work or of a passage within a work is indicated by a fraction (...). The denominator of the fraction indicates the basic note-value of the pattern, and the numerator indicates the number of such notevalues [sic] making up the pattern. Thus, a measure of the meter 3/4 consists of three quarter notes or their equivalent. The sign is the equivalent of 4/4; is the equivalent of 2/2. Informally, the numerator is sometimes taken as specifying the number of beats per measure [in this dissertation, we refer to this by the term division], and the denominator as specifying the note-value to receive one beat [in this dissertation, we refer to this by the term subdivision] (...). Meters in Western music are of two principal kinds: duple or triple, depending on whether the basic unit of pulse recurs in groups of two or three. The recurrence of groups of four pulses, as in 4/4, may be termed quadruple meter but is also a special case of duple meter. A meter in which this basic pulse is subdivided into groups of three, however, is said to be a compound meter. Thus, 6/8 is a compound duple meter because it consists of two groups of three eighth notes (three groups of two eighth notes would be written as 3/4 and would be a simple triple meter); 9/8 is a compound triple meter because it consists of three groups of three eighth notes.” (Randel, 2003, entry: Meter [Fr. mesure; Ger. Takt, Taktart; It. tempo, misura; Sp. tiempo, compás], definition: 1). 26, 283, 292

**note** “[a] symbol used in musical notation to represent the duration of a sound and, when placed upon a staff, to indicate its pitch; more generally (especially in British

**note value** “[t]ypes of notes are classed and named according to the relationship of their durations to one another and are sometimes termed notevalues [sic]. The symbol for indicating silence of a certain duration is termed a rest (...). [N]otes and rests in current use from largest to smallest, together with their names ([e]ach note or rest is twice as long as the next smaller one): [whole or semibreve, half or minim, quarter or crotchet, eighth or quaver, sixteenth or semiquaver, thirty-second or demisemiquaver, and sixty-fourth or hemidemisemiquaver].” (Randel, 2003, entry: Note, definition: 1). xviii, 157, 162–164, 167, 193, 284, 290, 291, 294

**passing note** “[a note] which connects two consonant pitches by stepwise motion and normally occurs in a metrically weak position.” (Randel, 2003, entry: Counterpoint [Lat. contrapunctus, fr. contra punctum, against note; contrepoint; Ger. Kontrapunkt; It. contrappunto; Sp. contrapunto], definition: II. Dissonance treatment in tonal counterpoint, 1. Passing tone). 159

**performance markings** “[w]ords, abbreviations, and symbols employed along with the notation of pitch and duration to indicate aspects of performance. These may be tempo indications, dynamic marks, technical instructions, marks for phrasing and articulation, and designations for the character of the piece or section.” (Randel, 2003, entry: Performance marks, definition: 1). 137

**pitch** “[a]ny point on the continuum of musical pitch, [which, in turn, is the] perceived quality of a sound that is chiefly a function of its fundamental frequency— the number of oscillations per second (...) of the sounding object or of the particles of air excited by it.” (Randel, 2003, entry: Pitch [Fr. hauteur; Ger. Tonhöhe; It. intonazione; Sp. entonación], definition: 2, 1). 26, 36, 39, 44–46, 50, 58, 62, 67, 68, 74, 124, 171
Glossary of musical terms

**pitch class** “[a] pitch without reference to the octave or register in which it occurs (...). Western tonal music uses twelve pitch classes, each of which is represented in each octave of the entire range of pitches.” (Randel, 2003, entry: Pitch class, definition: 1). 35, 45, 61, 67–70, 76, 80, 81, 86, 98, 99, 158

**pitch height** [also known as octave] according to Randel (2003, entry: Octave, definition: 1), an octave is “[a]n interval bounded by two pitches with the same pitch names and the higher of whose frequencies is twice the lower” and “[s]everal competing schemes for the designation of specific octaves are in use” (Randel, 2003, entry: Pitch names, definition: 1). In this dissertation, we assign the height 4 to the octave of the middle C (the C occurring roughly in the middle of the piano keyboard), the height 5 to the next higher octave, and so on. 27, 35, 86, 199

**progression** “[a] succession of two or more chords.” (Randel, 2003, entry: Progression, definition: 1) -not to be mistaken with GTTM’s progression. 4, 59, 140

**quarter-note** [also known as crotchet] see note value. 53, 54, 162, 163

**Riemannian functions** see functional harmony. 22

**Roman numeral notation** see chord function, harmonic analysis and scale degrees. 95

**rondo** “[a] multisectional form, movement, or composition based on the principle of multiple recurrence of a theme or section in the tonic key. In a standard rondo, the principal theme or section (usually symbolized A), known also as the refrain or rondo, alternates with subsidiary sections called couplets or episodes (symbolized B, C, etc.); it then returns at or near the end to complete the movement. All statements of the refrain are normally in the tonic key, whereas the couplets or episodes favor contrasting tonalities.” (Randel, 2003, entry: Rondo [It.], definition: 1). 252

**root** “[i]n tonal harmony, the fundamental or generating pitch of a triad or chord. A chord sounded with the root as the lowest pitch (even if the remaining pitches
are not sounded as superimposed thirds) is said to be in root position. Otherwise, the chord is in inversion.” (Randel, 2003, entry: Root, definition: 1) -not to be mistaken with the root of a generative grammar. xvi, 4, 12, 22, 60–62, 68, 70–73, 96–99, 108, 145, 157, 291

**root position** see root. 61, 172

**scale** “[a] collection of pitches arranged in order from lowest to highest or from highest to lowest.” (Randel, 2003, entry: Scale, definition: 1). 28, 35, 86, 114, 123, 159, 171

**scale degrees** “[t]he numbered positions of individual pitches within a major or minor scale. Because in Western tonality each pitch of a scale functions in a particular way with respect to the others, scale degrees are both numbered (traditionally with roman numerals) and named, as follows: I tonic, II supertonic, III mediant, IV subdominant, V dominant, VI submediant, VII leading tone or subtonic. The numbering and nomenclature are extensively used in harmonic analysis.” (Randel, 2003, entry: Scale degrees, definition: 1). 282, 290, 294

**semitone** [also known as half-step] “[t]he smallest interval in use in the Western musical tradition. There are twelve such intervals to the octave (...). The semitone is represented on the piano keyboard by the distance between any two immediately adjacent keys, whether white or black.” (Randel, 2003, entry: Semitone, definition: 1). 81, 165

**seventh chord** “[a] chord formed by the addition of pitches a third, a fifth, and a seventh above the lowest pitch or root.” (Randel, 2003, entry: Seventh chord, definition: 1). 68, 76, 186

**sixteenth-note** [also known as semiquaver] see note value. 53, 162–164, 193

**string quartet** “[a] composition for an ensemble consisting of four solo string instruments, normally two violins, viola, and cello.” (Randel, 2003, entry: String quartet, definition: 1). 1
subdivision  see metre. 160, 161, 288

subdominant chord  “[a chord built upon] the fourth scale degree of a major or minor scale, so called because it lies the same distance below the tonic as the dominant lies above the tonic, namely a perfect fifth. In harmonic analysis it is identified by the roman numeral IV.” (Randel, 2003, entry: Subdominant, definition: 1). 140

subdominant function  see chord function, functional harmony and tonal function. 22, 152, 153

suite  “[a] series of disparate instrumental movements with some element of unity, most often to be performed as a single work. The number of movements in a suite may be just large enough to constitute a series (three) or may be so great as to suggest that the work was intended to be treated as an anthology from which to make selections (…). Individual movements are almost always short and contrasting. A suite’s unity may result from nothing more than a common key or from its origins in a larger work, such as an opera or ballet, from which it is excerpted; unity may occasionally involve thematic connections and some sense of overall form. In some suites, the relationship among movements is defined by an extramusical program. The Baroque solo suite came close to having a specific pattern of dance movements at its core (allemande–courante–sarabande –gigue), but even then looseness of definition and variability of design were implicit in the term.” (Randel, 2003, entry: Suite [Fr., succession, following], definition: 1). 1

suspension  “a dissonant tone occurring in a strong metrical position, having been sustained (or ‘suspended’ or ‘prepared’) from an initial attack as a consonance and converted to a dissonance as a result of motion in another voice. It is most often resolved downward by step.” (Randel, 2003, entry: Counterpoint [Lat. contrapunctus, fr. contra punctum, against note; contrepoint; Ger. Kontrapunkt; It. contrapunto; Sp. contrapunto], definition: II. Dissonance treatment in tonal counterpoint, 3. Suspension) [embedded quote marks are originally written in Randel (2003) as " ", but in this definition are written as ‘ ’ for the sake of clarity]. 160
time-signature  “[t]he sign placed at the beginning of a composition to indicate its meter. This most often takes the form of a fraction, but a few other signs with origins in the system of mensural notation and proportions are also employed.” (Randel, 2003, entry: *Time signature*, definition: 1). xx, 53, 135, 157, 160, 161, 193, 209, 211, 212, 214, 215, 227


tonal function  *see* chord function and harmonic analysis. 11, 26, 283, 292, 294

tonal music  *see* tonality. 7, 8, 19, 24, 48, 64, 65, 75, 128, 131, 134, 171, 203, 246, 248, 250

tonal region  as a general rule (which admits exceptions), in the tonal system each work is in a single key, usually from the beginning of the work to its end. However, to get tonal variety and to build major forms, the composer can reinforce some degrees (diatonic or not) of the key to the point where, locally, these degrees function as provisional tonic. We call these provisional tonal centres regions (Roca and Molina, 2006, p.67). 45, 69, 70, 91, 247

tonality  “[i]n Western music, the organized relationships of tones with reference to a definite center, the tonic, and generally to a community of pitch classes, called a scale, of which the tonic is the principal tone; sometimes also synonymous with key. The system of tonality (sometimes termed the tonal system) in use in Western music since about the end of the 17th century embraces twelve major and twelve minor keys, the scales that these keys define, and the subsystem of triads and harmonic functions delimited in turn by those scales (…), together with the possibility of interchange of keys (modulation). A piece embodying this system is said to be tonal.” (Randel, 2003, entry: *Tonality*, definition: 1). 4, 28, 35, 36, 293
Glossary of musical terms

**tonic**  see **scale degrees**. 15, 58, 60, 66, 73, 74, 92, 157, 172, 192, 193, 197, 198

**tonic chord** a chord built upon the first scale degree. xvi, 11–15, 59, 60, 68, 72, 74, 80, 94, 97–99, 101, 139, 140, 144, 146, 147, 151, 217, 230, 240

**tonic function** see **chord function**, **functional harmony** and **tonal function**. 22, 147

**triad** “[a] chord consisting of three pitches, the adjacent pitches being separated by a third, and thus the whole capable of notation on three adjacent lines or three adjacent spaces of the staff; also termed the common chord.” (Randel, 2003, entry: *Triad [Fr. triade, accord parfait; Ger. Dreiklang; It. triade, accordo perfetto; Sp. tríada, acorde perfecto]*, definition: 1). 36, 67, 68, 108, 158, 173, 184, 186, 198

**tritone** “[a]n interval consisting of three whole tones.” (Randel, 2003, entry: *Tritone [Lat. tritonus]*, definition: 1). 12

**voice-leading** “[t]he conduct of the several voices or parts in a polyphonic or contra-puntal texture.” (Randel, 2003, entry: *Voice leading*, definition: 1). 26, 34, 81, 173, 199, 203, 248

**whole-note** [also known as semibreve] see **note value**. 53, 162
References


References


296


References


References


References


References


References


