Contour Levels: An Abstraction of Pitch Space based on African Tone Systems

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Aaron Carter-Ényi, MM, MM, BM

Graduate Program in Music

The Ohio State University

2016

Dissertation committee:

David Clampitt, Advisor

David Huron

Udo Will
Copyright by
Aaron Carter-Ényi
2016
ABSTRACT

Based on data from two years of fieldwork in Nigeria, a new methodology for contour analysis is presented with two motivations: 1) extend contour theory into an applied computational approach appropriate for a wide range of symbolic and recorded music; 2) develop a new discretization of pitch, similar to solmization but without an association to a scale or tonal qualia, that can be used to measure pitch prominence (or markedness) in both music and speech.

As an alternative to the conventional contour matrix for a segment of cardinality $n$ which compares pitches at all degrees of adjacency up to $n-1$, a continuous matrix is introduced, with unspecified cardinality and a fixed number of degrees of adjacency. The continuous matrix is a series of contour slices. Each slice compares a pitch to the pitch before and after up to the degrees of adjacency. The elements in each contour slice (a column in the continuous matrix) can be summed creating a measure of relative pitch height, a contour level.

The analysis implementation is based on a relationship between contour recursion and segmentation of pitch series. Thematic unity, as provided by contour recursion, is presumed to be intentional on the part of the producer and salient to the receiver. Non-overlapping iterations of a highly recursive contour are both semiotically and structurally important in a wide variety of monophonic signals. The analysis is made more robust by
searching for transformations and using reductive processes that make it possible to compare segments of different cardinalities.

Contour level analysis is applied to the phenomenon of “tone-and-tune”, wherein a single pitch series carries both linguistic and musical or paralinguistic communication. First the concept of a toneme (a pitch contrast in speech) is explored. Phoneticians and phonologists have described the toneme with paradigmatic (context-independent) and syntagmatic (context-dependent) features, but neither seems to satisfactorily formalize phonological equivalence of tone. Shortly before he died, prominent linguist Nick Clements asked “Do we need tone features?”, concluding that if we do, the ones we have are not working. A cue is taken from the folk heuristic and widely used pedagogical device for the Yorùbá language: Low-Mid-High tones are called Do-Re-Mi. It quickly becomes clear that the comparison with solmization has nothing to do with a tonal system and everything to do with relative pitch. Contour levels are proposed as a formal heuristic for the toneme that captures the relevant pitch relativity of the do-re-mi folk heuristic, while freeing it from the misleading Western tonality association. The rich oral poetry tradition of Southwestern Nigeria is explored using this approach.
DEDICATION

This document is dedicated to my family who inspired my scholarly pursuits and supported me through sometimes harrowing but ultimately rewarding research trips.
ACKNOWLEDGMENTS

Many people deserve thanks for their role in an undertaking spanning several years and two continents, notably my advisor David Clampitt and other mentors at Ohio State University including Adélékè Adéèkó, Johanna Devaney, David Huron, David Odden, Shari Speer and Udo Will, as well as peers, namely Nathaniel Condit-Schultz. This research is closely tied to previous graduate studies with John Silantien, John Nix and Nico Schüler. Linguists at-large have offered invaluable insight over the years, particularly Roger Blench, D. Robert Ladd and Will Leben. Michael MacAskill figured out the complex randomization of stimuli for the tone perception study in PsychoPy. Most of all I must acknowledge my wife, Quintina Carter-Ényi, and our colleagues in Nigeria: David Àiná, Austin Emielu, Anthony Mereni, Anthony Okoro and Sam Chukwu, who recently passed away.
VITA

2002 ................................................................. University of Texas High School
2006 ................................................................. B.M. Theory, Southwestern University
2009 ................................................................. M.M. Theory, Texas State University
2011 ................................................................. M.M. Conducting, Univ. Texas San Antonio
2012 to present ................................................ Graduate Teaching Associate, School of Music, The Ohio State University

Publications

Carter-Cohn, Aaron. “Interview with Ken Metz and an Analysis of His Work Songs from Mother Goose,” Texas State University at San Marcos, 2009.


Fields of Study

Major Field: Music
Specialization: Music Theory
TABLE OF CONTENTS

Abstract ............................................................................................................................... ii

Dedication .......................................................................................................................... iv

Acknowledgments ............................................................................................................... v

Vita ..................................................................................................................................... vi

Table of Contents .............................................................................................................. vii

List of Tables ................................................................................................................... xiii

List of Figures .................................................................................................................. xix

1. Introduction ..................................................................................................................... 1

  1.1 Key Concepts in Niger-Congo Tone Languages and Music ................................. 8

  1.2 Scope and Aims ...................................................................................................... 24

  1.3 Interdisciplinarity, Terminology and Notation ....................................................... 27

2. The Do-Re-Mi Folk Heuristic ....................................................................................... 34

  Introduction ................................................................................................................... 34

  2.1 Sarah Ann Glover and the Tonic Sol-Fa Method ................................................... 35

  2.2 Samuel Àjàyí Crowther and African Tone Systems .............................................. 43

  2.3 Origins of the Musical Analogy .............................................................................. 51

Discussion ..................................................................................................................... 54

Conclusion .................................................................................................................... 60
3. Theories of African Tone Systems ................................................................. 64
   Introduction .................................................................................................. 64
3.1 Paradigmatic Tone Levels ....................................................................... 65
3.2 Tone Level Normalization ........................................................................ 68
3.3 On the Musical Analogy (again) .............................................................. 70
3.4 The Paradigmatic versus Syntagmatic Debate ........................................ 71
3.5 Problems with the Syntagmatic Model .................................................. 74
3.6 Problems with the Paradigmatic Model .................................................. 79
3.7 Prototypical Paradigmatic and Non-Prototypical Languages .......... 81
3.8 Open Questions ....................................................................................... 83
4. Paralinguistic Intonation in Yorùbá .......................................................... 86
   4.1 Methodological Disclaimer ................................................................. 89
   4.2 Sermons from All Saints Anglican Church - Yaba ......................... 91
   4.3 Segmentation and Feature Extraction .............................................. 94
   4.4 Visualizing Intonation ...................................................................... 101
   4.5 Descriptive Statistics for Intonational Variability ......................... 111
   4.6 Intonational Model ........................................................................... 126
   4.7 Conclusion ......................................................................................... 128
5. Perception of Tone Intervals

5.1 Minimal Pairs and Pilot Study

5.2 Stimuli Preparation and Verification

5.3 Experimental Design and Results

5.4 Evidence for Syntagmatic Direction

6. Initialization and Temporal Scales

6.1 Tone Level Initialization

6.2 Lexical Priority

6.3 Lexical-Order Comparisons

6.4 Segmentation and Orders of Windowing

6.5 Scale-Dimension Temporalities

6.6 Phonological Equivalence in Music

7. Contour Theory and Pitch Perception

7.1 Origins of Contour Theory

7.2 Degrees of Adjacency and Perception
10. Realization in a Dynamic Register ................................................................. 285
   Introduction ........................................................................................................... 285
10.1 Modeling Intonation ......................................................................................... 288
10.2 Intonation Prototypes and Register Transformations ....................................... 292
10.3 Applying Register Transformations ............................................................... 306
10.4 Constraints: Ese Drum Example ...................................................................... 313
10.5 Musical Gestures in Jazz Improvisation ......................................................... 319
11. Interlevel Cycles and Robust Signals ............................................................... 328
   Introduction ........................................................................................................... 328
11.1 Interlevel Cycles ............................................................................................. 332
11.2 Interlevel cycle for a Two-level system: LH .................................................. 337
11.3 Interlevel cycle for three-level systems: LMH ............................................. 345
11.4 Interlevel cycles for larger level systems ...................................................... 352
11.5 Robust Signals ............................................................................................... 355
11.6 Contour Transformation ................................................................................. 363
12. Vocal Arts in Sub-Saharan Africa .................................................................... 370
   12.1 Vocal Arts and African Cultures ................................................................. 370
   12.2 Tone and Tune Elaborated .......................................................................... 377
13. Oríkì Ògún ......................................................................................................... 390
   Introduction ........................................................................................................... 390
13.1 Oríkí in the past ............................................................................................. 391
13.2 Oríkì Ògún: past in the present ....................................................................... 393
LIST OF TABLES

Table 1.1: tonal variants of the homophone /sI/ ............................................................... 10
Table 1.2: M-initial homophones of /i.gba/ ...................................................................... 12
Table 1.3: Tone combinations in a three-level system ..................................................... 13
Table 1.4: Two versions of a carrier sentence .................................................................. 15
Table 1.5: Measurements for each syllable segmentation of the three homophonous 7-
syllable phrases ................................................................................................................. 16
Table 1.6: Ìgbò tongue twister .......................................................................................... 21
Table 1.7: Tonal counterpoint in “Oríki Ejiré” (praise of twins) by Mayowa Adeyemo
(2013)................................................................................................................................ 21
Table 1.8: Intended and unintended meanings of /oha kele ike jesu/ ............................... 22
Table 1.9: Contour Adjacency Series for Coronation, recomposition and folk melody... 23
Table 1.10: Formal notation to be used ............................................................................ 32
Table 3.1: Example sentence in Jukun from Welmers (1973:81)................................. 66
Table 3.2 ........................................................................................................................... 76
Table 3.3 ........................................................................................................................... 76
Table 3.4 ........................................................................................................................... 78
Table 3.5: monosyllable homophones from Crowther (1852:4)................................. 84
Table 4.1: Measurements for each syllable segment of three repetitions of the same 11-
syllable phrase ................................................................................................................... 98
Table 4.2 ............................................................................................................................ 104
Table 4.3 ............................................................................................................................ 107
Table 4.4: Owadayo Yorùbá Sermon 2 ........................................................................... 111
Table 4.5 ............................................................................................................................ 113
Table 4.6 ............................................................................................................................ 113
Table 4.7 ............................................................................................................................ 114
Table 4.8 ............................................................................................................................ 115
Table 4.9 ............................................................................................................................ 116
Table 4.10 ......................................................................................................................... 117
Table 4.11: Yorùbá Sermons .......................................................................................... 121
Table 4.12: Nigerian English sermons ............................................................................ 122
Table 4.13 ......................................................................................................................... 123
Table 4.14: Summaries (British English not included) ................................................... 124
Table 5.1: homophone pairs used in the study ................................................................. 136
Table 5.2 ............................................................................................................................ 142
Table 5.3: Welch’s t-tests for stimuli in the set: /a.kwa/ HH-LH female-voice ............... 149
Table 5.4: Welch’s t-tests for stimuli in the set: /a.kwa/ HH-HL female-voice ............... 150
Table 8.1: Analysis of subject sug-segment by Dowling and Fujitani (1971:XX) ......... 213
Table 8.2: Interval series in generic intervals ................................................................. 213
Table 8.3: A ranking of contour slices in Schoenberg by multiplicity ......................... 224
Table 8.4: A ranking of contour slices by contour level (windowed relative pitch height) ................................................................. 225

Table 9.1: The cardinality saturation point for the Schoenberg miniature is 9 ................. 251

Table 9.2: Input Parameters for the Segmentation Algorithm ........................................ 253

Table 9.3: Segment evaluation criteria ......................................................................... 256

Table 9.4: Number of comparisons in mod C+SIM (degree limited) and C+SIM (n-1) for various cardinalities and degrees of adjacency around the focus. ............................ 258

Table 9.5: Input parameters and evaluation of resulting analyses, ranked lowest to highest in terms of evaluation product (last column). ...................................................... 260

Table 9.6: The location and similarity for all iterations of the models in SEG ............... 278

Table 9.7: Calculations of coverage (cov) from #S and #IS ........................................... 278

Table 9.8: The location and similarity of all iterations of CS2. Sound change terms are applied from linguistics: syncope indicates the omission of a medial sound, epenthesis is the addition ..................................................................................................................... 279

Table 11.1: All-ascent homophone groups in Yorùbá .................................................... 330

Table 11.2: Cayley table for a two-item binary state GIS .............................................. 332

Table 11.3: Interlevels in a two-level system ................................................................. 335

Table 11.4: Syntagmatic states and state changes for 4-cycle (bold indicates completion) ......................................................................................................................... 338

Table 11.5: Paradigmatic states and state changes for 4-cycle (bold indicates completion) ......................................................................................................................... 338

Table 11.6: /a.kʷa/ homophone group ............................................................................ 340
Table 11.7: /i.ɡba/ homophone group

Table 11.8: Inventory of n-level systems

Table 11.9: Example of parallelism from Olatunji (1984)

Table 11.10: Example of homophone change from Túbi (1955:49)

Table 11.11: Example of meaningless contrast from Olatunji (1984)

Table 11.12: Contour Adjacency Series for couplet in Table 11.11

Table 12.1

Table 12.2: Hymn to Speech comparison

Table 12.3: Chorus to Speech comparison

Table 13.1

Table 13.2: Text for Adéyêmo’s “Oríki Ògún” 7/27/13 audio-only recording

Table 13.3: Similarity and Correlation calculated in comparison to 7/27 1st Perf. (Video)

Table 13.4

Table 13.5: Correlation between tones (encoded LMH=012) to MIDI pitch values, and correlation between contour level reductions of each

Table 13.6: CONTCOM C+SIM

Table 13.7: C+SIM by phrase between 7/27 performances

Table 13.8

Table 13.9

Table 13.10: Comparison of CL2 (+1/–1) values for corresponding tones and realized pitch (1 = match, 0 = different)
Table 13.11: Phrase 18........................................................................................................ 410
Table 14.1 ................................................................................................................................... 414
Table 14.2: Track Information for First Generation Artist.................................................. 421
Table 14.3: Barrister ............................................................................................................. 423
Table 14.4: Comparison of CL2 (+1/–1) values for corresponding tones and realized pitch  
(1 = match, 0 = different)........................................................................................................ 423
Table 14.5: K1 ....................................................................................................................... 425
Table 14.6: Comparison of CL2 (+1/–1) values for corresponding tones and realized pitch  
(1 = match, 0 = different)........................................................................................................ 426
Table 14.7: Òṣùpá........................................................................................................... 429
Table 14.8: Comparison of CL2 (+1/–1) values for corresponding tones and realized pitch  
(1 = match, 0 = different)........................................................................................................ 429
Table 15.1: /akwa/ homophone group ............................................................................ 440
Table 15.2: C+SIM between Call (Baritone) and Response by Voice Part.................... 444
Table 15.3: CAS Similarity (phrase-bounded and only where there are complete verticalities, n=39)........................................................................................................... 450
Table 15.4: Statistical analysis of harmonic intervals ....................................................... 451
Table 15.5: CAS Similarity (phrase-bounded and only where there are complete verticalities, n=39)........................................................................................................... 455
Table 15.6: Phrase-initial and phrase-final tone and pitches (in Baritone Solo)............ 455
Table 16.1: Allophones in Ìgbò dialects (Clark 1990:17)............................................... 462
Table 16.2: Allophones in Yorùbá dialects (Crowther 1852:1)....................................... 462
Table 16.3: Disyllables that form tone-varied minimal pairs .......................... 473
Table 16.4: Initial Presentation of Disyllables in Ọjúkwụ’s “Tétá” ......................... 478
Table 17.1: Measurements for each syllable segmentation of the two homophonous 7-
syllable phrases ........................................................................................................ 485
Table 17.2: Symmetry of diatonic collections ............................................................... 492
Table 17.3: Values for Figure 17.3 ........................................................................ 494
Table 17.4: Value categories for 3-, 5- and 7-level reductions ............................... 496
Table 17.5: Sound categories in two cultures ............................................................ 500
LIST OF FIGURES

Figure 1.1: Diagram of spatial gestalts from Lehar (2003) .................................................. 1
Figure 1.2: Initial presentation of the subject of BWV 847 .................................................. 2
Figure 1.3: Modulated version of the subject (relative major) ............................................ 2
Figure 1.4: Tonal answer from BWV 847 .......................................................................... 3
Figure 1.5: Made-up melody with the same simple contour pattern ................................. 3
Figure 1.6 ............................................................................................................................ 4
Figure 1.7: Map of Africa from the World Atlas of Language Structures Online (Maddieson 2013) with annotations, blue arrows indicate Niger-Congo Expansion, red arrow indicates locus of Swahili (a non-tonal Bantu language) ........................................ 5
Figure 1.8: Map of language families in Africa.................................................................. 6
Figure 1.9: Alternate meanings for the homophone /ike/ in Ìgbò ........................................ 9
Figure 1.10: Three distinct pitch trajectories .................................................................... 10
Figure 1.11: Tones (noteheads) for ìké (HH, left) and ìkè (HL, right) within high (blue) and low (yellow) frequency bands .............................................................................. 11
Figure 1.12: Frequency banding of tones .......................................................................... 13
Figure 1.13 ........................................................................................................................ 14
Figure 1.14 ........................................................................................................................ 15
Figure 1.15: The same phrase with one tone altered “Mo mú [igbà/igbá] wá fún è”. Pitch: 60=C4................................................................................................................................. 16

Figure 1.16: talking drum set (Wikimedia commons)...................................................... 18

Figure 1.17: ùdù set in the Catholic chapel of Imo State University, Owerre (taken June 2014)................................................................................................................................ 20

Figure 1.18: Coronation hymn-tune.................................................................................. 22

Figure 1.19: Re-composition to fit tonemes ..................................................................... 23

Figure 1.20: Folk melody for “Ọhà kèlé” text.................................................................... 23

Figure 1.21: Language Map of Benin, Nigeria and Cameroon......................................... 25

Figure 2.1: “Ut queant laxis” on a four-line staff ............................................................ 36

Figure 2.2: Sarah Glover with the Norwich Sol-Fa Ladder chart..................................... 37

Figure 2.3: example of Tonic Sol-Fa notation.................................................................. 38

Figure 2.4: Portion of map showing West Africa from Johnston and Keane (1878)...... 44

Figure 2.5: Bishop Crowther (Date Unknown)................................................................. 45

Figure 2.6: Language use among respondents by state of origin from a 2014 survey of 1409 students and staff at public universities in southern Nigeria ................................. 48

Figure 2.7: (left) a drummer’s hand squeezing the tension cords of the dundún iyàálú (mother drum); (right) vocal apparatus (housed in the larynx) viewed from above........ 55

Figure 2.8: a Yorùbá Christmas song by David Äiná with staff and Sol-Fa notation...... 57

Figure 3.1 .......................................................................................................................... 66

Figure 3.2: The contour of the tone level sequence in Table 3.1 ...................................... 67

Figure 3.3: Jukun sentence in Table 3.1 with tone levels mapped to a C-major triad ...... 67
Figure 3.4: Illustrations of range-based proportional scaling from Patel (2008:45)........ 69
Figure 3.5 .......................................................................................................................... 73
Figure 3.6 .......................................................................................................................... 75
Figure 4.1: A complete Yorùbá sermon by Owadayo from All Saints' Anglican Church-
Yaba, analyzed by Melodyne.............................................................. 93
Figure 4.2 Three consecutive repetitions of the same phrase in a Yorùbá sermon:
“L’Órókọ Jésù ē ó rérin ayọ” (In the name of Jesus you will laugh with joy). ............ 98
Figure 4.3 ........................................................................................................................ 100
Figure 4.4: Owadayo’s Sermons..................................................................................... 102
Figure 4.5: Adesanya’s Sermons ................................................................................... 103
Figure 4.6: Adediji’s Sermons ........................................................................................ 104
Figure 4.7 ........................................................................................................................ 105
Figure 4.8 ........................................................................................................................ 106
Figure 4.9 ........................................................................................................................ 108
Figure 4.10: Owadayo Yorùbá Sermon 2 ................................................................. 110
Figure 4.11 ...................................................................................................................... 125
Figure 4.12 ...................................................................................................................... 127
Figure 4.13 ...................................................................................................................... 131
Figure 5.1: Tone-varied disyllable minimal pairs in Ìgbò (6 types, left) and Yorùbá (18
types, right). ................................................................................................. 135
Figure 5.2: images for /akwa/ (HH-HL) minimal pair.............................................. 137
Figure 5.3: Fundamental frequency plot (using YIN [12]) of the entire stimuli set for
/akwa/ modulated from HH to HL.............................................................. 138

Figure 5.4 ............................................................................................................. 143

Figure 5.5: images for /akwa/ (HH-HL) minimal pair........................................ 145

Figure 5.6: Results for /akwa/ HH to LH (left) and HH to HL (right) for female voice 147

Figure 5.7: images for /akwa/ (LH-HH) minimal pair........................................ 148

Figure 5.8: Fundamental frequency plot (using YIN) and MFCCs (using MIR toolbox) of
the entire stimuli set for /akwa/ modulated from HH to LH............................ 152

Figure 5.9: Fundamental frequency plot (using YIN) and MFCCs (using MIR toolbox) of
the entire stimuli set for /akwa/ modulated from HH to HL............................ 153

Figure 5.10: Sigmoid-curve regressions and inflection points for Ìgbò data........... 154

Figure 5.11: Sigmoid-curve regressions for Yorùbá.......................................... 155

Figure 5.12: Tone-varied minimal pairs in Ìgbò (6 types, left) and Yorùbá (18 types,
right)........................................................................................................... 157

Figure 5.13: /igba/ homophone collection on minimal pair circle with ascent triangle . 159

Figure 6.1 ........................................................................................................... 166

Figure 6.2 ........................................................................................................... 167

Figure 6.3 ........................................................................................................... 168

Figure 6.4 ........................................................................................................... 169

Figure 6.5 ........................................................................................................... 169

Figure 6.6 ........................................................................................................... 171

Figure 6.7 ........................................................................................................... 172
Figure 6.8 ........................................................................................................................ 175
Figure 6.9 ........................................................................................................................ 175
Figure 6.10: Phrase-Order Lexical Comparisons............................................................ 177
Figure 6.11: Section-Order Phrase Comparisons............................................................ 177
Figure 6.12 ...................................................................................................................... 179

Figure 7.1: an excerpted phrase from Schoenberg (Op. 19, No. 4) identified by Morris (1993), and a COM-matrix after Morris (1987) produced from the excerpt .......... 183
Figure 7.2: Kwakiutl Song (above) and Dahomean Song (below, Kolinski 1965:116). 191
Figure 7.3: (reprinted from Quinn 1999:445)................................................................. 199
Figure 7.4: COM-matrices for a segment of accumulating cardinality (third matrix adds a new minimum)................................................................................................. 204
Figure 7.5: Channel coding of a pitch series into working memory bins, based on Pollack (1952) and Miller (1956) ........................................................................................................ 206
Figure 7.6: Revised channel coding model with a fusing of middle bins...................... 207
Figure 7.7: C+ matrices for a segment of accumulating cardinality (after Quinn 1997) 209
Figure 8.1: the subject of the C-minor Fugue with its contour adjacency series (CAS) 212
Figure 8.2: a tonal answer of the C-minor Fugue subject with its CAS below .......... 212
Figure 8.3 ........................................................................................................................ 215
Figure 8.4 ........................................................................................................................ 216
Figure 8.5: Global matrix for Owadayo’s sermon with orders of window sizes (not precisely to scale)................................................................................................. 218
Figure 8.6: Second phrase of Schoenberg, Op. 19, No. 4, with 4–degree window around third pitch .......................................................... 219

Figure 8.7: COM-matrix, C+ Matrix and windowed C+ matrix for Phrase 2 of Schoenberg Op. 19, No. 4 .......................................................... 220

Figure 8.8: Windowed comparisons from a C+ matrix arranged as contour slices in a CONTCOM ................................................................. 221

Figure 8.9: The score with occurrences of slice [0;0;0;0] in boxes and the single occurrence of slice [0;0;1;0] circled. The least common slice corresponds to a prominent position in timing. ................................................................. 224

Figure 8.10: The continuous C+ matrix with [0;0;0;0] (the most frequent slice) in bold. ........................................................................................................................................ 224

Figure 8.11: The continuous C+ matrix with [0;0;1;0] (the least frequent slice) in bold. ........................................................................................................................................ 224

Figure 8.12: CONTCOM for Schoenberg, Op. 19, No. 4, with an added row for contour levels ........................................................................................................................................ 225

Figure 8.13: melody with sub-maxima boxed. .......................................................... 225

Figure 8.14: a Global Comparison Matrix (GLOBCOM) for Schoenberg Op. 19, No. 4. The bolded slices are the locations of the Global Maximum and Global Minimum ..... 229

Figure 8.15 ........................................................................................................ 230

Figure 8.16: All iterations of the C-minor Fugue subject ........................................ 231

Figure 8.17: (left) C+ matrix for the model versus (right) fuzzy C+ matrix for all iterations ........................................................................................................ 232
Figure 8.18: extraction of crisp values from a fuzzy C+ matrix................................. 233

Figure 8.19 ...................................................................................................................... 235

Figure 8.20: Same-pitch different-level anomalies....................................................... 236

Figure 8.21: the RCS reduction process for a C4 to C5 chromatic scale.................... 237

Figure 8.22: the RCS reduction process for a C-Major arpeggio.................................. 238

Figure 8.23: Contour slice reduction for a segment with many repeated pitches in the
middle of an ascent ......................................................................................................... 238

Figure 8.24...................................................................................................................... 239

Figure 8.25: All appearances of the first counter-subject............................................. 240

Figure 8.26: contour slices and contour levels for Counter-Subject 1............................ 240

Figure 8.27: Five-level reduction of the first counter-subject ...................................... 241

Figure 8.28: Iterations of episode material ................................................................. 241

Figure 8.29...................................................................................................................... 242

Figure 9.1: Schoenberg, Op. 19 No.4, segmented into phrases (from Morris 1993, 214)
......................................................................................................................................... 245

Figure 9.2: Morris’s phrase 2 decomposed into two similar segments, CSIM=0.80 .... 246

Figure 9.3: Frequency for each dyad type, the piece includes no adjacent repeated pitches
......................................................................................................................................... 248

Figure 9.4: score and CONTCOM with the most common melodic triad (CSEG <012>)
boxed............................................................................................................................... 248

Figure 9.5: CONTCOM with the most common melodic quatrad in bold. ................. 250

Figure 9.6: score and CONTCOM with the most common melodic quintad boxed. ..... 250
Figure 9.7: score and CONTCOM with the most common melodic hextad boxed........... 250
Figure 9.8: score and CONTCOM with the most common melodic septad boxed...... 250
Figure 9.9: A ground truth for the contour search algorithm, with two recursive segments
(one circled and one boxed)............................................................................................ 252
Figure 9.10: RCS reduction process for a C4 to C5 chromatic scale ......................... 254
Figure 9.11: RCS reduction process for a C-Major arpeggio...................................... 254
Figure 9.12: RCS reduction process for repeated pitches.......................................... 254
Figure 9.13: Schoenberg, Op. 19, No.4, includes two redundant contour slices and no
consecutive repeated pitches......................................................................................... 256
Figure 9.14: Workflow for the segmentation algorithm.............................................. 259
Figure 9.15: Analysis 1 is highest-ranked by EVAL Product, SEEGAP is turned on
(locations indicated on CONTCOM by arrows)....................................................... 263
Figure 9.16: Analysis 2: similar to Analysis 1 but with SEEPGAP turned off, allowing
segments to span rests (as from event 35 to 36)......................................................... 264
Figure 9.17: Analysis 3 is over-segmented because minimum cardinality is too low.... 265
Figure 9.18: Reduction of repeated pitches (DELREP) and incomplete slices.......... 267
Figure 9.19: Morris’s phrase 2, window allowed to change degree-orientation......... 268
Figure 9.20: Recursive Segments from Schoenberg piece with contour levels mapped to
staff.................................................................................................................................. 269
Figure 9.21: a mapping from the pitches of a parent segment iteration (starting at index 8)
to the CSEG class (center), and finally to the C+SEG_f (right)............................... 271
Figure 9.22: Voices in BWV 847 numbered in order of entrance............................... 273
Figure 9.23: concatenation of three voices ordered by entrance into P (n=750) .......... 274
Figure 9.24: CONTCOM for all three voices of Bach C-minor Fugue .................... 275
Figure 9.25: an appearance of Counter-Subject 2 (boxed) starting in the upper voice (V2) and proceeding to the middle voice (V1) (V1 is below V2 in the ordering) .......... 277
Figure 9.26: All appearances of CS2 .............................................................................. 279
Figure 9.27: C+matrix with fuzzy values for the four cardinality-twelve appearances of CS2, consistent subsegments are bolded ................................................................. 280
Figure 10.1: trend lines for tone levels (blue=high, green=mid, yellow=low)............. 288
Figure 10.2: intonation grid lines based on Eva Gärding’s work (Reprinted from Ladd 2008) ........................................................................................................................................... 289
Figure 10.3 ...................................................................................................................... 290
Figure 10.4: Piecewise constant register (Ladd 1990:40)............................................. 290
Figure 10.5: Phrase-order registers of Owadayo excerpt............................................. 291
Figure 10.6: (Left to Right) Two-level system; Three-level system; Generic 7–pitch space; 2–octaves of a pentatonic scale........................................................................ 292
Figure 10.7: Left-hand accompaniment at the opening of Mozart K. 545 ............... 294
Figure 10.8: Active region corresponding to Figure 10.7............................................. 294
Figure 10.9: Close (left) and open (right) spacing of fixed registers......................... 296
Figure 10.10: Fixed registers high (left) and low (right) within the range ................. 297
Figure 10.11: MODTRANS function mapping contour space to adjacent step-classes in scales ................................................................................................................. 298
Figure 10.12: Ascending (left) and descending (right) dynamic registers in continuous pitch space .................................................................................................................................................................................. 299

Figure 10.13: Conjunct mappings of 3-levels to pentatonic scale (box indicates mappings with overlap between higher octave and lower octave) ................................................................................................................. 300

Figure 10.14: Disjunct mappings [0,2,4] of 3-levels to pentatonic scale (box indicates mappings with overlap between higher octave and lower octave) ........................................... 300

Figure 10.15 ........................................................................................................................................................................ 301

Figure 10.16: Six prototypes for open/close function in continuous pitch space .......... 302

Figure 10.17: Compression-Ascent: Disjunct-to-conjunct mappings of 3-levels in lower octave and higher octave (no overlap between octave registers) .................................................. 303

Figure 10.18: Compression-Descent: Disjunct-to-conjunct mappings of 3-levels in higher octave and lower octave (no overlap between octave registers) ........................................... 303

Figure 10.19: Expansion using the open/close function in different scales ................. 304

Figure 10.20: warp function with augmenting lower interval (left) and augmenting upper interval (right) ......................................................................................................................................................... 306

Figure 10.21 ........................................................................................................................................................................ 306

Figure 10.22: CONTCOM₂(P) with contour levels below .............................................. 307

Figure 10.23 ........................................................................................................................................................................ 307

Figure 10.24: discretized dynamic register (R) for Alberti bass excerpt .................... 308

Figure 10.25: Mozart Alberti bass voice-leading modeled through J-function and warp function .................................................................................................................................................. 309

Figure 10.26: Contour figure for Alberti bass pattern .................................................. 310
Figure 10.27: Realization on a C-major triad ................................................................. 311
Figure 10.28: realization on an F-major triad in second inversion ...................... 311
Figure 10.29: Left-hand accompaniment at the opening of Mozart K. 545 .......... 312
Figure 10.30: Mappings of two tone-levels to a seven-pitch collection .......... 314
Figure 10.31 .................................................................................................................... 315
Figure 10.32 .................................................................................................................... 315
Figure 10.33 .................................................................................................................... 316
Figure 10.34: (a) mapping of 3-level contour space to two-octaves of a pentatonic scale (treble clef omitted) to a specified register, (b) mapping of three levels to 2–octave pentatonic scale without the constraint of an active region, (c) mapping of three levels to a smaller space .......................................................................................................................... 319
Figure 10.35: Instances of contour recursion in Gillespie’s solo .................. 320
Figure 10.36: Contour segments as they appear in the reduced continuous matrix ...... 321
Figure 10.37: Three-level contour reduction ......................................................... 321
Figure 10.38: Opening of the melody of “Blue Moon” by Rodgers and Hart .......... 322
Figure 10.39: Example of an elaborated gesture reduced to an underlying three-level contour .......................................................................................................................... 322
Figure 10.40 .................................................................................................................... 323
Figure 10.41: CLSEG segmented into three CSUBSEGs ........................................ 323
Figure 10.42 .................................................................................................................... 324
Figure 10.43: Reduced and normalized pitch segments in order of appearance with transformational diagram below each. Ellipses indicate other material between and R and RI are presented in prime form.

Figure 10.44: Realization in a dynamic register with expansion of the medial level...

Figure 10.45: Visualization of Figure 10.44...

Figure 11.1: Descent (blue) and ascent (yellow) triangles in the three-level 12-cycle...

Figure 11.2: Length 4 and length 6 uniform state-change cycles...

Figure 11.3: Syntagmatic 4-cycle...

Figure 11.4: Paradigmatic 4-cycle...

Figure 11.5: Aggregate cycle for two-level system...

Figure 11.6: Single sine-wave cycle representation of aggregate cycle...

Figure 11.7: Signal pathways...

Figure 11.8: Two-level 4-cycle with downstepped high satellite...

Figure 11.9: Two-level 4-cycle with satellites for raising and lowering High level...

Figure 11.10: 4–point star of zero-sum interlevels...

Figure 11.11: Dynamic interlevels (left) and static interlevels (right)...

Figure 11.12: Cross-cycle pairs...

Figure 11.13...

Figure 11.14: East and west pathways and MM-axis...

Figure 11.15: Three-level 12-cycle with same-paradigm diametric chords...

Figure 11.16: Geometric shapes formed by paradigmatic relationships...
Figure 11.17: 3D interlevel network, arrows indicate equivalent points in the wrap-around ................................................................. 355

Figure 11.18: Anomalies in contour levels (sums of contour slices): (left) consecutive repeated pitches, and (right) insufficient context for time-extrema ......................... 356

Figure 11.19: Preferred (green) pathways for a robust signal ........................................ 357

Figure 11.20: “–2” (left), “–1/+1” (middle) and “+2” (right) window reductions for HLMLHLM .................................................................................................................... 359

Figure 11.21: “–2” (left), “–1/+1” (middle) and “+2” (right) window reductions for HLMLMLH .................................................................................................................... 359

Figure 11.22: “–2” (left), “–1/+1” (middle) and “+2” (right) window reductions for MLHLHLM .................................................................................................................... 360

Figure 11.23: “–2” (left), “–1/+1” (middle) and “+2” (right) window reductions for MLHMHLM ................................................................................................................... 360

Figure 11.24: “–2” (left), “–1/+1” (middle) and “+2” (right) window reductions for LMHLHML .................................................................................................................... 361

Figure 11.25: “–2” (left), “–1/+1” (middle) and “+2” (right) window reductions for MHLMLHM ................................................................................................................... 361

Figure 11.26: “–2” (left), “–1/+1” (middle) and “+2” (right) window reductions for HMLHLMH .................................................................................................................... 362

Figure 11.27: Analysis of tonal counterpoint in Adeyemo’s 2013 performance ............ 367

Figure 11.28: Reduction of repeated tones ................................................................. 367

Figure 12.1 ...................................................................................................................... 371
Figure 12.2: Hymn 12: Spoken version of “Ase ifẹ orun” text (blue) and Narenza hymn tune (green) .................................................................................................................................................................................. 385

Figure 12.3: Spoken version of “Ase ifẹ orun” text (blue) and Swabia hymn tune (green) .................................................................................................................................................................................. 386

Figure 12.4: Hymn 10 “Krist’ ki’jọba Rẹ de” .................................................................................................................. 386

Figure 12.5: Chorus 1 .................................................................................................................................................. 388

Figure 13.2: Adeyẹmọ performing at the University of Lagos, April 13, 2013 .......... 395

Figure 13.3: David Àiná (back left), the Ogundepo family (front) and myself (back right) .................................................................................................................................................................................. 397

Figure 13.4: mean log pitch for 40-event windows and a hop size of 5 ................. 398

Figure 13.5 .................................................................................................................................................. 399

Figure 13.6: mean log pitch for 10-segment windows and a hop size of 2 ............. 402

Figure 13.7: Intonational Model for 3-level system .............................................. 403

Figure 13.8: Tonal counterpoint in Phrases 7 & 8 .................................................. 408

Figure 13.9: Phrase 18 ........................................................................................................ 411

Figure 14.1 .................................................................................................................................................. 416

Figure 14.2 .................................................................................................................................................. 421

Figure 14.3: Barrister ............................................................................................................................... 422

Figure 14.4: Phrase 5 ........................................................................................................................................ 422

Figure 14.5: Cover from K1’s Iṣẹ L’Ogun Iṣẹ .............................. 424

Figure 14.6: Track Information for Second Generation Artist ....................................... 424

Figure 14.7: K1 ........................................................................................................................................ 425
Figure 14.8: Tonal counterpoint in Phrases 1–2 ............................................................. 425
Figure 14.9: Cover for Òṣùpá’s Fuji Icon .................................................................... 427
Figure 14.10: Track Information for Third Generation Artist ................................... 427
Figure 14.11: Òṣùpá’s introduction to Fuji Icon “B” (“Tribute to my father”) ............ 428
Figure 14.12: Phrase 1 ................................................................................................. 428
Figure 14.13: Parallel phrases 3 & 11 .............................................................................. 430
Figure 14.14: Parallel phrases 6 & 14 .............................................................................. 430
Figure 15.1: MUSON Diploma Choir at the ISME Conference, Beijing, China, August 2010 ................................................................................................................................. 432
Figure 15.2: Sound relationships in Èkwúèmé’s Glees .................................................. 437
Figure 15.3: Vocal imitation: Mm. 1–2 of Ote Nkwu (1992) by Èkwúèmé ................. 439
Figure 15.4: “Kwango, kwango” (the sound of pounding yam) ..................................... 439
Figure 15.5: Verse 1: mm. 1–2 and Verse 3: mm. 17–18 of Obi Dimkpa (1980) by Èkwúèmé ......................................................................................................................... 441
Figure 15.6 ...................................................................................................................... 442
Figure 15.7: Comparison of Verse 2 and Verse 4, C+SIM = .800 ................................. 442
Figure 15.8: Praise singers outside the Èkwúèmé residence in Òkọ, Ánámbrá, Nigeria. Ògénè (metal gong) and ìdù (metal pot) are pictured ................................................. 443
Figure 15.9 ...................................................................................................................... 445
Figure 15.10 .................................................................................................................... 445
Figure 15.11 .................................................................................................................... 446
Figure 15.12 .................................................................................................................... 447
Figure 15.13 .................................................................................................................... 448
Figure 15.14 .................................................................................................................... 449
Figure 15.15 .................................................................................................................... 449
Figure 15.16: Parallel Harmony: mm. 5–6 of Obi Dimkpa (1980) TTBB Version by Èkwúèmé ......................................................................................................................... 452
Figure 15.17: Refrain of Obi Dimkpa (1980) SATB by Èkwúèmé ................................ 452
Figure 15.18: Glee-club harmony: Coda of Obi Dimkpa (1980) TTBB by Èkwúèmé, ii43 to ii°7 to I ........................................................................................................................ 454
Figure 15.19: UTSA Choir performing Obi Dimkpa, San Antonio, Texas, April 2011. 457
Figure 16.1: transcription and contour analysis of a Dohomean song (Kolinski 1965) . 466
Figure 16.2: Excerpt 1 of Dahomean song from Kolinski (1965) .................................. 468
Figure 16.3: Excerpt 2 of Dahomean song from Kolinski (1965) .................................. 468
Figure 16.4: Improvised harmony for “Oha Kele” ............................................................ 470
Figure 16.5: Distributions of harmonic motion ............................................................... 471
Figure 16.6: Frame from FILM performance (video link: <https://youtu.be/hspZiY3328s>) .................................................................................................................... 475
Figure 16.7: “Tètá” ostinato section, p. 1, (xylophone omitted) .................................... 476
Figure 16.8: “Tètá” p. 2 ................................................................................................ 477
Figure 16.9: Backup singers’ refrain .............................................................................. 481
Figure 16.10: A still from the music video for “Igwe” (2009) ........................................ 481
Figure 17.1: The same phrase with one tone altered “Mo mú [igbà/igbá] wá fún ì”..... 485
Figure 17.2: Curwen’s handsigns from the Standard Course (1880:iv) ....................... 490
Figure 17.3: “L’órúkọ Jésù ọ rẹrín a yọ” repeated three times ........................................ 494

Figure 17.4: Crisp, Fuzzy and Gestalt Equivalences .......................................................... 497

Figure 17.5: a tonal answer of the C-minor Fugue subject with its CAS below ............... 499

Figure 17.6: Cross-disciplinary and cross-cultural research............................................. 503
1. INTRODUCTION

Much Gestalt psychology has focused on the visual, or spatial. Take this object with what looks like a handle and an implement. In category A, the object appears in various rotations and depths. In category B, the parts of the object are rearranged. In C, the object is warped, and in D the image is enhanced or distorted.

![Diagram of spatial gestalts from Lehar (2003)](image)

Figure 1.1: Diagram of spatial gestalts from Lehar (2003)

If I were asked, *Which category has objects that do not share a common a Gestalt?* I would say category B. Recombining component parts produces something different from the original.
In music, which is experienced in time, we talk about temporal gestalts (e.g. Tenney and Polansky 1980). A classic example is the primary subject of Bach’s C-minor Fugue from the Well-Tempered Clavier (BWV 847).

![Initial presentation of the subject of BWV 847](image)

Fugues are highly self-similar. This subject appears eight times in the piece, in this original version and in various transformations. When the subject reappears in its original form we experience a temporal gestalt, based on pitch equivalence (abbreviated P). A transformation of the subject is exemplified by this modulated form of the subject in the relative major, E-flat.

![Modulated version of the subject (relative major)](image)

The generic intervals (e.g. seconds, thirds, fourths) remain intact, but the specific intervals measured in semitones have changed. Thus, this is a contour-equivalent transformation. Each produces an identical contour comparison matrix in which every note in the melody is compared to every other note in the melodic segment (see Marvin and Laprade 1987, or Chapters 7 and 8). I refer to this as complex-contour equivalence (C).
Unlike the modulated version above, the tonal answer is not a diatonic transposition of the subject nor does it produce the same contour matrix. It is adapted.

![Figure 1.4: Tonal answer from BWV 847](image)

A less rigid form of equivalence is simple contour (S), in which only notes next to each other are compared. In both the subject and the tonal answer, the first to second note is a descent, the second to third is an ascent, and so on. Both versions produce the same series of ups and downs. However, simple contour agreement does not guarantee melodies are perceptually invariant. Take this melody that I made up based on the same rhythm, implied harmony and simple contour pattern as the tonal answer.

![Figure 1.5: Made-up melody with the same simple contour pattern](image)

Although there is some semblance, for me, the made-up melody is not similar enough to belong to the same Gestalt.
I classify pitch (P), transpositional (T) and complex-contour (C) equivalence as crisp categories, and simple contour (S) as fuzzy. These forms of equivalence are either too rigid or too lax to describe the relationship between Bach’s subject and tonal answer. An alternative model comes from African tone systems.

Ethnolinguistic Cultures of Sub-Saharan Africa

Figure 1.7 is a map of Africa from the World Atlas of Language Structures Online. Each dot represents one language from a sample of 68 languages.¹

¹ The World Atlas sampled 527 languages globally.
There is a cluster of tone languages in Cameroon and Nigeria, an area with some of the greatest linguistic diversity in the world. Proto-languages from this region were carried west and southeast (indicated by the blue arrows) forming two branches of the Niger-Congo language family. Out of 68 languages included in this sample, 63, over 90%, are tonal. Swahili, a lingua franca in East Africa, falls into the “No Tones” category (its location is indicated with a red arrow in Figure 1.1).
In all, the Niger-Congo family includes around 1000 languages, with hundreds in Nigeria alone. Where there is a distinct ethno-linguistic culture there are often distinct musical practices.

George Herzog’s 1934 article “Speech Melody and Primitive Music” was the first to my knowledge to discuss the phenomenon of “tone-and-tune”: the reflection of spoken lexical tones in sung melody. In Herzog’s circumspect account of Jabo in eastern Liberia, he cautions that a strict following of speech by melody is not implied (1934:466). Others have taken stronger positions. Scholarship on the music of the Yorùbá and the Ìgbò of southern Nigeria supports a tone-tune relationship, but studies of geographically close but phonologically divergent languages such as Hausa (by Leben 1983) and Ewe (by Agawu
1988) have questioned the relationship. A recent study by Murray Schellenberg found tone-tune correspondence in music of the Shona, a southern Bantu group of Zimbabwe (2009). Field recordings from Kenya (2013) as well as discussions with researchers of the eastern side of the Bantu expansion suggest the phenomena is not prevalent there. This explains the lack of “tone-and-tune” studies from East Africa. This variation does not hinge on the presence of tone conventions in the language so much as whether tone forms essential contrasts for word comprehension. David Odden points out that tone is not always contrastive (email correspondence 2012) and Robert Ladd suggests that phonological equivalence of tones is the same in all languages (2008b). This dissertation focuses on cultures of Nigeria where constrictive pitch patterns are essential to language comprehension. Specifically, I will address Igbo and Yoruba of southern Nigeria, two Niger-Congo tone languages that continue to thrive amid many endangered languages. In these languages, around 50% of disyllable words can only be distinguished from other words by lexical tone contrast.

Out of 520 living languages in Nigeria, SIL's Ethnologue reports 70 are “in trouble” or “dying.” Seven documented languages have gone extinct in the last decade (Lewis et al. 2015). English is the national language of Nigeria, requisite in schools and workplaces. People continue to speak indigenous languages with friends and family as well as use them in worship and the arts. Indigenous-language vocal arts, including music, poetry and drama, remain vibrant and are key to future vitality of African languages.

West African cultures have had a tremendous impact on our global musical culture. Particularly those that inhabit what was known by Europeans as the “Slave
Coast” into the late nineteenth century. The coast from the Bight of Benin to the Bight of Biafra, present-day Benin, Nigeria and Cameroon, was the site of multiple trans-Atlantic slave ports including Ouidah, Badagry and Calabar. Today, practices of Yorùbá religion, including Santeria or Candomble, are as active in Brazil, Cuba and Haiti as they are in Nigeria. Niger-Congo cultures have had a major influence on music of the Americas and Caribbean, and from there, the world.

1.1 Key Concepts in Niger-Congo Tone Languages and Music

Chapters 2 through 5 explore the elusive concept of the toneme, a pitch contrast in speech. In English we are more familiar with segmental phonemes (vowels and consonants). By comparing the two words of a minimal pair (e.g. /pit/ vs. /bit/) we can determine that an unvoiced stop, /p/, and a voiced stop, /b/, are both segmental phonemes in English distinguished by voicing. In another language they might not be distinct, they might be allophonic labial stops which do not produce distinct words.

1.1.1 What is a Toneme?

In addition to segmental phonemes, duration, stress and tone may also be contrastive. Each of these has an analog to musical features. Contrast between features of segmental phonemes is like variation in timbre or articulation. Timing and stress often work together in speech (e.g. pre-sent versus pre-sent) much like rhythm and dynamics in music. Native English speakers are not accustomed to using tone contrastively at the lexical level, but as much as seventy percent of world languages have lexical tone (Yip
English speakers do articulate and attend to pitch contrasts on the larger intonational level (as in rising intonation for a question) and in musical melody.

An example of lexical tone contrast is the homophone /ike/ in Ìgbò. We know there is contrastive tone, a toneme, present in this word because it changes the mental representation of Ìgbò speakers from strength to buttocks (see Figure 1.9).

![Figure 1.9: Alternate meanings for the homophone /ike/ in Ìgbò](image)

Study data presented in Chapter 4 demonstrates that altering the relative pitch of the second syllable from roughly equal to the first (as in ike n. strength) to substantially lower than the first (as in ikè n. buttocks) shifts the dominant response to an audio stimulus from the left image to the right image.

Tonemes come in many different shapes and sizes. In Mandarin, each toneme has a distinct pitch trajectory within a single voiced segment, such as a vowel or nasal.

---

2For lack of a better term, homophone is used to refer to words that have the same segmental phonemes but vary in tone.
sonorant (/m/ or /n/). Figure 1.10 shows the four pitch trajectories in standard Chinese (Mandarin). There are as many as eight in other Chinese languages, such as Cantonese.

The homophone /sɪ/ forms three different meanings with the pitch trajectories in Figure 1.10 are applied. Languages such as Mandarin are sometimes referred to as contour tone languages.

<table>
<thead>
<tr>
<th>Mandarin</th>
<th>Tone</th>
<th>English Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>sǐ</td>
<td>high level</td>
<td>personal; private</td>
</tr>
<tr>
<td>sì</td>
<td>falling</td>
<td>number four</td>
</tr>
<tr>
<td>sǐ</td>
<td>falling-rising</td>
<td>death</td>
</tr>
</tbody>
</table>

Table 1.1: tonal variants of the homophone /sɪ/

In Ìgbò, the shape of the pitch trajectory within the segment is relatively flat and is not intentionally contrastive. In Niger-Congo tone languages each syllable is abstracted to a target pitch, not a pitch trajectory. This is reflected in both linguistic theory and in the common do-re-mi heuristic explored in Chapter 1. There may be portamento present, but it is generally considered not to be important to perception. The target pitch is a
realization of an underlying tone level. In the word *iké* (strength), both syllables are at a high tone level. In the word *ikè* (buttocks), the first syllable is at the high tone level and the second syllable at the low tone level.

So what is a toneme in a language like Ìgbò? Is it the relationship between the syllables? The pitch trajectory between the two syllables for the words *iké* (HH) and *ikè* (HL) is different. A syntagmatic approach ascribes a disyllable word a single toneme, flat tone or no change (0) for *iké* (strength) and falling tone or descent (–) for *ikè* (buttocks). Or, do High and Low tones have pitch within a respective frequency band of a speaker’s range? An autosegmental approach ascribes a toneme to each tone-bearing segment, usually a syllable (see Figure 1.11).

Figure 1.11: Tones (noteheads) for *iké* (HH, left) and *ikè* (HL, right) within high (blue) and low (yellow) frequency bands

In Figure 1.11, I use noteheads on a partial musical staff to illustrate the concept of frequency bands. In autosegmental theory, tones are very much like notes. One hypothesis is that tones recur at the same pitch or frequency band within speaker (see discussion of Welmers 1973 in Chapter 3). This is not the only explanation of

---

3 Although in early accounts, descriptions of Niger-Congo languages did suggest the pitch trajectory of each tone was relatively flat, that is not the implication here (see Chapter 3).
autosegmental tone but it is one that continues to influence some scholarship (including Deutsch et al 2003, 2006). To draw an analogy to pitches in music, A4 is not just 440 Hz it has a frequency band around it of 50 cents in either direction. 438 Hz and 443 Hz still belong to the pitch category A4, they are just slightly flat or sharp. If a vocalist or string player aims for 440 she may only reach the target frequency momentarily or only come close to it, but the performed note may still be accepted as A4. The concept of a frequency band for the High tone level is similar. It suggests that every time a speaker speaks a syllable on high tone it will be within a frequency band or near enough an absolute pitch referent that it unambiguously belongs to the category “high”.

What about Yorùbá, a three-level tone language: is the toneme the direction of change between syllables or does each tone fall within a frequency band?

<table>
<thead>
<tr>
<th>Word</th>
<th>Tone</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>igbá</td>
<td>mid-high (MH)</td>
<td>calabash</td>
</tr>
<tr>
<td>igba</td>
<td>mid-mid (MM)</td>
<td>200</td>
</tr>
<tr>
<td>igbà</td>
<td>mid-low (ML)</td>
<td>climbing rope</td>
</tr>
</tbody>
</table>

Table 1.2: M-initial homophones of /i.gba/

Yoruba is phonologized with three levels, low-mid-high, with a neutral level between high and low.
For a three-level system, the motivation for frequency banding is greater because there are multiple descending and ascending tone level combinations. Mid to High, Low to Mid and Low to High are all ascending.

1.1.2 TONE-VARIED HOMOPHONES

In a three-level tone language like Yorùbá, there are nine possible two-tone combinations \(3^2\).

```
<table>
<thead>
<tr>
<th>_H</th>
<th>_M</th>
<th>_L</th>
</tr>
</thead>
<tbody>
<tr>
<td>H_</td>
<td>HH</td>
<td>HM</td>
</tr>
<tr>
<td>M_</td>
<td>MH</td>
<td>MM</td>
</tr>
<tr>
<td>L_</td>
<td>LH</td>
<td>LM</td>
</tr>
</tbody>
</table>
```

Table 1.3: Tone combinations in a three-level system

This brings forth the possibility that a single disyllable homophone could have nine tonal variations and distinct meanings. However, there are no homophones with nine versions. Many homophones are vowel-initial and I have found no vowel-initial words with high-tone on the first syllable. This suggests a constraint against high-tone appearing on the
first syllable of vowel-initial words. This constraint reduces the number of possible variants to six for many disyllable homophones. The homophone /igba/, which forms a group of five words, has the highest number of variants to my knowledge.

![Tone Generator](image)

**Figure 1.13**

There are also words, such as the similar word *igbi* (n. wave, surf), that have no tone-varied homophones.

The first grammar and dictionary for Yorùbá was published by Samuel Ajayi Crowther in 1852. Crowther was the first to note the Yorùbá language is “very musical” and to develop an orthography for an African language that incorporated tone marks. Throughout the colonial era, a healthy dose of choral singing was part of the civilizing agenda. Many post-colonial cultures have lively tonic-solfa (do-re-mi) practices to this
day. Among the Yorùbá, solmization became a heuristic for speech tone levels, do-re-mi syllables for low-mid-high tone levels.

If we take the carrier sentence “I brought you a” and alternate the lexical tone for /igba/ between MH and ML, it forms two distinct sentences: “I brought you a calabash” and “I brought you a climbing rope.”

<table>
<thead>
<tr>
<th>Yorùbá</th>
<th>Tone Sequence</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mo mú igbá wá fun ᕙ</td>
<td>MHMHHHHM</td>
<td>I brought you a calabash</td>
</tr>
<tr>
<td>Mo mú igbà wá fún ᕙ</td>
<td>MHMLHHHM</td>
<td>I brought you a climbing rope</td>
</tr>
</tbody>
</table>

Table 1.4: Two versions of a carrier sentence

The do-re-mi heuristic suggests the pitch pattern could follow the notation in Figure 1.14.

1.1.3 AUDIO ANALYSIS

Analyses of recordings for the two versions of the carrier sentence appear in Figure 1.15 and Table 1.5. In this and other analyses of brief recordings, the signal is manually segmented into syllables based on a transcription by one or more fluent speakers using Melodyne’s Direct-Note-Access (DNA) GUI interface.
Figure 1.15: The same phrase with one tone altered “Mo mú [igbá/igbá] wá fún ẹ”. Pitch: 60=C4

<table>
<thead>
<tr>
<th>Syllable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text</td>
<td>Mo mú i- gbá wá fún ẹ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tone Level</td>
<td>Mid</td>
<td>High</td>
<td>Mid</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>Mid</td>
</tr>
<tr>
<td>Mean Pitch (ST)</td>
<td>47.3</td>
<td>48.5</td>
<td>46.9</td>
<td>49.2</td>
<td>48.4</td>
<td>48.0</td>
<td>46.3</td>
</tr>
<tr>
<td>Interval Direction</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>Pitch Difference (ST)</td>
<td>1.2</td>
<td>-1.6</td>
<td>2.2</td>
<td>-0.8</td>
<td>-0.4</td>
<td>-1.7</td>
<td></td>
</tr>
<tr>
<td>Segment Slope (ST/sec)</td>
<td>0.1</td>
<td>-1.5</td>
<td>1.0</td>
<td>-1.3</td>
<td>-2.1</td>
<td>-5.8</td>
<td>0.6</td>
</tr>
<tr>
<td>Duration (sec)</td>
<td>0.403</td>
<td>0.283</td>
<td>0.467</td>
<td>0.398</td>
<td>0.381</td>
<td>0.413</td>
<td>0.396</td>
</tr>
<tr>
<td>Power (normalized)</td>
<td>0.06</td>
<td>0.36</td>
<td>0.12</td>
<td>0.77</td>
<td>0.46</td>
<td>0.12</td>
<td>0.11</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Syllable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text</td>
<td>Mo mú i- gbá wá fún ẹ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tone Level</td>
<td>Mid</td>
<td>High</td>
<td>Mid</td>
<td>Low</td>
<td>High</td>
<td>High</td>
<td>Mid</td>
</tr>
<tr>
<td>Mean Pitch (ST)</td>
<td>49.2</td>
<td>50.2</td>
<td>46.8</td>
<td>43.5</td>
<td>44.4</td>
<td>47.9</td>
<td>46.0</td>
</tr>
<tr>
<td>Interval Direction</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>0</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>Pitch Difference (ST)</td>
<td>1.1</td>
<td>-3.5</td>
<td>-3.3</td>
<td>0.9</td>
<td>3.5</td>
<td>-1.9</td>
<td></td>
</tr>
<tr>
<td>Segment Slope (ST/sec)</td>
<td>2.0</td>
<td>-2.6</td>
<td>-8.0</td>
<td>-2.3</td>
<td>4.6</td>
<td>-4.0</td>
<td>-0.1</td>
</tr>
<tr>
<td>Duration (sec)</td>
<td>0.465</td>
<td>0.320</td>
<td>0.405</td>
<td>0.277</td>
<td>0.510</td>
<td>0.555</td>
<td>0.450</td>
</tr>
<tr>
<td>Power (normalized)</td>
<td>0.38</td>
<td>1.00</td>
<td>0.33</td>
<td>0.16</td>
<td>0.12</td>
<td>0.13</td>
<td>0.15</td>
</tr>
</tbody>
</table>

Table 1.5: Measurements for each syllable segmentation of the three homophous 7-syllable phrases

The first table section shows the text of the “igbá” (M-L) variant with a mapping to conventional speech tone levels (Low-Mid-High). The Direction of Change row indicates whether the speech tone of a successive syllable is higher (+), lower (-) or the same (0) as the tone of a prior syllable. The second table section is measurements of data from a segmented recording of the first phrase spoken. The Duration row is length of each syllable segment in seconds. The Mean Frequency is calculated using the YIN algorithm.
In Melodyne, one may quickly add and adjust note (or in this case syllable) segment boundaries. Onsets and offset values (in seconds) are exported as MIDI data to be used in MATLAB to analyze the continuous audio signal. In the plots, the black line is a fundamental frequency extraction using the YIN algorithm (de Cheveigne and Kawahara 2002). The red circles are mean logarithmic pitch (in semitones, C4=60) for each syllable segment (based on YIN) displayed at the midpoint (so they do not always align with the continuous f0 value). Color-coded dotted lines connect H (blue), M (green) and L (yellow) tones.

The speech tones differ from musical notes in their trajectories at the very least. Fundamental frequency in speech is generally less stable than singing or instrumentalism. In the upper plot, the mean pitch values are consistent for each tone level. Mids align with mids and highs with highs. In the lower, this is not the case. The second H is higher than the L immediately before, but is lower than all other Hs and Ms. A segment’s mean pitch does not necessarily represent the target pitch or the perceived pitch. It seems most credible that the plateau of the segment (right before 2 sec in the bottom plot) is representative of the target pitch of the speaker. This pitch is still within the range of the M tones, not the other H tones. What is more consistent between the phonological tone and the realized fundamental frequency is the syntagmatic direction between adjacent syllables. Speech that places tones within frequency bands is clear and intelligible. However, speech that does not adhere to frequency bands is also clear and intelligible. The do-re-mi folk heuristic preferred by Yorùbá speakers does not align fully with a

---

(de Cheveigne 2002). The Segment Slope is calculated using a Discrete Cosine Transform after Devaney et al (2011). The third and fourth sections are equivalent data for the “igbá” (M-H) variant.
syntagmatic or paradigmatic approach. Do-re-mi are not frequency bands, nor are they neutral pitches that comprise an interval. The history and development of the do-re-mi heuristic is addressed in Chapter 2 before formal linguistic models are explored in Chapter 3 and tested in Chapters 4 and 5.

1.1.4 SPEECH SURROGACY

The musical analogies do not end with the do-re-mi heuristic, and certainly did not emerge with colonialism. Speech surrogates are found in many ethnolinguistic cultures in throughout the world. In West Africa, “talking drums” are common such as the dùndún of the Yorùbá. Talking drums are used to recite proverbs, send messages or engage in a dialogue with a poet or neo-traditional singer. In this clip of Sàkàrà music from Lagos Island (<https://youtu.be/3MRhA7HvxTc>), the dialogue between the singer and drummer as well as singer and backup singers is antiphonal. The lead singer and the backup singers never sing together.

Figure 1.16: talking drum set (Wikimedia commons)
The lead drum of the dùndún family is the ìyáàlú, or mother drum, an association between pitch range and drum size with age and gender. Dùndún players almost certainly associate higher pitch with increased tension (see Chapter 1). Similar to the cricothyroid muscle, which elongates and thins the vocal folds, squeezing the hand on the tension cords of the dùndún stretches the drumhead increasing the pitch.

Much of southwestern Nigeria is ilé Yorùbá (Yorubaland). Ìgbò speaking areas are found in the southeast. The Ìgbòs also have pitched idiophones, usually in tuned sets such as bells or fixed-tension drums. They do not use variable-tension drums. This reflects different characteristics of tone realization in each language. In Ìgbò, the pitch trajectories of tones are more stable. In this clip recorded in Ọkọ, Anambra State (<https://youtu.be/JePyvA5NCKA>), a two-lobed ogene is the lead percussive instrument. The two tone levels (High and Low) of the language can be easily mapped to the lobes of the ogene. The voices are not antiphonal, but heterophonic. The leader cues in the other voices and sings with them. The ùdù pot drum makes a low “oom” sound when struck with the padded beater.
Different sizes of ùdù can be combined in a set and used as a speech surrogate, or simply play a fantastic bass line.

1.1.5 INDIGENOUS VOCAL ARTS

Many Niger-Congo vocal art traditions are panegyrical, in praise of gods, ancestors, people, places, or animals. There is also oral history, narrative and romance. The vocal continuum includes speech, chant, singing, and ululation. This is a broad space, but there are many fine distinctions of styles within cultures, according to vocal register and timbre. Ensemble singing may be antiphonal, heterophonic, or more rarely homophonic.

If tone is highly contrastive in a language, then the vocal arts may have special features. In Ìgbò, the homophone /akwa/ has four tonal variants, all of which appear in the tongue twister in Table 1.6.
<table>
<thead>
<tr>
<th>Ìgbò</th>
<th>English</th>
<th>/akwa/ tone</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nwány n’ákw’ákwa</td>
<td>Woman sewing cloth,</td>
<td>HL</td>
<td>cloth</td>
</tr>
<tr>
<td>i n’ákw’ákwa</td>
<td>are you crying</td>
<td>HH</td>
<td>cry</td>
</tr>
<tr>
<td>n’ókúkó yir’ákwa</td>
<td>because a hen laid an egg on</td>
<td>LH</td>
<td>egg</td>
</tr>
<tr>
<td>n’énú ákwà i kwár’ákwa</td>
<td>cloth you’ve already sewn</td>
<td>HL</td>
<td>cloth</td>
</tr>
<tr>
<td>nó n’énú ákwà?</td>
<td>which is on top of the bed?</td>
<td>LL</td>
<td>bed</td>
</tr>
</tbody>
</table>

Table 1.6: Ìgbò tongue twister

Such tonal wordplay is common in tone language poetry and other vocal arts. Because of the importance of pitch contours in poetry, the distinction between poetry and song is less secure.

Ọlátunji describes a device in Yorùbá poetry he calls “tonal counterpoint” (1984:30). This is exemplified by two phrases with parallel syntactic and metric structure that rhyme phonically but have inverted tone on the final word(s) (see Table 1.7).

<table>
<thead>
<tr>
<th>Yorùbá</th>
<th>Tone</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ejiré ñbá bí, ñbá jó jó jó</td>
<td>MMHMHHMHHHH</td>
<td>If I give birth to twins, I will dance!</td>
</tr>
<tr>
<td>Ejiré ñbá bí, ñbá yó yó yó</td>
<td>MMHMHHMHLLL</td>
<td>If I give birth to twins, I will rejoice!</td>
</tr>
</tbody>
</table>

Table 1.7: Tonal counterpoint in “Oríkì Ejiré” (praise of twins) by Mayowa Adeyemo (2013)

This poetic device is addressed at the end of Chapter 11 as a contour transformation.

Chapters 12–14 address indigenous and neo-traditional vocal arts from both an ethnographic and formal analytical perspective.

1.1.6 TONE-TUNE MISMATCH

In the nineteenth century, European missionaries made a lot of positive contributions to African societies. They contributed to the end of the slavetrade, created Latinic
orthographies for Niger-Congo languages, and established parochial boarding schools, which for better or worse proliferated Western ways on the continent. Along the way, there were many well-intentioned failures. Notably, hymn texts were translated metrically into African languages without awareness or regard for lexical tone. If one is not mindful of the tonemes when composing melodies for Ìgbò lyrics, tone-tune mismatch may occur (Èkwùèmé 1974b:337). An example is the Ìgbò translation of “All hail the power of Jesus’ name”, “Oha kele ike Jesu”, which is typically sung to Diadem or Coronation.

Figure 1.18: Coronation hymn-tune

Ọhà means all, kèlé means hail, and ìkè means strength or power. The syntax is a noun, a verb, a noun, and a noun. However, when this phonic content is sung to the hymn tune Coronation, it becomes a different noun, verb, noun, noun: Trees hail the buttocks of Jesus.

<table>
<thead>
<tr>
<th></th>
<th>Ĭghà</th>
<th>kèlé</th>
<th>ìkè</th>
<th>Jè-sù</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>V</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
</tbody>
</table>

Table 1.8: Intended and unintended meanings of /oha kele ike jesu/

The melody of Coronation can be re-composed to fit the tonemes of the Ìgbò translation.
This is not the beloved hymn-tune *Coronation* and is not a very interesting melody.

Another option is to compose a new melody altogether.

Tone and tune comparisons have typically used adjacent comparisons of pitch height analogous to Friedmann’s contour adjacency series (CAS) and sympathetic to a syntagmatic approach to tone.

Table 1.9: Contour Adjacency Series for Coronation, recomposition and folk melody

<table>
<thead>
<tr>
<th>Seg.</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text</td>
<td>ṑ-</td>
<td>hà</td>
<td>kè-</td>
<td>lé</td>
<td>i-</td>
<td>ké</td>
<td>Jé-</td>
<td>sù</td>
</tr>
<tr>
<td>CAS</td>
<td>NaN</td>
<td>0</td>
<td>NaN</td>
<td>+</td>
<td>NaN</td>
<td>0</td>
<td>NaN</td>
<td>-</td>
</tr>
<tr>
<td>Hymn</td>
<td>S</td>
<td>D</td>
<td>D</td>
<td>M</td>
<td>M</td>
<td>R</td>
<td>D</td>
<td>R</td>
</tr>
<tr>
<td>CAS</td>
<td>NaN</td>
<td>+</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>Rev.</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>R</td>
</tr>
<tr>
<td>CAS</td>
<td>NaN</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>New</td>
<td>D</td>
<td>D</td>
<td>M</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>R</td>
<td>D</td>
</tr>
<tr>
<td>CAS</td>
<td>NaN</td>
<td>0</td>
<td>NaN</td>
<td>+</td>
<td>NaN</td>
<td>0</td>
<td>NaN</td>
<td>-</td>
</tr>
</tbody>
</table>

The recomposition of the hymntune made for perfect agreement between underlying tone and realized tune in terms of syntagmatic direction, but it was not very aesthetic. As
Herzog originally suggested, perfect agreement may not be necessary. The folk melody does not have the same CAS as the tones of the text. Instead, the contour within each disyllable word is preserved. Stopping short of Èkwùèmé’s proposition that tone sequence determines the melody (1974b), a standard for best practice in text setting in a language with contrastive tone might be as follows: preserve directional relationships between adjacent tones in a word, especially in the case of words where lexical tone is contrastive (there are tone-varied homophones with other meanings). This is especially important for the initial presentation of each word. Tone-tune mismatch was not only a problem for nineteenth-century missionaries in Ìgbòland, but throughout West Africa (Euba 2001:121). After considerable exegesis of tonemes (Chapter 2–6) and melodic contour (Chapters 7–11), tone-to-tune mappings are revisited in Chapters 12–16.

1.2 Scope and Aims

This dissertation is a product of over five years of research on both the language and the music of the Ìgbò and Yorùbá, including nearly two years of fieldwork in southern Nigeria and domestic ethnography in immigrant communities in the United States. The map in Figure 1.21 shows the location of the Yoruboid and Igboid language clusters. Both clusters are in the West Benue Congo branch of the Niger-Congo A family (Greenberg 1963).
Lagos was my home-base for the majority of my 20 months in Nigeria (September to December 2011, February 2013–July 2014). Much of that time I stayed at a residence hall on the University of Lagos campus in Akoko, Yaba, Mainland Lagos, but I frequently traveled throughout southern Nigeria. Research sites include Anambra, Enugu, Imo and Rivers states in the southeast (Igbò land) and Kwara, Ogun, Oṣun states in the southwest.

Unless an exception is noted, none of the data presented, either linguistic or musical, is from monolingual speakers of either Ìgbò or Yorùbá. There are two reasons for this. The first is that all research sites were either urban areas, rural gatherings of city dwellers, or academic environments. Urban-dwellers generally have to be multilingual and formal English is required in the university setting. The second reason is that establishing informed consent is less complicated and more credible with bilingual

---

5 By Ulamm (Own work) [GFDL (http://www.gnu.org/copyleft/fdl.html) or CC BY-SA 3.0 (http://creativecommons.org/licenses/by-sa/3.0)], via Wikimedia Commons
speakers. All of the voice recordings are either presentations, performances or interviews with public speakers or vocal artists. The computer-key responses (for the study in Chapter 5) are from university-affiliated people, including students, staff and faculty. Data from focused monolingual speakers might yield different results, however such speakers are increasingly rare, particularly among ethnolinguistic groups that are highly urbanized like the Ìgbò and Yorùbá.

My initial interest was in the mapping of speech tones to musical melodies. As I dug farther into the tone-tune phenomenon, I realized that there were many open questions about the nature of tone contrast in these languages. In order to understand the realization of tone in music, including song and speech surrogacy, I sought a better understanding of the realization of tone in speech. The question evolved from is there a tone-to-tune mapping in a piece to what is the depth of the mapping? This led to the consideration of many forms of equivalence and degrees of similarity. Several years of thinking about phonological equivalence led to a modified approach to contour analysis presented in Chapters 8–11. My perspective on musical melody and polyphony has changed, not just with regard to African music, but toward music in general. This reflects significant growth for a musician who grew up as a pianist and became a harmony-focused composer-theorist somewhat oblivious to melodic nuance and invention.

In this dissertation, I will present a theory of lexical tone production and perception from the perspective of a music theorist, based in large part on contour theory and analysis literature; and, offer preliminary conclusions about aesthetics and form in Nigerian vocal arts and music, which may be extensible to music of the African diaspora. Throughout, I demonstrate methods for analysis of pitch features in speech and music.
Yorùbá sermons are compared to speeches by John F. Kennedy, Martin Luther King, Barack Obama and Donald Trump. In addition to Fújì singer Saheed Osupa and choral composer Laz Èkwúèmé, I consider the music of J. S. Bach, W. A. Mozart, Arnold Schoenberg, Richard Rodgers and Dizzy Gillespie. Contour levels reflect a general theory of pitch organization and perception in speech and music, based on my understanding of Niger-Congo tone systems.

1.3 Interdisciplinarity, Terminology and Notation

A breadth of sources and trans-disciplinary goals make it necessary to clarify terminology and formal notation. The primary sources of ambiguity are words that invoke different nuances of meaning for scholars of different fields and specializations. The terms I will use that have similar but not uniform cross-disciplinary usage include: paradigmatic, syntagmatic, segment, tone, contour, intonation, range and register.

1.3.1 TERMS

Paradigmatic and syntagmatic inform my understanding of the other terms. The concepts of paradigmatic and syntagmatic date back to Ferdinand de Saussure (1857–1913) and have been applied extensively to subfields of linguistics, from phonology to syntax. They have also been applied in music theory, in the analysis of both post-tonal (McCreless 1991) and tonal music (Rings 2006, 2011).

In my understanding, a paradigm is an associative category. Two objects belonging to the same paradigm are in some way equivalent. For some interpretations of
sound segments, such as segmental phonemes, the categories correspond to absolute values or a range of values (e.g. vowel formants). For other interpretations of sound segments, such as pitch height or duration, the category does not correspond to stable values. A syntagm is a relationship or pairwise comparison between two objects. This may be as generic as a binary equivalence relation (=,≠), ternary categories (>,=,<), or as specific as a vector with direction and distance (+9 semitones). For much of this dissertation, syntagmatic direction is of primary concern, which is equivalent to the concept of a ternary contour comparison of pitch height (+,0,−) or the directional component of a melodic interval. In short, my use of the terms paradigmatic and syntagmatic is consistent with phonologists such as Laura Dilley (2005), which in turn is similar to music theorists, not syntacticians.

My default use of *segment* will be to refer to portions of an audio signal on the order of a single note or syllable; however, in later chapters (Chapters 8–16) it may also refer to groupings of segments, e.g. melodic segment or contour segment. I avoid the use of the term *segment* to refer to a single segmental phoneme (e.g. /p/). Though it is cumbersome, I will consistently refer to what is more traditionally a segment or phoneme in linguistics as a segmental phoneme to avoid ambiguity. The segmental phoneme is not of central importance because I am most concerned with tone-bearing segments which are typically syllables with a sonorant nucleus (vowel or nasal) carrying the pitch. To further illustrate the difference between paradigmatics and syntagmatics, note that a single tone-bearing *segment* cannot have a syntagmatic value, but two or more adjacent segments, as in a melodic segment, can.
To be consistent with linguistic theory, I operationalize a *tone* as the paradigmatic tone level (e.g. Low, Mid or High) of a tone-bearing unit, usually a syllable consisting of a single vowel (V) or consonant-vowel (CV). In my mind, it is not interchangeable with *toneme* because there is evidence to suggest that the contrast is between tones not in and of tones themselves (see Chapters 4–5). As we shall see, syllable segments with the same tone may have different pitch. Thus, *tone* describes relative pitch height, in the context of other tones, a speaker’s range and paralinguistic intonation, more than absolute pitch. My notion of a *toneme* is most fully developed in Chapter 11 with the concept of an *interlevel*.

In linguistics, contour may refer to pitch trajectories within a single tone-bearing segment as in Mandarin (as in Figure 1.10). Tone-bearing segments may be divided into multiple tones to reify a pitch trajectory as a sequence of discrete level tones as in autosegmental phonology, but I will not be dealing with Mandarin or other “contour” tone languages further. Unless otherwise noted, each tone-bearing segment or syllable (which may consist of a single vowel) will be discretized as having a single tone and measured with mean pitch in semitones with reference to C4 as 60 (the conventional MIDI encoding for pitch). *Contour* refers to the syntagmatic relationship(s) between two or more adjacent segments in a grouping, consistent with the term’s usage in music theory.

In linguistics and music, *intonation* is a fluctuation of pitch that does not alter the paradigm (unless it is out of control). The allowable size of effect is quite different. In music, C4 is still C4 even if it is 30 cents sharp. In Yorùbá, high tone is high tone whether someone is speaking with elevated or depressed intonation, but the effect is
potentially much larger, with the placement of a high tone varying by as much as an octave within one speaker’s range (see Chapter 4). The similarity between intonation in music and language is that very gradual changes in intonation are less noticeable to most listeners. The dissimilarity is that intonational variation is desirable in speech but usually not aesthetic to musical performance (though there are exceptions, such as Tahitian choirs).

Range and register are fairly apparent. Range is the pitch extent of a voice or instrument and register is some portion of that range. In voice (singing) pedagogy, register often overlaps with a mode of production and the portion of the range in which that mode of production works, e.g., head voice, chest voice, whistle. Here, registers are not constrained to specific sections of the range, they are simply any portion of the range. No breaks or divisions in the range are acknowledged here, though that might be a useful refinement in the future. Operationally, I will use register as a discretization or time-segment of intonation. If intonation is patterns of pitch on a large-temporal scale, a register is a slice of that. The register is an active region within the total range (see Chapter 4 and Chapter 10).

1.3.2 NOTATION

In general, I will follow basic conventions within phonology and music theory. Segmental phonemes in IPA will be indicated by forward slashes, e.g. /a/, and phonetic transcriptions, also in IPA, will be indicated with square brackets, e.g. [a]. Little attention is paid to small variations in vowel quality, but the phonemic-phonetic distinction
becomes important with underlying tone versus realized tone. For example, /nì orúkọ/ (‘in the name of’ in Yorùbá) is realized as [l’órúkọ]. Tone marking conventions vary significantly from language to language. In Ègbò, high tone is the productively unmarked or neutral tone, and some orthographies leave it unmarked. I mark it (´) to be consistent with the Yorùbá orthography which marks high tones and leaves the middle (neutral) tone unmarked. To differentiate between mid tone in Yorùbá and downstepped high tone in Ègbò, the latter is marked with a macron, e.g. [ā].

A number of formalisms in music theory overlap in their notation. In musical set theory, Forte’s Adjacency Interval Series (AIS) (e.g. <1,3,2,6>) and Interval Class Vector (ICV) (e.g. <111111>) both use angle brackets, but differ in comma usage. Of these two, only AIS will be used and it is modified with the interval that wraps around in pitch class space excluded. In contour theory, Friedmann's Contour Adjacency Series (CAS) uses angle brackets and no commas (e.g. <+–+–>), as does Marvin and Laprade's CSEG equivalence classes (e.g. <03142>). The CAS is distinct from other tools in the use of operands (‘+’, ‘0’ and ‘–’). However, I will often use Quinn’s binary C+ Ascent values instead of the ternary contour values: 1 for + and 0 for 0/–. This will be in matrices so it should not be confused with other binary values that will be used. Lewin’s binary states are unique in appearance (e.g. @010) but the binary state changes are not (e.g. <010>) (1995). The binary state changes look very similar to a number of other tools, so I will modify the notation to be unique by using a delta (Δ010).
<table>
<thead>
<tr>
<th>Name</th>
<th>Abbreviation</th>
<th>Notation Example</th>
<th>Meaning of Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adjacency Interval Series</td>
<td>AIS</td>
<td>&lt;1,2,3&gt;</td>
<td>A tetrachord with 1 semitone between the 1st and 2nd pitch, 2 between 2nd and 3rd, 3 between 3rd and 4th.</td>
</tr>
<tr>
<td>Contour Adjacency Series</td>
<td>CAS</td>
<td>&lt;+–+–&gt;</td>
<td>A melodic segment of cardinality five with a rise, a fall, a rise, then a fall.</td>
</tr>
<tr>
<td>Binary State (Inventory)</td>
<td>@</td>
<td>@010</td>
<td>An inventory of three types with only the middle type present.</td>
</tr>
<tr>
<td>State Change</td>
<td>Δ</td>
<td>Δ010</td>
<td>A state change to only the middle type, if applied to the binary state above it would leave the null state: @000.</td>
</tr>
</tbody>
</table>

Table 1.10: Formal notation to be used

Because the distinction between paradigmatic and syntagmatic representations is key, this is the basis for a distinction in the notation. Angle brackets will be used for ordered series of intervals (syntagms) including the AIS and CAS, and square brackets will be used for ordered series of objects, whether absolute or abstract.
§

I. Tone
2. THE DO-RE-MI FOLK HEURISTIC

Introduction

In the Nineteenth Century, an English woman, Sarah Ann Glover (1785–1867), believed singing was for the public good and a Yorùbá man, Samuel Àjàyí Crowther (1809–1891), thought speech tones should be preserved in writing. Their stories illustrate how diversity in thought sometimes struggles to have an impact, but can ultimately shape human consciousness; and, that distinct ideas with disparate aims may be creolized in a period of rapid social change. While the outcome shows a positive side of the missionary field, bringing people and ideas together, the transmission of Glover’s and Crowther’s ideas was mediated by the overlapping political, social and cultural hegemonies of the colonial era. Crowther was celebrated in the English-speaking world as evidence the civilizing agenda—and colonialism—was good for all involved. Glover's innovations in music education have been misattributed to a few different men. This chapter draws on evidence from ethnographic work, field recordings, language surveys, and literature from a variety of disciplines. All of this information contributes to one answer to the question: why is do-re-mi the preferred heuristic for speech tone levels among bilingual Yorùbá speakers and teachers of Yorùbá language?

The presence of speech surrogates, such as the talking drum (dundún), indicate language and music have long had a close relationship in Yorùbá-speaking areas. Throughout Sub-Saharan Africa, missionary activity introduced Western forms of
literacy for both language and music concurrently. In an ethnolinguistic culture with a fuzzy boundary between language and music, a culture where drums can speak, it is not surprising a musical model was (and is) used to fill a void in the Western concept of what a language could be.

2.1 Sarah Ann Glover and the Tonic Sol-Fa Method

Solmization originated during the rise of musical literacy in the Carolingian era, musical literacy necessitated a way to teach musical literacy. Around 795 CE, Charlemagne wrote to the Abbot of Fulda requesting the bishoprics and monasteries ‘undertake the task of teaching’ because ‘without knowledge it is impossible to do good’ (Treitler 1984:135). Two hundred years later, Italian music theorist Guido d’Arezzo (c. 990-1030) introduced a more precise staff notation along with solmization syllables to make it comprehensible, ut-re-mi-fa-so-la, forming a hexachord. The growth in monastic education and literary tradition of hymns explain Guido's choice of an existing musical text and tune, “Ut queant laxis,” as the basis for his hexachordal (six-note) solmization system (Boynton 2003:100).
In Guido’s time, a plurality of notations existed with the common goal of transmitting texts with efficiency and fidelity (Treitler 1984:139, 207). Guido's staff notation and solmization were widely adopted throughout medieval Christian Europe. Staff notation continued to evolve, but the hexachordal system continued to be used for centuries. On the other side of the continent, English choristers at cathedrals and the Chapel Royal learned to recite the Guidonian gamut\(^7\) forwards and backwards well into the nineteenth century (Rainbow 1967:14–5). In this same period, a related but modified solmization system was introduced by an Sunday school teacher in Norwich, England.

\(^6\) By Romainbehar (Own work) [Public domain], via Wikimedia Commons <https://commons.wikimedia.org/wiki/File:%2AUtQueantLaxis-Arezzo.svg>

\(^7\) An earlier six-syllable (hexachordal) solmization system.
Sarah Ann Glover (1785–1867) had a conviction that teaching should emphasize practice and theory should be derived from practice, not the other way around (Bennett 1984:28). By reducing complexity—‘inadequate representation of the scale on the staff’, ‘non-accidental sharps and flats’, and the ‘contrivance of clefs’—Glover could implement practice swiftly (Glover 1982:16–7). Glover’s system Anglicized Guido’s syllables and added a syllable (te) for the seventh degree of the diatonic scale. Do-re-me-fah-sole-lah-te (Glover’s spelling) were used to sing the major scale and the same syllables starting from lah were used for the minor scale. Like Guido, her method included both a notation and a solmization but with an even more direct connection between the two. In Glover’s

---

8 By Unattributed (Bibliothèque Nationale de France) [Public domain], via Wikimedia Commons <https://commons.wikimedia.org/wiki/File%3ASarah_Ann_Glover.jpg>
notation, the pitches are represented by the first letter of the syllable and accompanied by a rhythmic tablature of dots and lines (see Figure 2.3).

```
\textbf{Figure 2.3: example of Tonic Sol-Fa notation}
```

Glover's attempts to apply her Sunday School 'experiments' in day schools were met with resistance, some believed teaching music at charity schools might be detrimental to the public good. It was the influence of John Curwen (1816–1880) that overrode these concerns. Curwen transformed Glover's Sol-Fa method into a movement aligned with the temperance cause, ensuring no one could associate Sol-Fa singing with societal ills (Bennett 1984: 29). By the 1850s, the ‘Sol-Faists’ were a community of thousands.

Curwen struggled with not only teaching music but reading it himself until a friend called his attention to Glover's simple and direct method in the spring of 1841; he found he was able to teach children sol-fa notation within a few weeks, and campaigned for the method with voracious zeal (Bennett 1984:27). Curwen and Glover soon began a correspondence. They disagreed whether tonic sol-fa notation was an end in itself or path to full music literacy, Curwen was satiated with the former while Glover stressed the latter (Bennett 1984:33).

Curwen was less concerned with students of Sol-Fa developing full musical literacy in the professional sense. This is consistent with Sol-Fa as a popular movement outside of the academy. According to Alexander Ellis, the contemporary English translator of the work of German physicist Hermann von Helmholtz (1821–1894), the "glory" of Sol-Fa was learning a sense of relative pitch (Olwage 2010:202). In the 19th
century, fixed-do solmization was preferred in the continental conservatories and, to some extent, the English academies of music. Curwen conceived an affinity between the Sol-Fa system and just intonation. His passion for just intonation was no doubt influenced by von Helmholtz, whom Curwen frequently cites in his own treatment on acoustics and music theory, *Musical Statics* (first published in 1874). Von Helmholtz and Curwen were both dogmatic about just intonation. Von Helmholtz characterizes the professional vocalist as a fallen singer while lauding the just singing of “unpracticed” a cappella part-singers (Helmholtz 1912:326). Similarly, Curwen used an anecdotal assessment of the black Fisk University Jubilee Singers, who visited England in the 1860s, as evidence that singing in just intonation was “natural” (Curwen 1897:82). Helmholtz insisted that all vocalists and string players, in the absence of corrupting (tempered) influences sang pure intervals. He chose a rather exceptional example, citing the playing of renowned concert violinist Joachim. George Bernard Shaw also wrote about Joachim’s distinctive tuning of intervals, calling it the “Joachim Mode” (Hui 2012:76).

Curwen, who regarded von Helmholtz as a genius, would have been thrilled when von Helmholtz was moved to add a new appendix to the 1892 edition of *On the sensations of tone* called “Just Intonation in Singing” (Helmholtz 1912:422–8). This was after a trip to England where he attended performances at the Crystal Palace at Sydenham. In the appendix von Helmholtz lauds the Sol-Faists for singing “by natural, and not by tempered intervals.” The evidence was “differences and disturbances” caused when the choir was accompanied by a tempered organ versus a specially-crafted enharmonic organ (1912:427). Despite the many instances of “natural” just intonation von Helmholtz and Curwen witnessed, Barbour suggests that just intonation may be a myth:
... just intonation has been defined in Stainer and Barrett's Dictionary as “singing or playing in perfect tune”, “the correct sounding of intervals in singing or playing”. However one may deplore the lack of precision in this definition, it must be admitted that some such meaning exists in the minds of the general musical public... Choral societies and string quartets, freed from the tyranny of fixed intonation, are supposed to use just intervals, thus interpreting music as intended by the composers. Such a conception is essentially false. There is no system of tuning that has the virtues popularly ascribed to just intonation. Neither singers nor violinists use just intonation. Furthermore, as it is usually defined, just intonation is a very limited, cumbersome and unsatisfactory tuning system.” (Barbour 1938:48)

Despite any anachronistic doubts about its relationship to just intonation, Tonic Sol-Fa took “its place alongside that greatest of Victorian truths – Christianity – in the ‘civilizing mission’ at home and in the colonies” (Olwage 2010:196).

Movable-do systems9 have been adopted widely in English and American music education. However, Glover, and to a lesser extent Curwen, have largely been neglected in histories of the period while John Hullah (1812–1884) is often celebrated for his impact on Victorian music education. Unlike Glover and Curwen, Hullah was a musical prodigy and was trained at the Royal Academy of Music. A main selling point of the Sol-Fa method was its ease and effectiveness. Hullah was not interested in amateur music-making, his priority was to raise the standard of professional musicianship in England and to improve the nation’s reputation as a producer of high culture. Hullah campaigned against Tonic Sol-Fa in the schools and advocated for a fixed-do system as was taught in much of continental Europe at the time (Leinster-Mackay 1981:165). Despite its tangible impact on amateur singing culture in England, music historian George Grove omitted the Sol-Fa movement in his telling of the Victorian Musical Renaissance (Olwage 2010:193).

---

9 Moveable-do is based on relative pitch. It enables the same solmization syllables to be used to sing equivalent scales in different keys (e.g. the key of C, the key of G). There are twelve major keys in the Western tonal system.
A century later, Bernarr Rainbow drew attention to the efforts of Glover and Curwen in *The Land Without Music* (1967). In American music education, Zoltán Kodály is given credit for a system he adapted. Sometimes Curwen is cited as his source, but rarely Glover. In several recent books and articles, this oversight has been corrected if not fully recognized. The issues Glover raised persist to this day. In a series of papers presented to the Music Theory Society of New York State, Martin and Surace asked such questions as ‘Does one “see” functional relationship on the musical staff?’ (1978:24).

In the late Nineteenth Century the Tonic Sol-Fa Society joined forces with the missionary movement, similar to the earlier alliance with the temperance movement in the 1840s. In 1857, Curwen began to publish testimonials and reports from stations in Barbados, India, Africa and China, including one from Old Calabar in present day Nigeria. A missionary to China, John Fryer, reported that Tonic Sol-Fa ‘formed a bond of union between teachers and pupils’ (McGuire 2009:130). Although Sol-Fa was part of missionary activities before, the first missionary trained by Curwen was Robert Toy, who was sent to Madagascar in 1862 by the London Missionary Society (LMS) (Southcott 2004:3). In Madagascar and elsewhere, learning Tonic Sol-Fa, along with European dress and language, became a rite of passage to conversion and an important symbol of control recognized by the colonizers (McGuire 2009:128).

According to Olwage (2010), in the Cape Colony (now South Africa), choral singing was a method of disciplining colonized peoples, demonstrating they could be civilized through the work of missionaries. The novel *The Black Peril*, published in the Cape Colony in 1912, depicted white women having sex with black men, swelling fears

---

that blacks might come to dominate the society. During this time of heightened anxiety, missionaries assured the white-male power class that blacks could be controlled with Christian means, partially through choral concerts (Olwage 2005:35). Two decades earlier, a tour by a black choir from the cape colony had a similar effect on English audiences, convincing them non-Europeans could be civilized through missionary education including a healthy dose of Sol-Fa singing. Halfway through the performance, the choir changed from indigenous dress to Victorian clothing, intended and likely received as a ‘serious demonstration’ of progress (Erlmann 1999:128).

By the early Twentieth Century, the Sol-Fa method was present at many English and American-run missions, often with a diet of simple hymns like those composed by the American evangelist Ira Sankey (1840-1908). The loss of World War I led to the internment of German missionaries. As a result, many former German missions now taught Tonic Sol-Fa (Busse Berger 2013:482). Erich von Hornbostel was one of the first musicologists to express some anti-colonial sentiment, perhaps out of sympathy from his experiences with imperialism as an Austrian. In an article intended for a general audience with a commentary on missionary work, ‘African negro music,’ von Hornbostel states that while Africans can be taught to sing hymns ‘such as they are’ or encouraged to produce songs in the European fashion, this would mold Africa as a ‘mere European colony’ much like the Americas. Instead, he suggests Africans be ‘encouraged to sing and play in their own natural manner’ (von Hornbostel 1928:61–62). Few followed von Hornbostel's recommendation, but Busse Berger researched a missionary who did: Franz F. Rietzsch worked in Tanganyika (now Tanzania) in the 1930s, was a member of the
German folk-song movement, Singbewegung, and had training in comparative musicology (2013:482).

The 1950s and 1960s independence movement in Africa largely brought the protestant missionary era to an end in Africa. Now, many Africans are evangelists of both Christianity and Tonic Sol-Fa. Music educator Robin Stevens writes that the method continues to thrive in locales as far as Fiji where it is a mainstay of community singing and a highly effective ‘alternative to staff notation’ (2003).

2.2 Samuel Àjàyí Crowther and African Tone Systems

For centuries, ports at Badagry and Ouidah in the Bight of Benin were busy with human trafficking to Cuba, Haiti, Brazil and elsewhere. Many of the people traded at these ports belong to what is now known as the Yorùbá ethnic group. As a result, aspects of Yorùbá culture, including the language, religion, food, and music, are now found throughout the Americas in addition to their ancestral home in present-day Nigeria and Benin. The act abolishing British involvement in the slave trade was passed by their parliament in 1807. Soon after, the Royal Navy began intercepting outbound slave ships along the West African coast.
On one of these ships was a young man, Àjàyí, who would be reborn as Samuel Crowther (1809–1891) in the missionary community in Freetown, Sierra Leone.

Samuel Crowther's career was unique. A kidnapped slave in 1821, a rescued slave in 1822, a mission school boy in 1823, a baptized Christian in 1825, a college student in 1826, a teacher in 1828, a clergyman in 1843, a missionary to the country whence he had been stolen in 1845, the founder of a new mission in 1857, the first negro bishop in 1864—where is the parallel to such a life? (Page 1908:vi)

Crowther received his Doctorate in Divinity from the University of Oxford in 1864. His identity as a ‘Black Bishop’ piqued the interest of evangelical communities in London and New York where multiple biographies were published. His legacy in Nigeria is closely tied to the publishing of the first Yorùbá primer and vocabulary in 1843 and a complete orthography, grammar and dictionary in 1852. Crowther went on to publish primers and vocabularies for Ìgbò (1857) and Nupe (1860, 1864).
Like Crowther, many of those freed from slave ships and educated in Freetown eventually returned to their homelands. In the Niger territory, they were known as Sàròs (after Sierra Leone) and occupied esteemed but ultimately restricted positions within the colonial system. As bureaucrats, clergy and educationists, the repatriates utilized dual identities of being Africans and Western-educated Christians to mediate between the colonists and the colonized. Crowther made colonial administration and evangelism all the more possible with his orthography. On the other hand, the early adoption and sustained use of a Romanized orthography has contributed to an ethnolinguistic culture that continues to be robust.

Crowther’s first task was to create a Pan-Yorùbá identity by collapsing the dialects:

---

11 By Ernest Edwards [Public domain], via Wikimedia Commons <https://commons.wikimedia.org/wiki/File%3ABishop_Samuel_Ajayi_Crowther_B.png>
Among the purest Yorùbá speakers, there are no less than three modes of pronouncing some words; namely, the Capital-or Oyo-pronunciation, and two Provincial dialects-the Ibapa and the Ibollo. People from all parts of Yorùbá are now together in the Colony of Sierra Leone, and each party contends for the superiority of its mode of utterance. (Crowther 1852:1)

Because Crowther recognized Òyó as the capital, Òyó dialect became the model for Standard Yorùbá (SY). The Church Missionary Society (CMS) became the arbiter of Yorùbá literacy. In 1875, the CMS convened a conference to standardize the Romanized orthography, a disappointment to Muslim Yorùbá scholars then and now (Ogunbiyi 2003:77). Despite this, Yorùbá people are united by their language and are not prone to the intra-ethnic religious conflict consuming northern Nigeria in the Twenty-First Century.

The success of the civilizing mission among the Yorùbá was lauded by general reports of the missionary field:

Yariba is the every-day Language of teaching and preaching of a large Mission at Lagos and Abeokúta. The whole Bible is in the course of publication... The Yariba people are full of energy, and from their ranks several men have already sprung up of high attainments, and we may look forward to this Language being one of the most important in Western Africa. (Cust 1883:207)

While Cust was a bit off on the spelling, his prediction about Yorùbá’s virility in the future was correct. English is the national language of Nigeria, but Yorùbá people continue to take pride in their language and cultivate a pan-Yorùbá identity around it, ignoring smaller differences in dialect and custom, and sometimes major differences in religious belief. Findings from our 2014 survey indicate Yorùbá language use with family is high among respondents who identify ethnically as Yorùbá. Figure 6 consists of four bar graphs of language use data by state. According to the results, Kwara State is exceptional among predominantly Yorùbá states in that roughly half of respondents of all ages reported using Yorùbá primarily in casual settings (such as conversation with
friends). This is the only state for which Yorùbá language use in casual settings outweighed English language use.¹² Our ethnographic evidence supports the statistics: the city of Ilọrin is very welcoming but does not acquiesce on its linguistic identity. Formal business in schools, universities, banks and government offices is conducted in English, but Yorùbá is dominant in all other situations. Visitors and newcomers are expected to know some Yorùbá (or learn some fast) whether or not they are ethnically Yorùbá. At the University of Ilọrin, non-Yorùbá students are fluent in Yorùbá. In Lagos, the assumption is that non-Yorùbás do not speak Yorùbá and speaking Yorùbá is often used as an exclusionary code-switching device. The insistence on Yorùbá language use in Kwara state can be viewed as inclusive, not xenophobic. It should also be noted that Ilọrin is predominantly Muslim and this is often the case for highly focused Yorùbá language environments.

¹² Because the survey respondents come from an ethnically diverse but otherwise demographically homogeneous population (bilingual with Western-style education), it is presumed Yorùbá language use is higher among the general population in these states. Comparing data by state brings more useful interpretations than generalizations based on the entire data set.
Cust also notes a peculiar aspect of Crowther’s approach to orthography: Crowther insisted on the importance of tones in both the language and the orthography (1883:229). Cust’s survey of the missionary field makes it clear that Crowther’s approach to tone was novel at the time.

The Yorùbá language is very musical: certain marks to distinguish the tones thus become \textit{indispensable}. Two accents have therefore been used to point out this distinction, i.e. not to imply that a particular stress is to be laid on the accentuated syllable, but to mark a variety of intonation. The accents thus employed are, the
acute, indicating elevation of tone... [and the] grave, indicating depression of tone... (Crowther 1852:3; emphasis added)

When Crowther developed the orthography in the 1840s, there was no precedent for accommodating the strong presence of lexical tone found in many Niger-Congo languages. I found that over 40% of the disyllables in the concise and widely available Ìbàdàn University Press Yorùbá dictionary depend on tone to be differentiated from other entries. In a more thorough dictionary, such as Abraham (1962), I would expect this percentage to be higher. There is a comparison to lexical tone in Chinese in the preface to the 1852 edition. However, the Romanized script for Chinese with diacritics, Pīnyīn, was not developed until the Twentieth Century. It is not made explicit, but it is likely the polytonic orthography for Ancient and Medieval Greek pitch accents—(´) for high, (´) for low—were the closest model available to Crowther.

Crowther’s prescription has largely been sustained, but tone-marking remains alternately bewildering or irritating to many speakers and a recent dictionary advocates ‘eliminating the tonal signs’ (Fakinlede 2003:9). The concept of lexical tone was almost inconceivable to missionaries who only spoke European languages without contrastive tone. The missionaries approached orthography for the hundreds of languages in Africa from a very different perspective. Crowther had a distinct advantage of being a first-language (L1) speaker of Yorùbá who had immersive training in English. Of the European missionary-linguists, Johann Gottlieb Christaller (1827–1895) of Basel was unique in his careful consideration of West African tone systems (Bearth 1998:85). Written some years later and in description of Akan in the Gold Coast (Ghana), Christaller's description of Niger-Congo tonology is largely consistent with, and likely influenced by Crowther's.
The great variety of vowels is increased by different tones, every syllable of every word having its own relative tone, equal with or different from the neighbouring syllables, either high, or low, or middle, sometimes in successive degrees. (Christaller 1875:XVIII; emphasis added)

Some skeptics still doubt the necessity of marking tone for a language like Yorùbá or Akan. However, if one accepts that marking tone is necessary, the issue then becomes efficiency: should tone always be marked or just when it provides lexical contrast? There are many words of the same class, noun and noun, verb and verb, in which vowels or consonants are allophonic (non-contrastive) while tone is phonemic (contrastive). Are these the only words in which tones should be marked? What about the syllables themselves, should all syllables within a word carry diacritics? Crowther's grammar advocated for marking the first high or low tones in a sequence of consecutive high or low tone, such that the second syllable would only be marked in a four-syllable word of mid-high-high-high tones. This is problematic because a return to the mid (unmarked) level in the third or fourth syllable since cannot be indicated. In Bamgboše's orthography (first published in 1966), each occurrence of high or low is marked, solving the problem.

Christaller's system largely followed Crowther's in terms of leaving consecutive equal tones unmarked but introduced some complex rules for step tones and short vowels (Christaller 1875:15). While Rev. Christaller had a much greater sensitivity to tone than the vast majority of missionary-linguists, he adopted a laissez-faire attitude about when tone marks should be applied:

In common writing and in books for the people we mark the tone only in cases of ambiguity; but in grammar and dictionary, and for the study of the language by foreigners, an accurate designation of the tones and the stress is necessary. (1875:15)

However, in the preface he admits:
[The tone-marks] are also wanting on many words of this dictionary, either from uncertainty or oversight, or because the tones may be known from analogy or simple rules. (1875:XXV)

Crowther’s orthography and dictionary does not omit tones because he does not need to, he is a native speaker. Much like the misattribution of Sarah Ann Glover's Tonic Sol-Fa method, similar systems were widely used by linguists over the next century, often citing Christaller not Crowther. Welmers’s *African Language Structures* (1973), which will be drawn on in the next chapter, lists Christaller (1875) as the first of some “fortunate exceptions” to the “general disregard for tone” (1973:77), completely overlooking Crowther’s more original and thorough documentation of tone in a more prototypically tonal language.

### 2.3 Origins of the Musical Analogy

Unsurprisingly, the first reference to Sol-Fa and tone is by a musician, a former organist of the Christ Church Cathedral in Lagos, not far from the Musical Society.

Yorùbá is supposed to have only three tones. There are some who go further to assert that these three tones are fixed and can be represented by Do, Me and Soh. These ideas are quite erroneous. The positions of the tones may be principally three, but not only may each of these, especially the medium be slightly higher or lower, but the speech tones do not strictly follow the three Solfé tones. The system that I propose to use is that of a three-line Staff, with provision made for the use of the space as well as the lines, as in music. (Phillips 1952:1)

Phillips is the first, to my knowledge, to point out in published literature that the positions of the tones are not fixed. Do-mi-so is actually more similar to the spacing than do-re-mi, but both equate the speech tones with stable pitches. Drawing on a staff does give the option of indicating contours by connecting pitch events across the lines and spaces. This
is useful because there are circumstances in which tones are stable and others in which tones are sloped. In ‘The Assimilated Low Tone’, Bamgbose describes the circumstances under which low-tone has level pitch (after a low-tone) and is low-falling (after a mid or high), high becomes low-rising after low and mid is low-mid after low (Bamgbose 2010: 9). Although capable of indicating these pitch trajectories, Phillips's suggestion of using the staff is more elaborate than the simple diacritics so many already find cumbersome, and is more appropriate for transcription than a streamlined orthography. This musical staff has been used as Phillips suggested by music researchers such as Adégbítè. The comparison of tones to musical pitches is not restricted to musicians however, linguists have also contributed to the conflation. Writing on another Nigerian language, Jukun, Welmers states:

The three levels [of Jukun] are discrete throughout the sentence, and so precisely limited that playing them on three notes on a piano (a major triad does very well) does not appreciably distort the pitches of normal speech. (Welmers 1973:81)

While the tones may be understood if sung on do-mi-so, the actual phonetic implementation for these languages is quite apart from this: production and intonational effects such as downstep and high-rising put tone levels in constant flux (Laniran and Clements 2003:203). Despite the complexity of African tone systems and the manifold ways musical pitches are not like speech tones (see Ladd 2008b), the do-re-mi folk heuristic has gained traction with the Christianized public and academy alike. A popular text found in street markets and used in schools in Nigeria includes Hausa, Ìgbò and Yorùbá vocabulary (the three major indigenous languages of Nigeria) and was written to “generate unity.” Yorùbá is the only language in the text for which tones are marked: “accents are used over Yorùbá words to denote their sounds which are doo, ree, mii ... ree
has no visible tone mark except on nasalized syllables” (Odetunde 2009:1). A foreign text
by a bona fide linguist, Professor Antonia (Yetunde Folarin) Schleicher of Indiana
University, also uses the same folk heuristic:

Each unit in this book has a tone exercise to help you learn how these tones are
pronounced in different words. You can use the musical notes ‘do, reh, mi’ to
help you learn how to pronounce the tones: low tone is ‘do’; mid tone is ‘reh’;
high tone is ‘mi’. (Schleicher 2008:XV)

Although Schleicher is targeting a non-Yorùbá audience, school teachers and even
linguists in Nigeria use the do-re-mi heuristic to describe the tone levels and for
themselves, as a mnemonic aid in transcribing speech-to-text. Do they have experience
singing in a choir? Most likely, many Nigerians do. The current orthography, reflected in
Bamgbose's grammar (first published in 1966) makes the analogy more appropriate,
because the underlying tones are all conceived of as discrete (low, mid or high). To
manage this, he divides long syllables with contour tones into smaller units: ‘The so-
called glides ... recognised as additional tones by many scholars... are treated in this
system as separate tones occurring on a sequence of two syllables...’ (Bamgbose 2010: 6).

A new Yorùbá grammar by Rutgers University professor Akinlabi is forthcoming. Some
years back, in a chapter for laypeople on Yorùbá orthography, he states:

... there are three contrastive tones, a one syllable word may have a three way
pitch contrast... e.g. ko (H) (build), ko(M) (sing), ko(L) (reject). Therefore tones
are like consonants and vowels in Yorùbá, since they distinguish the meanings of
words like consonants and vowels do. (2004: 459–460)

Akinlabi’s research has covered complex and novel topics such as underspecification and
clitic assimilation of tone, all of which point to the relationality of tone. Yet, for a
description for lay people, Akinlabi’s illustration of phonemic tone does not deviate from
Crowther's description 150 years earlier. Both used monosyllables to introduce the
concept. Despite a century of research, the basic concept of the tonemes as ‘atomic units’ has not changed. Shortly before his death, tonologist Nick Clements questioned whether tone features were motivated at all, reasserting that they “do not serve the same functions as segmental features”\(^\text{13}\) (Clements \textit{et al} 2011:3). Akinlabi’s example of a monosyllable with contrasting tones, like Crowther’s example 150 years earlier, is misleading because it is not clear how an isolated syllable can have contrastive tone. Crowther’s and Christaller’s early descriptions refer to ‘relative’ tone and ‘elevation’ or ‘depression’ of intonation, implying the pitch relationship between nearby syllables is important. However the temptation to use a monosyllable example of contrastive tone persists.

Features of segmental phonemes, such as those that combine to create /a/ (+open, +back) are easily discretized because they are absolute (paradigmatic) features. However, tone features, may be perceived in terms of relative pitch (syntagmatic relationships).\(^\text{14}\)

Although the conflation of tone and tune has faults, the folk heuristic of \textit{do-re-mi}, implying moveable pitch relationships, has advantages over the atomic units \textit{low}, \textit{mid} and \textit{high}.

\textbf{Discussion}

For several years, I have worked with colleagues on transcriptions of Yorùbá vocal arts, including poetry and song, with particular attention to accurately recording tone.

Studying Yorùbá poetry is challenging because it includes \textit{iijinlẹ} (deep) language that is not only untranslatable, but has no synonymy within Yorùbá. Because there are words

\(^{13}\) Clements is referring to features of segmental phonemes (e.g. single vowels and consonants).

\(^{14}\) For further discussion of syntagmatic and paradigmatic tone intervals, see Dilley 2005.
not in dictionaries, these transcriptions are not only a record of a performance, but a record of the language. I often seek independent opinions on words, phrases or larger sections. Time and again, I have found bilingual speakers prefer to talk about tone as *do-re-mi*, not *low-mid-high*. The linguistic terms, *low-mid-high*, reflect an association between frequency, how fast or slow the sound wave is vibrating, and height. This conceptualization of pitch is shared with Western music theory, but conceiving of pitch as low or high is not found in all cultures. In Yorùbá culture, tension may be a better descriptor for pitch variance than height, This is suggested by talking drum (*dùndún*) performance practice.

Figure 2.7: (left) a drummer’s hand squeezing the tension cords of the *dùndún iyáálú* (mother drum); (right) vocal apparatus (housed in the larynx) viewed from above.

In *dùndún* performance, the lead drummer often engages in a dialogue with a poet-singer and can speak proverbs or common sayings. This is accomplished through a mapping of the tones of speech to pitches played on the drum. Pitch is changed by tightening or loosening grip on the tension cords connected to the drumhead (see Figure 2.7). The energy for the sound wave is supplied by striking the drumhead with a curved beater
(drumstick). A light presence of the hand on the cords produces mid-tone, squeezing them tightly produces high-tone, and releasing all tension is low-tone. The pitch-control mechanism of the talking drum is remarkably similar to the pitch-control mechanism of the human voice: variable tension. The vocal cords (or folds, Figure 2.7 right side) are pulled tight by the cricothyroid muscle, increasing frequency (pitch height).

Yorùbá speech is full of ideophones, words that evoke ideas or sensations. Despite the prevalent use of sound symbolism in Yorùbá, Ọkè, a word meaning ‘on top of’ or ‘up’ has low tones. In combination with voice (Ohùn), Ohùn-Ọkè means ‘high tone’, but it has low tones. Terms used to describe speech tone in Yorùbá are codified in Bamgbose’s Yorùbá Metalanguage (1990), and many of them appear in Abraham’s 1962 dictionary. They are not in colloquial use, nor do they appear in Crowther’s works. Most likely, they are later translations of linguistic terms from English into Yorùbá. In contrast to Ọkè, a word referring to height, words referring to tension better fulfill expectations of sound symbolism, or an ideophonic quality:

Tight (adj) hà, fún, le, mó, pinpin.
Tightly (adj) ni lilelile, gaga, daindain, ṣínṣín. (UI Dictionary¹⁵ 1991)

Many of these words have high tone and none of them have low tone. Is tension a better conceptualization of pitch variation than height within Yorùbá culture? Perhaps. Unless one is a researcher of the voice, the mechanism of the human voice is felt not seen. The dùndún is exceptional as an external embodiment of the human voice. The instrument is iconic within Yorùbá culture and in the past, before massive urbanization, the instrument was likely even more central to daily life. The dùndún is a tangible model of the voice

¹⁵ UI Dictionary henceforth refers to A Dictionary of the Yoruba Language (1991) published by the University of Ibadan press.
and it does not move up and down, it is tightened and loosened. Many discussions during fieldwork in Nigeria and much evidence in the literature on other cultures (e.g. Seeger 2004) suggest the pitch height concept is not universal. Voice range can be conceived of in terms of age, gender or size instead of height. The mechanism of the dùndún suggests tight tone may be closer to pre-colonial indigenous concepts of pitch than high tone.

In the Western classical music tradition, reading staff notation constantly reinforces the idea that pitch goes up and down, but Tonic Sol-Fa notation (see Figure 2.8 below) does not present pitch that way, there is no staff. It reads left to right in a chronological stream of do-re-mi-fa-so-la-ti (see Figure 9). Sol-Fa notation itself is agnostic about pitch height but Glover’s Sol-Fa ladder is not (see Figure 2). The notation (and not the ladder) was introduced in missions through Sub-Saharan Africa and still widely used (Southcott 2004, Olwage 2010).

![Figure 2.8: a Yorùbá Christmas song by David Àíná with staff and Sol-Fa notation](image)

Figure 2.8 shows both staff notation and Glover’s Tonic Sol-Fa notation (above and below the staff) for a Yorùbá Christmas carol. Although it appears in Àíná’s score, staff notation is not nearly as widespread a practice as solmization syllables and Tonic Sol-Fa notation. Tonic Sol-Fa notation is widely used by Nigerian choral composers and for amateur music-making by church and school choirs in southern Nigeria. When singing,
one feels the sensation of tightening (engaging the cricothyroid muscle) or thickening (engaging the Thyroarytenoid muscle) the vocal folds, which may or may not extend to a metaphorical notion of raising or lowering pitch. The action of singing the melody in Figure 3 is more than superficially analogous to the sequence of tightening and loosening grip necessary for playing the same melody on the talking drum.\(^{16}\) The height paradigm is now present in Nigeria. However, among the Christianized public, including language teachers and even professional musicians (who are very familiar with the musical staff), the do-re-mi heuristic is preferred. Whether one associates pitch change with height or tension, \textit{do-re-mi} fits (with the caveat that it does not imply specific frequency ratios in this context). A weakness of the pitch height concept is that it is used to describe many different pitch relationships and ranges. Within the field of linguistics, tone and non-Western languages are peripheral, outside of the mainstream. So, using height to describe small-scale pitch relationships in non-Western languages is unproblematic. But in the public sphere of a tone-language culture, where pitch plays so many roles in both language and music, it becomes much more problematic. In examination of four dictionaries, it is not clear whether \textit{ohùn-òkè} means a \textit{voice with a high range}, using the \textit{high part of one’s range} or a \textit{high tone} (which can be quite low within one’s range)—all very different meanings. Terms like \textit{ohùn-òkè} reflect an ongoing and important effort to develop a metalanguage for Yorùbá in Yorùbá (Adeeko 1992), an effort that is not always understood or appreciated by the general public. Using height to describe pitch may be part of the obstacle.

\(^{16}\) \textit{Dùndùn} players can often play diatonic (Westernized) melodies on their instruments, a rather astonishing feat because the drums are variable-pitch instruments not tuned to the diatonic scale.
My conjecture that adjusting tension is an historic and still vital conceptualization of changing pitch more familiar to Yorùbá speakers through the dundún talking drum still needs empirical verification. It arises out of consideration of a wide-variety of data, but does not reflect a line of inquiry I pursued during my fieldwork or that can be substantiated historically. The do-re-mi heuristic avoids the extrinsic analogy between pitch and height that does not resonante with all people, opting instead for the cross-cultural currency of the Tonic Sol-Fa movement. If one chooses to take the heuristic very literally and sing Yorùbá words with the tone levels fixed to do-re-mi, it works. If one uses the heuristic as intended by teachers, infusing the relative pitch relationships and contours of Tonic Sol-Fa singing into speech, it works even better.

It is not clear when the heuristic originated, but this is how I imagine it… Missionary churches were a meeting place of language and music, both Western and indigenous, but the latter was seen as a tool for evangelism, not forms of communication and art valuable in themselves. European missionaries were bewildered by the melodic speech of Yorùbá, much as stress language speakers struggle to learn tone languages now. In the late Nineteenth Century to early Twentieth Century, Yorùbá scholar-evangelists (in one or many locations synergistically) creolized their rich knowledge of two very different cultures into a notion that Europeans could understand. There is no evidence to suggest this originated with Crowther himself, but certainly among those who followed in his footsteps. I imagine a statement both confrontational and cathartic: ‘You know your do-re-mi that you’ve been evangelizing with? Well, that is how our language

17 Sol-fa was a staple of colonial education and it remains a pop culture phenomenon through The Sound of Music which is widely available in DVD stalls in street markets in Lagos.
is, and that is why your hymns don’t make sense!’ Rev. J. J. Ransome-Kuti wrote that no hymn tune can ‘express the words in a tonic language’ in the preface to the Yorùbá language Anglican hymnal (*Iwe Orin Mimo* 1923). Over time, the do-re-mi heuristic became a pedagogical device for Yorùbá instruction in parochial and public schools, particularly useful for learning to mark tones in written Yorùbá.¹⁸ *Low-mid-high* tone levels are acknowledged in secondary school, but *do-re-mi* is the mnemonic for learning and writing tone sequences. The standardization of the model in bilingual education reinforced colloquial use of *do-re-mi* in talking about tone. While the Yorùbá language is robust, it is competing in a multi-lingual environment of Hausa, Îgbò, Pidgin, Arabic, English, French and other languages. Language change is speeding up. I have often heard older speakers correct younger speakers, aghast they do not know the correct tones. As an òyìbó, I have often received patient coaching—even a little sensitivity to tone on my part is appreciated.

**Conclusion**

Crowther made one of the most important contributions to the field of linguistics. He made it clear that segmental phonemes are not enough to describe the lexicon of all languages. This very important observation questions the primacy of the segmental phoneme. Phonology is ostensibly the study of sounds, but is mostly the study of segmental phonemes. Methods for the analysis of other forms of linguistic contrast (within phonology) are not nearly as developed (Leben 2006). Before Crowther, African

---

¹⁸ Àmì ohùn mean tone marks, literally ìmì: mark and ohùn: voice (Bamgbose 1990).
languages were filtered through Western ears and minds, and conformed to Western ways of writing without any accommodation. In mandating that ‘elevations’ and ‘depressions’ of pitch be marked in the Yorùbá orthography, Crowther made a bold step forward. Unfortunately, his innovation of orthography was credited to Johann Gottlieb Christaller, who is credited with generating a more holistic and sensitive understanding of African languages (see Bearth 1998). The exclusion of Crowther in the narrative of linguistic theory is partially because of the dominance of Germany within the field at the time (Agwuele 2008), and because Crowther was an African. His fame within Anglophone evangelism lauded his work as a bishop, not as a linguist. Although Christaller’s 1875 Twi grammar does not cite Crowther’s 1852 Yorùbá orthography, the comparative study of related languages in the Introductory Notes draws comparisons to Yorùbá no less than seven times (1875:X-XXIV). While Crowther is celebrated within Yorùbá scholarship, he is not given due credit in the broader field of linguistics. I propose the Christaller method be redubbed the Crowther method. Since tone-marking has evolved, minor nuances between their work does not matter. History often celebrates who got there first. Hopefully we have progressed to the point where the fact that Crowther was an African does not prevent us from giving credit where credit is due.

Sarah Ann Glover’s modernization of Guido’s solfege was also misattributed in past scholarship, often credited to Curwen and Kodály. Several notable scholarly works (Rainbow 1967, Bennett 1984, McGuire 2009) have already made this correction. However, within American music education, the Tonic Sol-fa method is still considered a subset of the Kodály method. Continued reinforcement within scholarly literature is necessary to acknowledge Glover. Even more effort is needed (beyond this article) to
draw attention to Crowther’s contributions, not only to Yorùbá, but to the broader study of African languages.

Aside from suffering the consequences of intellectual and social dominance by white men in life and in death, there is a greater kinship between Glover’s and Crowther’s innovative approaches. Glover brought music education out of elite cathedrals and conservatories, adapting an antiquated method of solmization to new music and environments. Crowther developed an orthographic approach that accommodated other forms of linguistic contrast, instead of using the same cookie-cutter method. Though their physical paths never crossed, their ideas did, converging in an inter-continental and trans-disciplinary synthesis.

Four main points summarize this chapter:

1. The presence of speech surrogate instruments indicates language and music had a close relationship among Yorùbá-speaking cultures prior to the missionary era;

2. Western perspectives on language and music literacy were introduced simultaneously in many African cultures, key aspects include:
   a. The segmentation of sound using the Latin alphabet, and,
   b. The Tonic Sol-Fa method of solmization and notation;

3. Crowther’s innovation of marking tone in standard Yorùbá was later adapted by Christaller to Twi languages becoming a standard for nearly a century;

4. The *do-re-mi* heuristic creolizes indigenous knowledge with solmization into a pedagogical tool that is now widely used in secondary and tertiary education.

---

19 Crowther was in England in the 1850s. Glover was still alive, but to my knowledge they never crossed paths.
The *do-re-mi* heuristic may be seen as a gentle and democratic resistance to the pitch-height paradigm used in formal linguistics, i.e. *low-mid-high* tone levels. And, as Crowther taught us, we should maintain healthy skepticism about the primacy of segmental phonemes and their self-sufficiency.
3. THEORIES OF AFRICAN TONE SYSTEMS

Introduction

In his comprehensive study of *African Language Structures* (1973), Welmers describes the tone system of what he terms “discrete level tone languages.” In these languages, each syllable has at least one tone that is associated with a discrete tone level:

...each level tone is restricted to a relatively narrow range of absolute pitch (absolute for a given speaker under given environmental conditions) within a phrase, and these tonemic ranges are discrete—never over-lapping, and separated by pitch ranges which are not used—throughout the phrase, though they may all tilt downward at the very end of the phrase in a brief final contour. Thus, in a three-level system, high tone near the end of the phrase has virtually the same absolute pitch as a high tone at the beginning of the phrase, and is higher than any mid tone in the phrase... (Welmers 1973:81).

Despite the date, Welmers’ insight into tone systems remains fairly current, is still often cited, and deserves credit for taking positions on still open questions. In the encyclopedic *Music, Language and the Brain*, Patel calls Welmers’ account “provocative” but points out digital recording and fundamental frequency extraction was not available at the time so empirical verification is still needed (2008:44). In the coming chapters, new data will be presented that sheds light on key issues which Welmers raises. In this chapter, I explore open questions and still credible theories developed when interpreting data was predominantly aural.

Inherent in Welmers’ description is a paradigmatic model of tone, which was a growing trend in the 1970s. Though dormant for some years, the debate over using
paradigmatic or syntagmatic models of tone production and perception has been reinvigorated recently, notably in Laura Dilley’s dissertation (2005), as well as Leben (2006), Ladd (2008a) and Clements et al (2011). Dilley, for one, advocates for the integration of paradigmatic and syntagmatic representations.

3.1 Paradigmatic Tone Levels

There are several salient points in Welmers’ description, both credible and questionable, including: (1) tones are level; (2) levels are discrete; (3) levels are adaptable (to voice ranges); and (4) phrase-final lowering is present but does not compromise the discreteness of each level. Welmers’ use of the term level is twofold in my reading. First, the term level is an adjective to describe the shape of the pitch segment, which is characterized as flat not sloped. This differentiates the pitch trajectory of level tones from contour tones, an older taxonomy to differentiate types of tones and tone systems. In contour tone languages like Mandarin and Vietnamese, the pitch segment has a unique trajectory which leads to its classification, such as “rising” or “falling” tone. The second use of level is as a noun, “three-level system”, indicating multiple discrete pitch ranges. Because each tone has the same pitch trajectory (relatively flat), it is the placement of the pitch segment within one of the tone levels which gives it a categorization. Unless it is somehow neutralized, each pitch segment is associated with either high level, mid level or low level in a three-level system. For the rest of this dissertation I will favor the noun version of level, as in tone level, and eventually contour level. This does not imply that a tone-bearing segment has a flat trajectory but that it is associated with a referent tone.
level. Although it is not made explicit in Welmers’ writing, the association of tones to referent tone levels is a paradigmatic model.

Welmers provides an example in Jukun (Diyi) from eastern Nigeria to model the phonetic realization of tones in a three-level system.

<table>
<thead>
<tr>
<th>Jukun</th>
<th>Tone Level Sequence</th>
<th>English Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>Áku péré ni zé budyi à syi ni bi kéré.</td>
<td>HMLLHLMLMLMHMH</td>
<td>That person brought this food here.</td>
</tr>
</tbody>
</table>

Table 3.1: Example sentence in Jukun from Welmers (1973:81)

In his visualization, each tone is represented by a dash placed on one of three horizontal planes representing the levels.

[Figure 3.1]

Welmers notes phrase-final lowering occurs, as is common in many languages. It is not reflected in his illustration, perhaps because he deems the effect negligible. Figure 3.2 (below) is my own interpretation of Welmers’ description. The use of dots and lines is taken from musical contour theory (e.g. Kolinski 1965; Marvin and Laprade 1987). Here, each dot represents a tone (pitch segment) and is connected by a line. As in contour drawings of musical melodies, the relative duration of each pitch segment is not represented, each is a single dot, evenly-spaced. Solid lines connect lexical tones (belonging to a single word) and dashed lines connect inter-lexical tones (between words). The shading of the dots is changed at the beginning of each new word. The three
levels are indicated by color coding (not present in contour graphs): blue for H (high),
green for M (mid), and yellow for L (low). Warping of the levels on the right side
indicates phrase-final lowering, but true to Welmers’ account, the last two High tones do
not breach the Mid level.

In connection with the discussion from Chapter 2, Welmers also draws a musical
analogy:

The three levels are discrete... and so precisely limited that playing them on three
notes on a piano (a major triad does very well) does not appreciably distort the
pitches of normal speech. (Welmers 1973:81)

This is similar to the heuristic for Yorùbá tone levels among bi-lingual speakers
describing L, M and H using the tonic solfà syllables, Do, Re, Mi. However, Welmers’
metaphor suggests tone levels are farther apart in terms of logarithmic frequency: three to

---

20 This color coding will be used consistently for three-level systems.
four semitones, as in a diatonic third, instead of two semitones. The intervallic distances between the levels are not intended to be specific, but do-mi-so spacing is actually more realistic than do-re-mi. The problematic part of Welmers’ description is the insistence that each tone has a flat trajectory and that the pitch height of the levels that each tone associates to is discrete and largely invariant. Aside from the negligible effect of phrase-final lowering, in Welmers’ model, tones at the same level are limited to a narrow frequency-banded range, much like notes in musical scales. As in Figure 3.3, the first high tone in an utterance is at G4, then each subsequent high tone will be at G4, give or take a few cents, until the phrase-final lowering. After breathing, the tone levels would presumably reset with the high level centered on G4. The specific pitches in Figure 3.3, C4, E4 and G4, are not so central to Welmers’ argument as the discreteness of the levels: the pitch of each level is absolute. The key to a paradigmatic model of lexical tone is invariant pitch-height of the levels. The voice range is divided up into dedicated frequency bands that act as target frequencies for level tones.

3.2 Tone Level Normalization

Welmers describes levels as absolute for a given speaker, indicating the necessity for tone levels to adapt to each speaker. Normalization of contoured tones in Asian languages was the subject of phonetic studies by Earle (1975) and Rose (1987), but data was not presented for normalization of tone levels in African languages until the second edition of D. Robert Ladd’s Intonational Phonology (2008a). Using data supplied by Bruce Connell, Ladd calculates mean $f_0$ values for five speakers of Mambila, a four-level tone
language in eastern Nigeria. Based on Earle, Rose and the Mambila data, Ladd generalizes a theory of “speaker-specific normalized scales,” in which tone levels adapt to each speaker’s overall level (relative height) and span (the range of the speaker’s habitual frequency) (2008:197). Ladd is careful to note that scales of tone-levels are nothing like musical scales (2008:196).

Patel adapts Ladd's findings to a visualization of hypothetical data for a four-level system. The mid-low tone level is 20% above the low tone level, the mid-high tone level is 50% above the low tone level, and the high level is at the upper extent of the range. For example, Speaker A’s range is from 100 Hz to 200 Hz so mid-low is at 120 Hz and mid-high is at 150 Hz. In the analysis of the actual Mambila data, the spacing is closer to 30% for low-mid (2008a:196). Patel also expresses Ladd’s model as a formula:

\[ T = 100 \times \frac{(F-L)}{R} \]

Figure 3.4: Illustrations of range-based proportional scaling from Patel (2008:45).
F, L and R are measured in Hz as presented in Ladd and Patel, but could potentially be measured in logarithmic pitch (semitones) or ERB (equivalent rectangular bandwidth).\textsuperscript{21} F is the frequency of the interior tone level (\textit{low-mid} or \textit{high-mid} in a four level system), L is the frequency of the low tone level, and R is the range (the frequency of the high tone level minus L, the low tone level). With these, one can calculate T, which is the proportional height of the interior tone level within the tone range (from H to L). It is T that Ladd found to be consistent between speakers of Mambila.

The dissimilarity between normalized tone spaces and musical scales is made apparent by comparing the scaling of tone levels for Speaker A and Speaker C in Figure 3.5. The frequency ratio (interval) between H and L for Speaker A is 2:1 (200 Hz to 100 Hz). The ratio for Speaker C is 7:5 (140 Hz to 100 Hz). If tone level space were based on constant frequency ratios (intervals) like many musical scales are, Speaker A and Speaker C would share the same tone level space because their L is at the same frequency (100 Hz). Instead, the tone levels are normalized by setting L to 0 (accomplished with “F–L” in the equation above).

\textbf{3.3 On the Musical Analogy (again)}

The temptation to draw musical analogies between tone language speech and musical melody is irresistible. There is no better way to get the ball rolling in terms of conceptualization. But almost as soon as the light bulb starts to flicker on, you have to pull the plug. Some have argued for a musical proto-language (Mithen 2005) or that

\textsuperscript{21} These interpretations of fundamental frequency are not equivalent, and despite problems with log pitch, it is adopted for much of this dissertation.
auditory mechanisms for pitch perception evolved to serve language first (Deutsch, Henthorn and Dolson 2004). Such theories may some day explain the missing link between music and language, but the motivation for the musical analogy is usually to demystify a fuzzy concept with a concrete one. The fundamentals of music theory have been developed over two millennia. Scales and intervals are physically measurable and mathematically formalized. Welmers’ description is almost entirely in the image of music. First, tones are level: they have stable frequency for the duration of the tone-bearing segment like musical notes. Second, tone levels are discrete: they are invariant pitch referents much like C4, E4 and G4. Third, levels are adaptable to different speakers: this makes solmization fit better than absolute pitches, but is still similar to phenomena in music: simple melodies can be sung by any voice. Finally, phrase-final declination is present but does not compromise the discreteness of each level. This can be read as: tones may go slightly flat and still be at the same tone level.

Ladd richens the musical analogy by calling tone systems “speaker-specific normalized scales” but almost immediately refutes the comparison: “one important fact emerges… linguistic equivalence of pitch between speakers with different ranges is not based on anything like a musical scale” (2008a:196). Why refer to it as a scale at all? There must be some semblance to motivate the analogy.

### 3.4 The Paradigmatic versus Syntagmatic Debate

In the first chapter, the *toneme* was introduced as a contrastive unit of lexical tone, but the nature of the contrast was left open. Welmers’ *African Language Structures* (1973) was
at the cusp of a shift in the phonological understanding of tone contrast, from a
syntagmatic to a paradigmatic model. Growing documentation of African tone systems,
particularly those of Nigerian languages, woke up phonologists to new ways of thinking
about tone.

The classical theorization of tone is as a syntagmatic feature (see Fant, Halle and
approach comes from Dilley: a syntagmatic tone interval relates two sequentially-ordered
tones (2005:2). Ladd (2008a) calls this an initializing system because an utterance
consisting of a single pitch segment or multiple segments with approximately the same
frequency have no syntagmatic significance.

In the 1960s, generative phonology largely focused on paradigmatic features
belonging to segmental phonemes (e.g. Chomsky and Halle’s Sound Patterns of English
1968) and was less concerned with suprasegmentals (tone, stress and duration) that were
deemed as “little more than the residue of segmental transcription” (Ladd 2014:64). In
the 1970s the study of tone was invigorated by autosegmental theory, which relies on
paradigmatic tone intervals to relate each tone to a “speaker-specific referent level”
(Dilley 2005:2).22 As noted in the discussion of Welmers, the referent is usually thought
to be a frequency band within a speaker’s range. Because different people have different
ranges the tone levels normalize to speaker range. In this model, each syllable has a
atoneme based on where it falls within the frequency-banded range.

---

22 I will refer to tones at a paradigmatic tone level instead of paradigmatic tone intervals.
In 1973, Will Leben (an Africanist with field experience in Nigeria) completed his dissertation on *Suprasegmental Phonology* (1973) laid the groundwork for a new approach to studying tone. The shift came full force in 1976 with John Goldsmith’s dissertation *Autosegmental Phonology*, which Leben calls the last major milestone in the representation of tone (2006:1). The tone systems of African languages were both a model and motivation for autosegmental phonology and the Ìgbò language of southeastern Nigeria is the focus of application for an entire chapter. Less formal descriptive accounts (such as Welmers’) were already consistent with a paradigmatic view, so autosegmental phonology fit with trends in documentation by Africanist scholars. The greater change came in the analysis of contoured tones (as found in Mandarin). These were reinterpreted as sequences of level tones, e.g. *rising* tone as low-high.

Two publications from the last ten years, Laura Dilley’s dissertation *The Phonetics and Phonology of Tonal Systems* and D. Robert Ladd’s second edition of *Intonational Phonology*, provide exegeses of the paradigmatic versus syntagmatic debate within tonology. They come to different conclusions. Dilley advocates for the integration
of paradigmatic and syntagmatic representations using a flexible *tone interval*. Ladd maintains a preference for a paradigmatic model, exemplified by autosegmental phonology and largely consistent with Welmers’ description of *tone levels* at the beginning of this chapter.

### 3.5 Problems with the Syntagmatic Model

Dilley’s definition of a syntagmatic tone interval (relating two sequentially-ordered tones) is very close to the concept of a melodic interval within music theory. A melodic interval is a vector in pitch space, having both direction and magnitude. Dilley provides a formula with the variables $I$ for interval and $T$ for tone.

$$I_{1,2} = \frac{T_2}{T_1}$$

The subscripts indicate sequential ordering and are not restricted to adjacent tones, but true to the syntagmatic milieu of localized relationships, the two tones should be proximal. If $I$ is greater than 1, the tone in the numerator position is greater than the referent tone in the denominator position and the direction of the vector from $T_1$ to $T_2$ is an ascent in pitch. If $I$ is less than 1, $T_1$ to $T_2$ is a descent in pitch.

#### 3.5.1 Initialization of Syntagmatic Tone

We can alternatively formalize the directional component of the syntagm alone by using a pre-existing framework from music theory called the Contour Adjacency Series (CAS) (Friedmann 1985). A CAS is an ordered series of adjacent contour comparisons, + for higher, – for lower or 0 for equivalent pitch height. Though developed for music, the
CAS describes a basic feature of pitch relationships in any mode, whether speech, music or other sound. The CAS is a very simple syntagmatic method and it clearly demonstrates what Ladd calls an initializing approach. Figure 3.6 brings back the phrase in Jukun presented by Welmers and applies the CAS to the tone level sequence, e.g. HM yields a descent (–). Note the tone level sequence has 14 tones but the CAS has 13 contour values. For a melody or tone sequence with a cardinality of $n$ notes or tones there are only $n-1$ adjacent pairs.

<table>
<thead>
<tr>
<th>Jukun</th>
<th>Tone Level Sequence</th>
<th>Contour Adjacency Series</th>
</tr>
</thead>
<tbody>
<tr>
<td>Áku pèrè ní zè budyi à syi ní bi kéré.</td>
<td>HMLLHLMMLMHMH</td>
<td>$&lt; - - 0 + - 0 - + - + - 0 &gt;$</td>
</tr>
</tbody>
</table>

Figure 3.6

Ladd’s objection to a syntagmatic (or relational) model is that it requires initialization (2008:190). This is manifest in the CAS in Figure 3.6 which only has 13 adjacencies for 14 tones. A single tone utterance has an empty CAS. This means that a single syllable cannot have an intelligible toneme. Furthermore, a string of pitch segments at the same tone level are unspecified in terms of height. Under a syntagmatic model, there is no difference between HH, MM and LL. According to Ladd, tone level languages such as Yorùbá pose the greatest challenge to a syntagmatic approach (2008:191). An initializing approach leaves a single tone unspecified and the tone level of a single tone is understood. Indeed, Yorùbá dictionaries by Crowther (1852), Abraham (1962) and others, include monosyllable words differentiated by tone level (e.g. á, a à). There are also homophone pairs formed by disyllables in which the syntagm (contour) between syllables has no change (e.g. /igba/ MM and LL).
Finally, an initializing model is unappealing on formal grounds because it does not offer a one-to-one correspondence between the number of tone-bearing segments and the categorization of them. A one-to-one mapping (n inputs to n outputs) is preferable.

### 3.5.2 The Need for Magnitude

Another challenge posed by a three-level system like Yorùbá to a syntagmatic model is disyllable homophone pairs with the same direction between syllables (e.g. /igba/ MH and LH). This is not advanced by Ladd but is implied by his discussion of normalization of paradigmatic tone levels. Roughly 40% of Yorùbá and 60% of Ìgbò disyllable words form minimal pairs distinguished only by tone variation (Carter-Cohn 2013). Under a syntagmatic representation, the toneme for each disyllable consists of one contour comparison. Simple directional relationships (+,0,–) can account for the majority of tone-contrasted minimal pairs, but not all.

<table>
<thead>
<tr>
<th>Word</th>
<th>mimó</th>
<th>mimo</th>
<th>mimò</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gloss</td>
<td>holy, clean</td>
<td>built with mud</td>
<td>known</td>
</tr>
<tr>
<td>Tone</td>
<td>HH</td>
<td>HM</td>
<td>HL</td>
</tr>
<tr>
<td>Contour</td>
<td>0</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

Table 3.3
In the /mimò/ homophone group, mimò (HH) is distinct with static tone level, but the HM and HL tone patterns reduce to the same directional relationship.

As Welmers notes, tone levels have to adapt to different speaking ranges and his illustration of tone spacing (the major triad) deviates from the Yorùbá folk heuristic (do-re-mi) presented in the previous chapter. Because of the variability in tone-level spacing (further explored in Chapters 4 and 5), it is doubtful that interval magnitude between tone levels remains constant (e.g. 2 semitones as in Do-Re-Mi or even generic thirds like the major triad). Intervals in Western music theory are based on superparticular ratios (n+1:n) of frequency, e.g. 3:2 is a perfect fifth (P5). It is unreasonable to expect tone levels to conform to mathematical ratios, yet a syntagmatic model of tone that accounts for the difference between HM and HL requires accounting for interval magnitude. While precision on the order of semitones is not possible, it may be possible to refine the contour comparison, by using non-ternary contour. Huron’s *Themefinder* algorithm has a search parameter that aggregates directional categories (+,0,–) with broader categories of magnitude, the *step* and the *leap*. A step is a movement to an adjacent pitch in a scale and a leap is a movement to a non-adjacent pitch. The aggregated categories of refined contour are: leap up, step up, same, step down, and leap down (Huron et al 1999). These contour categories can be aptly applied to a tone language with three levels: low, mid and high. A movement from L to H is a level change of +2, a movement from M to H or L to M is a level change of +1, and so on. However, these categories are not sufficient for languages with more than three levels. Moving from the lowest to highest level in a four-level language is +3 levels.
Refined contour categories are not conventional musical intervals but include some indication of magnitude of change. When applied to music, the magnitude component (step versus leap) requires some contextual variation depending on the musical scale to which it is applied, but otherwise forms discrete categories. To accommodate speaker range variability, Clough’s concept of the generic step interval (1979) could be useful. Just like music listeners habituate to different scales, tone language speakers might habituate to different speakers, determining each speaker’s step interval magnitude and thereby distinguishing steps (e.g. MH and LM) versus leaps (e.g. LH). Such a model would not distinguish between MH and LM and HM and ML, nor does it explain how a single tone is understood. Furthermore, applying refined contour to tone languages suggests some equivalence between musical scales and tone-level systems, and as Ladd notes, this is problematic (2008:196).

To differentiate same-direction pairs under a syntagmatic model, some reliable indicator of magnitude of change, in addition to direction, must be included in the signal. Because of variability in speaking range, the distance between paradigmatic tone levels is not based on absolute frequency ratios. There is little certainty about the difference in magnitude between a syntagmatic tone interval that shifts the pitch one level and shifts the pitch two levels, just like there is uncertainty about the relative height and span of paradigmatic tone levels. If the Mambila data covered in Section 3.2 is extensible to other
languages, a movement from 100 Hz to 140 Hz could indicate a change of one level or an entire span. Along with initialization, magnitude of change makes a strictly syntagmatic model problematic.

### 3.6 Problems with the Paradigmatic Model

Dilley does not avoid paradigmatic or syntagmatic models, but acknowledges faults in both that lead her to an integrated solution: a *tone interval* that can be alternately paradigmatic (associated to a speaker-referent) or syntagmatic (relative to an adjacent tone). According to Dilley, English and similar languages challenge paradigmatic analysis because of significant variability in the frequency of phonologically-equivalent tones within a single speaker’s range (2005:11). Paralinguistic intonation raises and lowers speaking range moment-to-moment as an affective cue. Welmers description of tone levels in Jukun as “precisely limited” implies that they are not modified by paralinguistic intonation. Although he acknowledges the presence of late-phrase declination, he maintains that *high* tones at the end do not enter the frequency range of *mid* tones at the beginning. Diana Deutsch is another scholar that has considered contrastive tone and paralinguistic tone to be mutually exclusive in a language.

#### 3.6.1 Tone and Intonation

Deutsch, Henthorn and Dolson (2004) report tone language speakers (specifically Mandarin and Vietnamese) are able to speak the same word at the same pitch on different occasions. The same task is difficult for speakers of “intonation” languages (e.g.
English). From this, they draw the conclusion that learning a tone language during the critical period for language acquisition triggers absolute pitch perception. Ladd responded directly to Deutsch’s speculation about the role of absolute pitch in tone languages, calling the hypothesis unnecessary. Phonetic studies show tone languages “manage phonological equivalence in exactly the same way” as English (Ladd 2008b:7). Ladd has consistently argued against the mutual exclusion of tone and intonation. Paralinguistic modifications of pitch range within speaker are in addition to between-speaker variation illustrated by the normalization model (Section 3.2). Interactions between tone level normalization and paralinguistic intonation are not completely understood (Ladd 2008a:35–36). Although Welmers describes minimal declination, he did not acknowledge other paralinguistic intonation, and likely was not privy to it. This may mean languages Welmers studied (1) did not have paralinguistic intonation, (2) interlocutors he worked with did not use it when demonstrating speech, or (3) paralinguistic modifications are present but subtle enough to be imperceptible. Of these, the second and the third possibilities are the most plausible. When a speaker is demonstrating intelligible speech to a non-native speaker for teaching or documentation purposes, they are likely to emphasize linguistic contrasts and minimize paralinguistic affect. This is entirely appropriate. The last possibility is not only plausible but has the seeds of an explanatory account of the coexistence of tone and intonation in a language. Paralinguistic intonation may alter tone spacing, but it may be gradual or constrained to certain positions so as not to obscure the identity of the tone levels, like an unaccompanied choir that goes flat by a semitone over the course of several minutes. For the paradigmatic model, it is not essential the tone levels are invariant physically (in
terms of frequency) so much as the tone levels are invariant perceptually. However, if tone levels vary and are not strictly frequency banded within a speaker’s range, it is unclear how phonological equivalence is managed.

### 3.7 Prototypical Paradigmatic and Non-Prototypical Languages

Dilley makes a distinction between prototypical paradigmatic languages, such as Cantonese and Yorùbá, and non-prototypical languages such as English (2005:9). There are parallels here to Deutsch’s distinction between intonation and tone languages as well as an inclination toward absolute pitch paradigmatics. Dilley (2005:9–10) suggests frequency banding as the phonetic framework for prototypical paradigmatic languages citing Connell and Ladd (1990) as well as Laniran and Clements (2003), both of which will be elaborated. Dilley stresses the need for paradigmatic and syntagmatic features, but she does not believe they apply uniformly to all languages. In her view, some languages freely use intonation, such as English, and are better suited to syntagmatic representation. Other languages, exemplified by Yorùbá and Cantonese, are prototypically paradigmatic.

Dilley’s paradigmatic and non-paradigmatic taxonomy parallels Deutsch’s (less fancy) tone and intonation language dichotomy to some extent, but not fully. The difference is that there are languages such as Ìgbò that have a strong presence of lexical tone (60% of disyllables form minimal pairs) but Dilley classifies as non-prototypical. For Dilley, the presence of intonation in a language supports the use of syntagmatic features, whether or not lexical tone plays a major role in the language. She does not acknowledge, or perhaps

---

23 This taxonomy is post-autosegmental because the older distinction between level tone (e.g. Yorùbá) and contoured tone (e.g. Cantonese) languages present in Welmers (1973) has evaporated. With Autosegmental theory, contoured tones were reinterpreted as sequences of pitch targets (e.g. falling would be high-low).
is not aware of, the presence of intonation in Yorùbá, which is one of her cardinal examples of a prototypically paradigmatic language exhibiting frequency banding. Conversely, Ladd indicates Yorùbá does have paralinguistic intonation and rejects absolute pitch referents as a basis for tone levels (contrafactum Deutsch). Yet, he earlier supported frequency banding in Yorùbá (Connell and Ladd 1990) and still espouses a paradigmatic approach.

3.7.1 TONE TERRACING

A distinction in Niger-Congo tone languages that Welmers and others make is between level tone languages (such as Jukun, Mambila, and Yorùbá) and terracing languages, exemplified by Ìgbò. Clark (1978, 1990) offers a syntagmatic phonology for Ìgbò that allows for considerable shift in the absolute pitch height of high tones, even within an utterance. This is characteristics of tone-terracing languages in which tone levels are downstepped (lowered) repeatedly within an utterance, before resetting for a new utterance. Connell and Ladd (1990:4) describe pitch realization in Yorùbá as exhibiting frequency bands that are “more or less fixed throughout an utterance”. They only speculate as to whether tone levels remain discrete and non-overlapping in the presence of downstep. Laniran and Clements (2003) address Yorùbá specifically and do mention frequency banding as a utility to protect the integrity of tone levels (specifically the mid level). By and large however, they attest to the variability of the high tone level, which exhibits both downstep and raising without compromising the mid level (true to Welmers’ description of Jukun). Downstep is another challenge to static frequency bands, demonstrating the need for some gradual shifting or widening of Yorùbá’s high tone level at least. By empirically demonstrating the significant presence of downstep in Yorùbá, the distinction between level tone and terracing tone is weakened. The behavior of the high tone level, which is highly variable in the analysis by Laniran and Clements, mirrors
accounts of *high* tone level in Ìgbò terracing. Likewise, their description of the *mid* level in Yorùbá as fixed (with a discrete frequency band) resembles the *low* tone level in Ìgbò. Instead of having tone systems with different behavior (level tones and terraced tones), Yorùbá and Ìgbò may only differ in the number of tone levels: three versus two respectively. In both, the high level is flexible and a non-high level (*mid* in Yorùbá and *low* in Ìgbò) exhibits frequency banding within small phonetic time spans. Dilley does not list Ìgbò among prototypical paradigmatic languages. On the contrary, Dilley uses it as an example of a tone language in which syntagmatic features work, in accordance with Clark (1978, 1990) but with the caveat that paradigmatic features are also needed (Dilley 2005:15). If the distinction between tone levels in terracing languages like Ìgbò and ostensibly non-terracing languages Yorùbá is moot (a position long ago asserted by J. M. Stewart 1971), then a syntagmatic and paradigmatic integration should also be applicable to Yorùbá.

### 3.8 Open Questions

The coming chapters explore Ladd’s co-presence of intonation and tone and Dilley’s acquiescence of paradigmatic and syntagmatic features. Chapter 4 examines the effect of intonation on tone levels in Yorùbá. Chapter 5 presents evidence for the perceptual identification of lexical tones based on syntagmatic features for both Yorùbá and Ìgbò, with no evidence that the perception of tonemes is fundamentally different (i.e. paradigmatic or syntagmatic) in the two languages. Ultimately, my position is more pro-syntagmatic than Dilley’s because I support the application of syntagmatic tone intervals to Yorùbá, a language she classifies as prototypically paradigmatic. I do not dispute the existence of paradigmatic tone levels in such languages, however I question the extent to which frequency bands describe them. The relationship between Ladd’s perspective and my perspective on tone languages is peculiar. They are sympathetic in that lexical tone
and intonation are not mutually exclusive. He has seen the evidence I present, but maintains a preference for the autosegmental method. The intractable objection for Ladd and others to a syntagmatic model is the initializing problem: a single tone-bearing segment lacks the context necessary to form a syntagmatic tone interval (unless divided up into multiple tones as in a contour tone in Mandarin). This is a valid problem for which there is no syntagmatic explanation. So, if that is what you are looking for, there is no reason to read any further. I think it is very plausible that a combination of pitch and non-pitch features help to distinguish single-tone homophones as in Table 3.5.

<table>
<thead>
<tr>
<th>High</th>
<th>Gloss</th>
<th>Mid</th>
<th>Gloss</th>
<th>Low</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>bà</td>
<td>to overtake</td>
<td>ba</td>
<td>to lie in ambush</td>
<td>bà</td>
<td>to bespeak</td>
</tr>
<tr>
<td>lè</td>
<td>to appear</td>
<td>le</td>
<td>to be strong</td>
<td>lè</td>
<td>to be able</td>
</tr>
<tr>
<td>yè</td>
<td>to make much of</td>
<td>yè</td>
<td>to be fit</td>
<td>yè</td>
<td>to be out of place</td>
</tr>
<tr>
<td>kí</td>
<td>to salute</td>
<td>kí</td>
<td>to be thick</td>
<td>kí</td>
<td>to press</td>
</tr>
<tr>
<td>mọ</td>
<td>to be clean</td>
<td>mọ</td>
<td>to build</td>
<td>mọ</td>
<td>to know</td>
</tr>
<tr>
<td>kú</td>
<td>to die</td>
<td>kú</td>
<td>to come short</td>
<td>kú</td>
<td>to blow into dust</td>
</tr>
</tbody>
</table>

Table 3.5: monosyllable homophones from Crowther (1852:4)

However, as Crowther notes, most of these words occur with a pronoun at the mid-tone level (which may or may not be modified due to tone interaction). Crowther’s own description of the acute (´) and grave (´) accents that he introduced to the orthography of African languages is that they indicate “elevation” and “depression” of tone, respectively. These terms are inherently syntagmatic, unlike categories of “high” and “low”. There are indeed monosyllables with contrastive tone, and if pressed speakers can differentiate them using slope within the tone-bearing segment or knowing something about a speaker’s range and registral characteristics. However, as Crowther establishes at the very outset of Yorùbá language studies, monosyllables work in concert. After spending two years in Yorùbá-speaking areas, I cannot recall many (or any) instances of utterances consisting of a monosyllable homophone with direct lexical meaning. By and large, short utterances consist of expressive ideophones with very active pitch activity on a single
vowel. Neither the tone or phone are lexically objective, it’s more of an existential cue in conversation that lets a speaker know you are paying attention. The single tone problem raised by phonologists is a red herring because utterances with a single tone-bearing unit (a sonorant of some sort) rarely carry lexical meaning that relies on contrastive tone. If anything, such utterances are an avenue for studying paralinguistic intonation in the absence of lexical tone, and not vice-versa. The issue of tone sequences that are ambiguous because they do not use all tone levels (e.g. LL, MM and HH or LM and MH) is addressed in Chapter 11. First, I address the extent to which tone and intonation can coexist (Chapter 4) and whether there is perceptual evidence for the prototypical and non-prototypical distinction (Chapter 5).
4. Paralinguistic Intonation in Yorùbá

Introduction

Prosody is the “music of everyday speech,” paralinguistic cues enhance meaning through variation in pitch, timing, and stress (Wennerstrom 2001). If linguistic signals are carried on the same dimension of sound as paralinguistic cues, there is potential for one of these to be compromised, even contradicted. In English, speech becomes less intelligible when an agogic stress is placed on an unstressed syllable. Likewise, in Yorùbá, placing a high pitch emphasis on a low tone could make an utterance ambiguous or even contradictory (Rowlands 1969). Paralinguistic modification, however, does not normally obscure phonological contrasts in such a way (Ladd 2008a:3).

If tone levels of a three-level system are assigned to the pitches of a major triad and do not deviate from them (as in Welmers’ illustration at the opening of Chapter 3), lexical tone is very clearly articulated. This reflects frequency banding idealized by the paradigmatic model. Speech that strictly adheres to frequency bands is clear but also may sound contrived and unnatural and, as we shall see, invariant frequency banding does not reflect all fluent and intelligible Yorùbá speech. So, the question is to what extent tone levels can vary, as a result of affective (paralinguistic) modification, without obscuring word identification.

Few documentation projects and no perceptual studies (to my knowledge) have sought to determine the extent to which paralinguistic intonation may modify contrastive lexical tone. As discussed in Chapter 3, there are descriptions and theoretical models implying lexical tone and paralinguistic intonation are mutually exclusive. For example,
Deutsch, Henthorn and Dolson have proposed that comprehension of tone language speech relies on absolute pitch perception (2004:350). They offer phonetic evidence, not perceptual evidence, to support a position about speech perception, a strong fallacy. Ladd calls this hypothesis “unnecessary” because phonetic studies of tone languages demonstrate that phonological equivalence in tone languages is the same as it is in English, and not based on absolute pitch (2008b:7).

Local variation in pitch of phonologically equivalent tones in African languages has motivated complex rules for downstep, including for Yorùbá (see Laniran and Clements 2003). Downstep is distinct from late phrase declination in that it only affects the pitch height of a single tone level. I venture the full extent of tone level and intonation interaction and composite variation possible in tone languages is still unknown. Phonetic studies addressing fluctuation of tone levels analyze short excerpts of speech for the most part, from a single to a few utterances, not large corpora of natural speech (as has been done for English). Recordings for documentation and phonetic studies are often neutral in terms of emotional valence, either because of explicit instruction or simply the nature of working with a professional informant or study participant. Researchers do not (and perhaps should not) want to make the people on which research depends angry, or otherwise attempt to alter emotional states.

There is no phonological or physiological explanation for the rich within-speaker variation of tone-level spacing seen in this chapter. It is not a by-product of an accumulation of tone realization effects (such as down-step and high-raising) or variation in sub-glottal air pressure (like phrase declination). It reflects clear and deliberate affective gesture: prosody. Ladd acknowledges that linguistic and paralinguistic messages
may be carried on the same signal, even in Yorùbá, but he does not speculate how this is accomplished or provide much evidence for the phenomena:

… paralinguistic modifications of pitch range—such as raising the voice in anger—modify the realisation of lexical tones in languages like Yorùbá or Chinese, but do not normally obscure their identity. This is true even in languages in which the tone phonemes are distinctive levels (such as High, Mid, and Low in Yorùbá): in some way that we do not entirely understand, the phonological essence of the levels must be invariant relative to a phonetic frame of reference that can be modified for paralinguistic purposes. (Ladd 2008a:35–36)

A challenging theoretical problem in Ladd’s book is that he strongly advocates for a paradigmatic formalization (because of the single tone problem), but also acknowledges that tone and intonation can be present in the same signal, which poses a problem for paradigmatic model (as Dilley 2005 argues). The clearest explanation of paradigmatic tone levels is frequency banding. It is hard to imagine how frequency bands can be consistent with tone levels modified by paralinguistic intonation. Others have made paradigmatic tone and intonation mutually exclusive. The languages Ladd mentions, Yorùbá and Chinese (presumably Mandarin), are the same languages Dilley classifies as prototypical paradigmatic tone languages, languages she associates with frequency banding. Mandarin is also one of the languages Deutsch uses to advance her absolute pitch hypothesis. If Ladd is right, then Deutsch’s distinction between tone and intonation languages and Dilley’s distinction between prototypical paradigmatic and non-prototypical tone languages evaporate.

24 A syntagmatic model offers no explanation of how a single segment can have a toneme (see Section 3.5.1).
4.1 Methodological Disclaimer

Despite many descriptive accounts of West African languages, the phonetic realization of tone is still not well understood. This is even true of Yorùbá, a well-documented tone language spoken in southwestern Nigeria and the Benin Republic. The aim of this chapter is to establish the presence of both tone contrast (linguistic signals) and intonation (affective paralinguistic cues) in Yorùbá. While I regard the findings of this chapter empirical, it does not reflect the most rigorous methodology, as in being strictly a priori or using an experiment to test a hypothesis that reflects my own theory about tone and intonation. I do not have the means to implement (and have trouble conceiving) a study that could correlate changes in the emotional valence of a speaker (e.g. going from low arousal to high arousal) with changes in intonation. One experimental possibility would be to ask participants to read the same passage with different emotions, “as if” angry or “as if” sad, etc. If a speaker, read a passage then had some physical exertion (e.g. jogging on a treadmill) this might show a correlation between heartrate and prosodic cues. However, exposing a participant to a stimulus that is somehow upsetting or inflammatory, though it might get induce physiological or emotional changes, is problematic from a research ethics standpoint. I do not have the means to do any of this, neither in terms of facilities and resources or the wherewithal to propose such an experiment for research review.

A greater problem with applying behavioral science methods here is that, in all the scenarios mentioned above, a speaker would be reading. I cannot substantiate it in tone languages, but in English, intonation differs when one is reading aloud versus when one is talking (Esser and Polomski 1987). Certainly, among voice professionals this
distinction disappears, there are many people, such as actors, that are able to recite text as if it were spontaneous. To employ an actor to read the same text with different emotions would be a possibility, but would also have the strong possibility of demand characteristics, as if the experimenter were a director and the actor responding to what the director wants. Some verification others interpreted the emotion as intended, which could in fact be the experimental task, would be needed. So there are further, and perhaps more rigorous, possibilities for determining what factors correlate with or even cause changes in intonation and clearly distinguishing tone and intonation in the same signal.

The approach here is to analyze speech as if I were analyzing music. At the outset at least, I am concerned with large-scale structure, as one might observe in a symphonic movement. The analysis objects are creative, not purely functional. These recorded sermons and speeches have persuasive power that in large part relies on artistry: meaning it, or at least saying it like you mean it. In addition to analysis of large-scale intonation, a number of post-hoc analyses are applied making the label of exploratory study appropriate. Indeed, my thinking on tone levels has changed significantly in the process, but the hypothesis I am testing has not: I am still looking for any evidence of stable absolute frequency bands, but I have yet to find any. Due to negative results during initial attempts to find frequency bands in my data, the hypothesis that frequency bands frame tone realization shifted from the positive to the null hypothesis. I did not resist this shift because it was not a hypothesis I supported, it is simply the status quo. There is substantial evidence to disregard the hypothesis that there are speaker-specific frequency bands. Phonologically equivalent tones may occur at the same pitch, they also may not. What this means for a paradigmatic model is reflected upon in future chapters. With the
knowledge that tone levels are not restricted to frequency bands, it is all the more possible to explore alternate hypotheses, based on syntagmatic and quasi-paradigmatic features (see Chapters 5–11).

Affective intonation is not something that has to be, but can be. If someone becomes more engaged and energetic, angry or excited, intonation is likely to change. At the end of this chapter, I posit a theoretical model of this, but I cannot “prove” it. I can merely state there is no evidence to suggest emotional valence cannot have an effect on range and span within speaker. Astonishing intonational variation can be found in languages with essential lexical tone contrasts. Case in point: Yorùbá sermons from All Saints Anglican Church in Yaba, Lagos, Nigeria.

4.2 Sermons from All Saints Anglican Church - Yaba

In the rest of this chapter, results are presented for a macro-analysis of Yorùbá sermons. Recorded sermons are a rich resource for the study of intonation in enhanced (but intelligible) speech. Sermons do not use everyday speech, but expressive speech with a high degree of affectation. In religious communities the world over, religious leaders often build their reputation on strengths in oratory. Speech can be beautiful or compelling in any language, and the speaker’s command of prosody is a major factor.

The sermons are from Sunday morning services at All Saints Anglican Church-Yaba in Lagos, Nigeria. All recordings were made in 2012 to 2013 and are from the mixing console of the church's sound system. The priests were speaking into a wireless dynamic microphone. The church keeps a hard-drive with recordings of services. My
colleague David Aina of Lagos State University is the music director at All Saints.

Through him, I was able to secure permission to copy and study these sermons. In all, six Yorùbá sermons are analyzed. These come from four different priests: Rev. Adediji, Rev. Adekunle, Rev. Adesanya, and Rev. Owadayo. English sermons were also collected for three of these speakers.

Transcriptions (including tone marks) and translations were completed for the majority of the sermons. This verifies the sermons are intelligible Yorùbá speech. Sermons, as enhanced public speaking or recitation, may have a different degree of intonational variation than less formal speech. However, these recordings represent articulate and intelligible speech that does not distort the tonology of Yorùbá. It is presumed that lexical tone in the sermons is neither more or less important than, nor the realization different than, tones in other Yorùbá speech.

The initial research question was: how are tones spaced within Yorùbá? Is it as indicated by my friends in Lagos (i.e. do-re-mi)? Is it like Welmers describes (i.e. do-mi-so)? Or, is it like Ladd's range-based proportional scaling (i.e. big range = wide spacing, small range = close spacing)? When I first loaded a recorded sermon into Melodyne, the question changed significantly to: How dynamic is speaking range in Yorùbá?
Figure 4.1 is a screen-shot of the Melodyne workspace with a sermon by Owadayo loaded, about 10 minutes in length. Melodyne automatically segments audio into note-like blobs. The segmentation is quite effective for monophonic signals, elaborated below. Frequency banding did not characterize this recording, instead an intonational arch could be observed spanning the extent of the recording. Tones at the beginning and end occupied a different range than tones in the middle. I did not take this as evidence large-scale intonational arches are a common phenomena (though they proved to be). I did, however, see this as evidence against frequency banding. Fortunately, I had more sermons by Owadayo and his colleagues that served as reserve data to confirm this is not an isolated phenomenon.
4.3 Segmentation and Feature Extraction

Many challenges to the documentation and phonetic study of African languages in the past are no longer an obstacle. The technology to record speech and accurately and efficiently analyze it is now widely available and even portable. This has been the case for some time, but audio signal processing has improved greatly in the last decade. Connell and Ladd (1990) and Laniran and Clements (2003) both present analyses based on fundamental frequency extractions of Yorùbá speech from recordings of immigrant speakers in the UK and US respectively. Both acknowledge some gentle shifting of tone level during the course of an utterance, but not enough to question the paradigmatic model. The amount of data presented here is much greater, with hours, not seconds or minutes of speech. This reflects advances in technology and processing power that were not available one to two decades ago. It is now possible to process hours of speech in minutes. In particular, a rather miraculous software (also very secretive in its workings) is Melodyne by the Celemony corporation, released in 2008. Melodyne is designed to correct poor intonation in singing, but is increasingly being used in analysis and research (e.g. Wild and Schubert 2008, recent analysis projects by Jay Rahn). In Chapter 5, it is used to produce experimental stimuli.

4.3.1 Melodyne’s Segmentation

An aspect of my method that will no doubt be questionable to some is occasional reliance on Melodyne’s automatic note segmentation feature. When audio is loaded into the software, this segmentation is performed automatically, chopping up a continuous
signal into pitch-time segments that can be moved around in the Graphic User Interface. Throughout this dissertation, when I am closely analyzing a text and recording, the segmentation is done manually (as described in Chapter 1). The vast majority of what is presented in this chapter, is not. Is Melodyne’s segmentation more or less accurate than a human? It is certainly more consistent than a human. The algorithm is not public knowledge, but it is clear that each audio segment Melodyne produces has a vowel nucleus and pitch change does not trigger the formation of a new segment. For instance, a sharply sloped pitch trajectory may fall within a single note segment. On the other hand, an abrupt pitch change as found in bel-canto singing may trigger a new note segment, most likely because of a decrease and increase in periodicity and amplitude. Such pitch discretization is not found in speech, not even in tone level languages. The “level” tones of Niger-Congo languages are not actually flat pitch trajectories.

One issue is whether Melodyne’s automatic segmentation is equally equipped for Yorùbá, which has syllables after the fashion of consonant-vowel, and English (also examined) which has diphthongs. If anything, Yorùbá, with consistent CV structure, poses less of a challenge to an automated pitch segmentation than English. I have spent considerable time comparing Melodyne segmentations to Yorùbá transcriptions and found them to be quite accurate. As for the English data, there are existing studies of inter-vocalic duration and variation and through comparison to published findings, I can verify the segmentation is reasonable and consistent, if not acutely accurate.

Melodyne enables correction of the segmentation through clicking and dragging and so on. This was not done for the large sermon and speech files (roughly 10-30 minutes in length, ranging from 1500 to 4000 pitch segments). After the segmentation,
Melodyne applies a pitch reduction, assigning each pitch segment to the logarithmic pitch scale (C4, G2, etc) based on mean frequency (this has been verified by independently calculating mean f0 for each pitch segment). Like other algorithms for fundamental frequency extraction (e.g. the auto-correlation algorithm YIN by de Cheveigne and Kawahara 2002), Melodyne’s method produces sub-harmonic octave errors (observable in Figure 4.1). There is a correction tool for octave errors but this also was not used. In the interest of quick data preparation, these were highlighted and deleted to avoid the same signal producing octave errors in MATLAB. The crucial interpretation of the data provided by Melodyne was the onset and offset values for each audio segment. After importing these values into MATLAB, the YIN algorithm was used for f0 extraction to have continuous data for the recording (to calculate slope for the pitch segment, and so on). I cannot verify it, but I have a suspicion that YIN and Melodyne may use the same auto-correlation method, and at the very least would react similarly to the same signal, i.e. produce octave errors on the same pitch segments. Thus, pitch segments that produced octave errors in Melodyne were deleted before exporting a MIDI file that served as a list of onsets and offsets for MATLAB implementations. It is slightly messy data, but there is also a lot of it. The comparisons made are between data prepared the same way, not data prepared in different ways.

4.3.2 Feature Extraction in MATLAB

Before getting into a large-scale view of the data, it is important to understand the individual units on which the analysis is based: different features at the pitch segment
(smallest) order. Each column in Table 4.1 represents a pitch segment corresponding to a V or CV syllable and each row one of four features as well as pitch difference. There are three groups of five rows, one for each of three sequential iterations of the same phrase shown in the caption for Figure 4.2. This is a recording excerpt from one of two Yorùbá sermons by Peter Owadayo. The figure shows the f0 analysis for each repetition of the phrase (the black line), the mean frequency and time midpoint of the pitch segment (the red circles) and dotted lines connecting all High tones to each other (blue), all Mid tones to each other (green) and all Low tones to each other (yellow). The three plots in Figure 4.2 correspond to three shaded blocks of Table 4.1. The values for pitch (in semitones) and duration (in seconds) are in familiar units, but the amplitude value (normalized power, corresponding to stress) is not. This is a mean (across the syllable/pitch segment) of the period-smoothed instantaneous power returned by the YIN algorithm. It is period-smoothed and instantaneous because a value is returned for each sample of the audio signal (in this case 44.1 kHz), but it does not reflect within period fluctuations because it is period-smoothed. The means for all pitch segments are normalized to a value of 1 for the maximum (instead of the very small values returned in the power parameter of YIN). The peak power for the entire recording excerpt is found in the very first pitch segment of the model (first version) of the phrase. Amplitude compression may have been applied at various points between the microphone receiving the signal and the recording to the computer hard drive.

There is a value for each feature for each pitch segment, but these values are not particularly useful in isolation. Vowels correspond to absolute referents (formant structure), but tone, timing and stress do not have cross-speaker referents.
Figure 4.2 Three consecutive repetitions of the same phrase in a Yorùbá sermon: “L’Órúkọ Jésù ẹ ó rẹ́rín ayò” (In the name of Jesus you will laugh with joy).

<table>
<thead>
<tr>
<th>Text</th>
<th>L’ó-</th>
<th>rú-</th>
<th>kọ</th>
<th>Jé-</th>
<th>sù</th>
<th>ẹ́</th>
<th>ó</th>
<th>rẹ́</th>
<th>rín</th>
<th>a-</th>
<th>yò</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speech Tone</td>
<td>High</td>
<td>High</td>
<td>Mid</td>
<td>High</td>
<td>Low</td>
<td>Mid</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>Mid</td>
<td>Low</td>
</tr>
<tr>
<td>Direction of Change</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Mean Pitch (ST)</td>
<td>58.2</td>
<td>58.4</td>
<td>56.7</td>
<td>58.7</td>
<td>48.0</td>
<td>54.6</td>
<td>54.6</td>
<td>51.8</td>
<td>50.3</td>
<td>49.0</td>
<td>45.1</td>
</tr>
<tr>
<td>Pitch Difference (ST)</td>
<td>0.1</td>
<td>-1.7</td>
<td>2.0</td>
<td>-10.6</td>
<td>6.5</td>
<td>0.0</td>
<td>-2.8</td>
<td>-1.5</td>
<td>-1.3</td>
<td>-3.9</td>
<td></td>
</tr>
<tr>
<td>Slope (ST/sec)</td>
<td>1.4</td>
<td>-4.7</td>
<td>1.1</td>
<td>-7.3</td>
<td>-13.2</td>
<td>16.6</td>
<td>-7.0</td>
<td>0.0</td>
<td>-2.1</td>
<td>-1.6</td>
<td>-8.4</td>
</tr>
<tr>
<td>Duration (sec)</td>
<td>0.349</td>
<td>0.234</td>
<td>0.570</td>
<td>0.900</td>
<td>0.245</td>
<td>0.176</td>
<td>0.266</td>
<td>0.266</td>
<td>0.269</td>
<td>0.590</td>
<td>0.513</td>
</tr>
<tr>
<td>Power (Normalized)</td>
<td>0.79</td>
<td>0.13</td>
<td>0.35</td>
<td>0.39</td>
<td>0.16</td>
<td>0.47</td>
<td>0.77</td>
<td>0.63</td>
<td>0.28</td>
<td>0.29</td>
<td>0.18</td>
</tr>
<tr>
<td>Mean Pitch (ST)</td>
<td>58.7</td>
<td>58.4</td>
<td>56.5</td>
<td>59.6</td>
<td>53.5</td>
<td>56.4</td>
<td>56.9</td>
<td>55.6</td>
<td>54.6</td>
<td>53.9</td>
<td>48.2</td>
</tr>
<tr>
<td>Pitch Difference (ST)</td>
<td>0.3</td>
<td>-1.9</td>
<td>3.1</td>
<td>-6.2</td>
<td>2.9</td>
<td>0.5</td>
<td>-1.3</td>
<td>-1.0</td>
<td>-0.7</td>
<td>-5.6</td>
<td></td>
</tr>
<tr>
<td>Slope (ST/sec)</td>
<td>1.0</td>
<td>-0.1</td>
<td>-2.1</td>
<td>-1.1</td>
<td>-5.1</td>
<td>21.0</td>
<td>-3.2</td>
<td>-2.2</td>
<td>-0.8</td>
<td>-0.8</td>
<td>-11.3</td>
</tr>
<tr>
<td>Duration (sec)</td>
<td>0.283</td>
<td>0.185</td>
<td>0.422</td>
<td>0.663</td>
<td>0.279</td>
<td>0.193</td>
<td>0.422</td>
<td>0.313</td>
<td>0.243</td>
<td>0.518</td>
<td>0.615</td>
</tr>
<tr>
<td>Power (Normalized)</td>
<td>0.79</td>
<td>0.13</td>
<td>0.35</td>
<td>0.39</td>
<td>0.16</td>
<td>0.47</td>
<td>0.77</td>
<td>0.63</td>
<td>0.28</td>
<td>0.29</td>
<td>0.18</td>
</tr>
<tr>
<td>Mean Pitch (ST)</td>
<td>60.6</td>
<td>60.5</td>
<td>58.9</td>
<td>61.8</td>
<td>56.3</td>
<td>58.6</td>
<td>59.0</td>
<td>58.2</td>
<td>57.4</td>
<td>57</td>
<td>52.4</td>
</tr>
<tr>
<td>Pitch Difference (ST)</td>
<td>-0.1</td>
<td>-1.6</td>
<td>2.9</td>
<td>-5.4</td>
<td>2.2</td>
<td>0.4</td>
<td>-0.7</td>
<td>-0.8</td>
<td>-0.3</td>
<td>-4.7</td>
<td></td>
</tr>
<tr>
<td>Slope (ST/sec)</td>
<td>1.1</td>
<td>-1.0</td>
<td>-1.3</td>
<td>1.0</td>
<td>-7.6</td>
<td>7.1</td>
<td>-2.2</td>
<td>-1.0</td>
<td>-1.7</td>
<td>0.4</td>
<td>-12.1</td>
</tr>
<tr>
<td>Duration (sec)</td>
<td>0.333</td>
<td>0.195</td>
<td>0.534</td>
<td>0.385</td>
<td>0.396</td>
<td>0.100</td>
<td>0.309</td>
<td>0.400</td>
<td>0.214</td>
<td>0.385</td>
<td>0.871</td>
</tr>
<tr>
<td>Power (Normalized)</td>
<td>0.83</td>
<td>0.08</td>
<td>0.30</td>
<td>0.51</td>
<td>0.23</td>
<td>0.41</td>
<td>0.79</td>
<td>0.35</td>
<td>0.18</td>
<td>0.30</td>
<td>0.17</td>
</tr>
</tbody>
</table>

Table 4.1: Measurements for each syllable segment of three repetitions of the same 11-syllable phrase

25Is pronounced like é (H) in all cases, however both University of Ìbádàn and Abraham's dictionary have ẹ (M) as you, it, etc. Abraham has ẹ́ as pronunciation of letter ẹ in reciting the alphabet.
It is unclear which events stand out in the signal until we compare the values in each dimension within the phrase and to analogous points across phrases. From this, one might draw a conclusion that there are stable referents for this speaker, or not. Whether one subscribes to a paradigmatic or syntagmatic view, this initial stage of orienting a framework to a new speaker is inherently relational, in communication and analysis.

Table 4.1 is organized by recurrence of the syllables in three lexically identical sentences. This alignment through the columns makes patterns apparent. For instance, in all three versions, the first syllable has the highest mean power in the phrase, and the second has the lowest. The fourth syllable is longest in the first and second versions, but not the third where both the syllable before and the last syllable are longer. Presumably, since Yorùbá is a tone level language, there should be some commonalities in the pitch structure. Comparisons within and across phrases indicate the first High tone (58.2) does not represent a constant referent within the phrase (the last High tone is 50.3) or across phrases. Instead, the first H in the third phrase is 60.6 and the first mid is 58.9, both higher than the first H in the first phrase. The within phrase variation is consistent with late-phrase declination. This is posited by Welmers (1973) but there is more declination than he indicates, to the extent that high tone is lower than mid tone. In this excerpt, the levels are not discrete. Laniran and Clements (2003) also propose that high tone may vary but that mid tones are frequency banded. The between phrase variation is something different which has not been described before. This phenomenon can more clearly be

26 The Direction of Change row indicates whether the speech tone of a successive syllable is higher (+), lower (-) or the same (0) as the tone of a prior syllable. The second table section is measurements of data from a segmented recording of the first phrase spoken. The Duration row is length of each syllable segment in seconds. The Mean Frequency is calculated using the YIN algorithm (de Cheveigne 2002). The Segment Slope is calculated using a Discrete Cosine Transform after Devaney et al. (2011). The third and fourth sections are equivalent data for the “igbá” (M-H) variant.
observed in Figure 4.3 which includes all three phrases (as they are in the recording) in one plot.

Figure 4.3 is a contour plot for the same three-phrase excerpt of a Yorùbá sermon by Peter Owadayo. Similar to plots in Figure 4.2, the y-axis is log pitch and the x-axis time. The dotted blue line connects the pitch approximations for the high tone level, the green line connects mid tones and the yellow line connects low tones. If there are frequency bands present, they are not static within the phrase. The pitch resets with each new utterance, but tone levels are not necessarily assigned to the same absolute pitch height. Each new repetition has a compressed span in comparison to the prior utterance. In the last, the range is also raised considerably. One possible explanation is the phonemic and tonemic information is established in the first utterance and so does not have to be so precise in the repetition. In this view, repetition of an utterance could be used as a modulatory device for speech intonation, just as transposition of sequential material can be used for modulation in music. I think this is a strong possibility, and as demonstrated in the final analysis section of this dissertation, there is more flexibility in the musical
setting of repeated text than initial presentations of text. In those musical settings, contour (the pattern of ups of downs) is not preserved, but in Owadayo’s three-phrase repetition, contour is keenly preserved. Furthermore, this variation in range and span of tone levels from one phrase to the next is not isolated to repeated utterances.

4.4 Visualizing Intonation

If speaker-specific absolute pitch referents (or frequency bands) are a part of the phonetic framework of a language, I would expect mean pitch for any small part of a recording by the same speaker to be similar to mean pitch in another part of the recording, indicating that their speaking range has not gone up or down. Likewise, if there are stable frequency bands, I would expect standard deviation from mean pitch to be consistent, indicating span has not changed. Laniran and Clements (2003) have suggested that High tone level varies with an utterance but mid does not. Even with local variation in High tone, and the additional effect of late-phrase declination, one would still expect mean pitch and standard deviation for windows much larger than an utterance (e.g. 100 syllables) to be fairly invariant. Such a result would be consistent with the prototypical paradigmatic classification of Yorùbá, and Deutsch’s hypothesis about absolute pitch in tone languages. On the other hand, if the effect seen in Figure 4.3 is not isolated, frequency bands (if present) are at the very least variable in terms of range and span. In this analysis, mean pitch and standard deviation are considered as indicative of the register a speaker is actively using within a window of the recording. The extent to which tone and intonation can be differentiated in the signal by the measures presented is discussed.
4.4.1 MEAN FREQUENCY PLOTS OF ALL SAINTS SERMONS

Figure 4.4 shows the trend in mean pitch over the course of three sermons by Peter Owadayo of All Saints Anglican in Yaba, Lagos. The “L’oruko Jesu” excerpt in Figure 4.3 is from “Owadayo Yorùbá 2”. The y-axis value is mean pitch for windowed samples of 100 consecutive events, considerably larger than the 33-segment “L’oruko Jesu” excerpt. The x-axis is the chronological order of the windows. A hop size of 20 segments was used. The average total length of the three sermons is 1735 segments. In “Yorùbá 2”, Owadayo starts low with a mean pitch around Bb2 and reaches a mean pitch of around middle C (C4=60). “Owadayo Yorùbá 1” shows a similar trend, a large-scale arch shape over approximately 10 minutes. Notably, his English sermon leaves off with a high pitch centroid while his Yorùbá sermons both have a gentler denouement.
utilized to the same extent in the English sermon. Segment standard deviation (calculated for the pitch values for all segments, not windows) is 4.59 semitones and the maximum standard deviation for any single window is 4.89 semitones. This is within the range of values for the Yorùbá sermons, 4.35 and 4.66 semitones for segment standard deviation.

Two sermons by Rev. Adediji, one Yorùbá and one English, are totally disparate in mean pitch trends at first glance. This false impression is somewhat alleviated by noting, in Table 4.2, that the English sermon is 4025 segments, while the Yorùbá sermon is 1881 segments. For the other speakers, time-warping helps to observe corresponding arch shapes and so on, but in this case it is a bit misleading.\textsuperscript{27} Not completely however. The windows are the same size, so it is striking that Adediji reaches higher pitch centers than ever reached in the Yorùbá sermon at multiple points. The highest standard deviation for any single 100-segment window in the Yorùbá sermon is 4.99 semitones while the highest for the English sermon is 8.36 semitones. There is an explanation for the

\textsuperscript{27} One way to combat this visually misleading aspect of the time-warp is to adjust window sizes based on the number of segments for the recording. Use the same window size (100 segments) regardless of length was chosen as the better option.
difference in length: the 8:30 AM Yorùbá service is budgeted at 1 hour, but the 10 AM English service follows no such schedule, often going well into the afternoon.

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Language</th>
<th>Duration</th>
<th>Segments</th>
<th>Min</th>
<th>Max</th>
<th>Segment $\mu$</th>
<th>Segment $\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adediji</td>
<td>English</td>
<td>25'37&quot;</td>
<td>4025</td>
<td>37</td>
<td>70</td>
<td>54.180</td>
<td>6.895</td>
</tr>
<tr>
<td>Adesanya</td>
<td>English</td>
<td>8'21&quot;</td>
<td>1536</td>
<td>46</td>
<td>70</td>
<td>56.056</td>
<td>4.599</td>
</tr>
<tr>
<td>Owadayo</td>
<td>English</td>
<td>13'31&quot;</td>
<td>1757</td>
<td>42</td>
<td>68</td>
<td>58.852</td>
<td>5.101</td>
</tr>
<tr>
<td>Adediji</td>
<td>Yorùbá</td>
<td>10'27&quot;</td>
<td>1881</td>
<td>36</td>
<td>66</td>
<td>53.449</td>
<td>5.386</td>
</tr>
<tr>
<td>Adekunle</td>
<td>Yorùbá</td>
<td>12'40&quot;</td>
<td>2557</td>
<td>40</td>
<td>66</td>
<td>56.528</td>
<td>4.531</td>
</tr>
<tr>
<td>Adesanya 1</td>
<td>Yorùbá</td>
<td>6'18&quot;</td>
<td>1181</td>
<td>44</td>
<td>66</td>
<td>56.626</td>
<td>4.349</td>
</tr>
<tr>
<td>Adesanya 2</td>
<td>Yorùbá</td>
<td>8'26&quot;</td>
<td>1730</td>
<td>47</td>
<td>69</td>
<td>57.093</td>
<td>4.656</td>
</tr>
<tr>
<td>Owadayo 1</td>
<td>Yorùbá</td>
<td>10'26&quot;</td>
<td>1754</td>
<td>40</td>
<td>65</td>
<td>55.043</td>
<td>5.692</td>
</tr>
<tr>
<td>Owadayo 2</td>
<td>Yorùbá</td>
<td>9'25&quot;</td>
<td>1694</td>
<td>40</td>
<td>65</td>
<td>55.952</td>
<td>5.680</td>
</tr>
</tbody>
</table>

Table 4.2

Without exception, windows at the beginning of the sermons had very low mean pitch. They usually begin with a prayer. The highest mean frequencies were typically found in the second and third quarter (between .25 and .75 on the plots). Whether the sermons finished on a high centroid or returned to a near-initial state was not consistent for all sermons by the same speaker, nor for all English sermons or all Yorùbá sermons. If the mean went back down, the sermon also ended with a prayer, similar to the beginning.
Figure 4.7 includes one Yorùbá sermon by each of four priests. In comparison, Adesanya’s mean pitch range is high and narrow. Adediji’s mean pitch range is lower than the others, but he can go higher as demonstrated by his English sermon. Adekunle (not pictured before) and Owadayo have very similar extents. The variation in mean pitch is quite striking. All but Adesanya have a minimum to maximum mean pitch difference greater than an octave (+12 semitones). This means the initial mean frequency doubles at the peak mean frequency.

4.4.2 COMPARISON TO AMERICAN AND BRITISH ENGLISH

Dilley (2005) takes a more liberal view of paradigmatic and syntagmatic tone models, but she still acknowledges a class of tone languages she calls prototypical paradigmatic languages. These include Yorùbá and Mandarin. The notion is these languages strictly adhere to absolute referents specific to each speaker and overall range variation is constrained. However, there are non-prototypical languages (with or without lexical tone) with more flexibility for intonation. In English, paralinguistic intonation is richly varied.
While the last section made within speaker comparisons of Nigerian English and Yorùbá, a comparison to American and British English speakers is also illustrative. I found no indication that bilingual speakers in Nigeria constrain pitch in one language and not the other, but do these broad trends differ from analogous speech (such as sermons) in the English of non-Yorùbá speakers?

The first cross-dialect comparison is to sermons by the immediate-past and current Archbishops of Canterbury, Rowan Williams and Justin Welby. Williams’ sermon is from the 2012 General Synod. Welby’s is his inaugural sermon from 2013.

Figure 4.8 includes both sermons. The scaling of the y-axis values is the same as previous plots, but the extent is less (44 to 54 semitones). To characterize Williams’ and Welby’s intonation as flat is accurate. In each, mean pitch varies by less than two semitones—perhaps we should apply frequency banding! Between the two, Welby is a high talker and Williams a low talker. Comparing Anglican sermons in Nigeria and the UK would seem to be appropriate for gauging intonation use. However, there are some key differences in Anglicanism in the UK and Africa. Among other things, the latter is much more lively in terms of sound. The sermons by the Archbishops are also read, which may account for the relatively static mean pitch windows.
Instead of relying on this single comparison to dismiss the null hypothesis that intonation is unconstrained in English and constrained in Yorùbá, I went further. Nigerian Anglican sermons may have more within-speaker diachronic variation in speaking range than British Anglican sermons, but surely cannot match the degree of prosodic excitation in American politics. To test this, I selected two iconic speeches of the twentieth century: John F. Kennedy’s inaugural address from 1961 and Martin Luther King’s “I have a dream” speech from 1963.\(^{28}\) I also extracted audio from videos of two more recent speeches: Donald Trump’s 2014 speech for the American Conservative Union and President Barack Obama’s 2016 speech advocating for more stringent gun control.

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Year</th>
<th>Duration</th>
<th>Segments</th>
<th>Min</th>
<th>Max</th>
<th>Segment $\mu$</th>
<th>Segment $\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kennedy</td>
<td>1961</td>
<td>13'52””</td>
<td>1721</td>
<td>45</td>
<td>66</td>
<td>56.888</td>
<td>2.941</td>
</tr>
<tr>
<td>King</td>
<td>1963</td>
<td>15'59””</td>
<td>1955</td>
<td>49</td>
<td>70</td>
<td>60.373</td>
<td>2.453</td>
</tr>
<tr>
<td>Williams</td>
<td>2012</td>
<td>14'37””</td>
<td>1035</td>
<td>37</td>
<td>98</td>
<td>45.184</td>
<td>3.141</td>
</tr>
<tr>
<td>Welby</td>
<td>2013</td>
<td>11'06””</td>
<td>1509</td>
<td>42</td>
<td>59</td>
<td>52.999</td>
<td>3.009</td>
</tr>
<tr>
<td>Trump</td>
<td>2014</td>
<td>18'18””</td>
<td>2924</td>
<td>42</td>
<td>69</td>
<td>55.289</td>
<td>4.557</td>
</tr>
<tr>
<td>Obama</td>
<td>2016</td>
<td>37'59””</td>
<td>3837</td>
<td>39</td>
<td>64</td>
<td>50.452</td>
<td>4.660</td>
</tr>
</tbody>
</table>

Table 4.3

In addition to removing apparent octave errors, clapping was removed from the speeches. Table 4.3 provides general summaries of the American speeches and British sermons. Speeches by Trump and Obama are longer and the standard deviation for segment pitch (calculated from the entire pitch series) is higher for these two speakers. As can be observed in the plot of British sermons, Williams has a very low voice.

\(^{28}\) I do not consider King a politician, but this speech had a political agenda: motivating the passage of the Civil Rights Act.
The American speeches do not exhibit the arch common in the Nigerian sermons. They do not start and end with prayers. The speeches exhibit much more intonational variation than the British sermons. All four end with a windowed mean pitch that is five or more semitones higher than they started. The clearest pattern is found in MLK, who articulates a long, nearly uninterrupted, ascent. JFK, Trump and Obama also reach peak mean pitch in the last quarter of the plot. Obama’s gun-control speech includes several arches. It is considerably longer than the others. The sharp rise to the highest mean pitch in Obama’s speech (just after time tick 0.8 in the plot) corresponds to a tearful recounting of the Sandy Hook mass shooting in 2012, at about 29 minutes in the speech which can be found on YouTube. Some have speculated Obama’s tears were theatrical and not evidence of authentic grief. I cannot evaluate Obama’s internal emotions, but there is an immediate transformation of his voice at this moment, most clearly seen in the change in pitch centroid. (There may also be changes in overall amplitude, but these are mostly lost to amplitude compression). The peak mean pitch in Obama’s speech comes at a plaintive moment, which is comparatively subdued. This use of high intonation differs from Kennedy, MLK, Trump and the Nigerian priests, all of whom pair high and loud
consistently in a confident, strident and intense voice. Obama also does this, but when he uses this strong voice, it does not reach the same pitch height as the plaintive voice. I cannot quantify the trends in amplitude because of the nature of the amplitude-compressed audio I am working with. This is a qualitative assessment based on listening. Qualitatively, perhaps due to its subject of gun violence, Obama’s speech is fundamentally different. Both King’s and Obama’s speeches are goal-directed calls to action, but intonation is used in a greater variety of ways in the latter.

4.4.3 STANDARD DEVIATION WITHIN SEGMENT WINDOWS

The intonational architectonics of these recordings can be seen clearly in the plots of windowed mean pitch. They indicate shifts in pitch center are not constrained in Yorùbá any more than in English. However, mean pitch tells us nothing about another important aspect of speech register: span. In addition to pitch center, extent is important to tone level dispersion (see Section 3.2). Minimum and maximum pitch for the recordings (in Table 4.2 and Table 4.3) do not reflect the span in which most segments fall, these may be outliers (perhaps even octave errors that were missed in the preparation). Standard deviation is a better indicator of normative pitch span (50% of pitch segments within the sample fall within two times the standard deviation). Calculating mean pitch for windows of the pitch series revealed mean pitch (for 100 segment windows) varies within speaker by as much as an octave within a single speech or sermon. Thus, the standard deviation for all pitch segments within a recording (also in Table 4.2 and Table 4.3) is not representative of the span with an utterance.
One question I had was whether height and span could be correlated. Figure 4.10 indicates that this is not the case in Owadayo’s second Yorùbá sermon \( r=0.174 \). The mean of Pearson’s \( r \) correlation coefficients between mean pitch and standard deviation across all Yorùbá sermons was \( r=0.184 \). For American and British speeches and sermons, it was \( r=0.086 \). This reflects weak positive and negative correlations for each individual recording. No strong correlations \( (r > 0.6, \text{ using Pearson’s } r) \) were found between mean pitch and standard deviation with the exception of Adesanya’s English sermon \( r=0.871 \).

Figure 4.10 shows a more typical trend that is also quite logical: when mean pitch is changing rapidly the standard deviation is greater. Table 4.4 includes values for mean and standard deviation for four windows within Owadayo’s second sermon, the windows with the lowest mean pitch, the highest mean pitch, the lowest standard deviation and the highest standard deviation. The window with maximum standard deviation \( (6.15) \) is found during a sharp decline in mean pitch, right before the end.
The window with the highest mean pitch (60.6) and lowest standard deviation (2.28) of segment pitch from the mean is the same. This was found two-thirds of the way through the sermon, at index 0.67. Peak pitch centers near time values 0.25, 0.35, 0.6 and 0.9 all have very low standard deviation, around 2.5 semitones. Standard deviation at the lowest mean pitch (at index .04, near the beginning) is also quite low (2.85). Another logical generalization about the relationship between height and span is: when one is speaking at an extreme in one’s range (especially in the upper register), span is smaller.

### 4.5 Descriptive Statistics for Intonational Variability

A challenge in this analysis is finding quantitative representations of what can be observed in the plots as well as patterns in the data that are not apparent in the 100-segment windowing. There are no established methods for characterizing intonation trends and variability (Patel 2005:61). Instead of using maximum and minimum pitch to characterize range, I have suggested that windowed mean pitch and standard deviation is a good indicator of height and span of the speech register. Together, height and span describe the vocal register in use. It occupies a portion of the total voice range of the individual. In general, total voice range is not so variable, though temperature, humidity and other physical factors effect it. One extension of approximating the active register for tone languages is that the neutral (or mid) tone level is likely to be found around a mean.
or median pitch. More research needs to be done to assess whether windowed mean pitch and standard deviation from that pitch can be used to approximate the location of tone levels in a speaker’s range. However, several of the data treatments demonstrated in the next few pages show promise in the description of register height and span as well as interval contour and magnitude variability.

4.5.1 HEIGHT AND SPAN

I observed in the visualization of Owadayo’s second Yorùbá sermon (Figure 4.10) that pitch center changed a lot but standard deviation (relating to span) remained fairly constant, whether he was speaking in the bottom octave or upper octave of his range. Greater spans were found when occupying the middle range, or perhaps more accurately, moving from low to high or high to low range.29 This characterization of the plot is consistent with summative treatments of the data, specifically the mean of the windowed standard deviation and the standard deviation of the windowed standard deviation. Table 4.5 includes these two measures as well as the standard deviation of windowed mean pitch for window sizes of 100 and 10, with hop sizes of 20 and 2 respectively. Consistent with the visualization in Figure 4.10, the values for 100-segment windows for Owadayo (averaged for both Yorùbá sermons) include a mean standard deviation of approximately 3 semitones and very little variance from that across all windows (.398). This indicates that span is fairly consistent. Similar values were found for the other Yorùbá sermons with 100-segment windows. For 10-segment windows, the mean spans are a little smaller

29 An issue not addressed here is register changes when going to the upper or lower part of one’s range, what bel-canto singers call the “passagio” or break.
(µ (σ)) and there is slightly more variation (σ (σ) > .5) in span size. Individual Yorùbá sermons varied in their architecture, but by and large, similar descriptive summaries of height and span are found for all of the priests studied.

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Window = 100; Hop = 20</th>
<th>Window = 10; Hop = 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>µ (σ)</td>
<td>σ (σ)</td>
</tr>
<tr>
<td>Adediji</td>
<td>3.937</td>
<td>0.414</td>
</tr>
<tr>
<td>Adekunle</td>
<td>2.921</td>
<td>0.387</td>
</tr>
<tr>
<td>Adesanya</td>
<td>3.429</td>
<td>0.400</td>
</tr>
<tr>
<td>Owadayo</td>
<td>3.175</td>
<td>0.393</td>
</tr>
<tr>
<td>µ</td>
<td>3.365</td>
<td>0.398</td>
</tr>
</tbody>
</table>

Table 4.5

The similarity within language is not sustained across languages. The summaries are quite different for Nigerian English sermons by three of the same priests (Table 4.6). The contrast is impressive for Rev. Adediji who uses a greater span when speaking in English. This divergence is also reflected in the mean pitch plots for his two sermons (Figure 4.6). Adesanya and Owadayo also use larger spans in English, and have slightly more variation in span for 100-segment windows (σ (σ) > .5). However, Adesanya and Owadayo have less variation in mean pitch when preaching in English.

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Window = 100; Hop = 20</th>
<th>Window = 10; Hop = 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>µ (σ)</td>
<td>σ (σ)</td>
</tr>
<tr>
<td>Adediji</td>
<td>5.712</td>
<td>1.272</td>
</tr>
<tr>
<td>Adesanya</td>
<td>4.031</td>
<td>0.705</td>
</tr>
<tr>
<td>Owadayo</td>
<td>4.274</td>
<td>0.549</td>
</tr>
<tr>
<td>µ</td>
<td>4.672</td>
<td>0.842</td>
</tr>
</tbody>
</table>

Table 4.6

The within-speaker variation between Table 4.5 and Table 4.6 is evidence for cross-language variation. Because it is observed within-speaker, this change can be attributed to
differences between Yorùbá and Nigerian English and not between-speaker variation.

Across the board, the priests have a mean standard deviation at least 1 semitone greater for English than Yorùbá. One possible explanation is a greater degree of freedom in intonation at the local level (phrase to phrase) in Nigerian English.

Between-speaker cross-dialect comparisons within English reveal a bigger difference diachronically than geographically. Trump and Obama in Table 4.7 and Adesanya and Owadayo in Table 4.6 have remarkably similar values. Kennedy and King have much narrower spans than Trump and Obama.

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Window = 100; Hop = 20</th>
<th>Window = 10; Hop = 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>μ (σ)</td>
<td>σ (σ)</td>
</tr>
<tr>
<td>Kennedy</td>
<td>2.565</td>
<td>0.425</td>
</tr>
<tr>
<td>King</td>
<td>1.832</td>
<td>0.436</td>
</tr>
<tr>
<td>Trump</td>
<td>4.016</td>
<td>0.649</td>
</tr>
<tr>
<td>Obama</td>
<td>4.054</td>
<td>0.532</td>
</tr>
<tr>
<td>μ</td>
<td>3.117</td>
<td>0.511</td>
</tr>
</tbody>
</table>

Table 4.7

The variation in mean pitch (σ (µ)) is markedly lower for both American speeches and British sermons. The Archbishops, Williams and Welby (in Table 4.8), both stick close to the same mean pitch (σ (µ) of .36 and .46) as seen in the windowed mean pitch plot. In addition, Welby distinguishes himself from all the other recordings analyzed by maintaining the same span throughout (σ (σ) = .308). The earlier speculation that frequency banding may describe high and low levels in the British Anglican sermon

---

30 Levels are also used to describe English intonation in Autosegmental theory, though they are not lexically contrastive.
holds true for Welby. Of course these high and low levels do not constitute lexical contrast.

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Window = 100; Hop = 20</th>
<th>Window = 10; Hop = 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>μ (σ)</td>
<td>σ (σ)</td>
</tr>
<tr>
<td>Williams</td>
<td>3.020</td>
<td>0.992</td>
</tr>
<tr>
<td>Welby</td>
<td>2.956</td>
<td>0.308</td>
</tr>
<tr>
<td>µ</td>
<td>2.988</td>
<td>0.650</td>
</tr>
</tbody>
</table>

Table 4.8

In all of the recordings, span (approximated with standard deviation) remained consistent from window to window. Nigerian English sermons had the largest spans (μ (σ)) and Yorùbá sermons had the greatest variation in mean pitch (σ (µ)).

4.5.2 CONTOUR

One theory of lexical tone contrast is that it is syntagmatic, based on relationships between adjacent tone-bearing segments (see Chapter 3). Figure 4.3 offered an instance of the same tone sequence adapting to changes in height and span within speaker. In this transformation, contour is preserved but frequency bands for tone levels are not. With accumulating evidence that pitch usage can vary as much in Yorùbá as English on the large scale, it also important to acknowledge differences on the small scale. The Yorùbá and Nigerian English sermons are more angular in terms of pitch contour, with more frequent changes in direction, than the American or British English speech examined. A friend told me that my English was more melodious when I returned from 18 consecutive months in Nigeria, and that he had observed a similar effect among other friends who had extended stays in Sub-Saharan Africa, where tone languages are common. Whether
paradigmatic tone levels or syntagmatic tone intervals form a better model of lexical tone in Yorùbá, tone languages require changes in direction for lexical identification. English does not. Such changes are motivated in Yorùbá but not Nigerian English, or the English of Americans living in Nigeria, yet the contours of Yorùbá and other tone languages in Nigeria influence Nigerian English (both the “Queen's” English and less formal varities of Pidgin English). Angularity is best characterized by the prevalence of changes in pitch direction, local maxima and minima. The number of “turns”\(^{31}\) is summarized for each speaker in Table 4.9 (excluding British English). In Yorùbá and Nigerian English data, pitch segments that are at turning points in the pitch series account for more than half of the segments. There is more variability among American English speakers, with Obama having an angularity that is on par with the Nigerians. Trump is not far off.

<table>
<thead>
<tr>
<th>Yorùbá</th>
<th>Nigerian English</th>
<th>American English</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Speaker</strong></td>
<td><strong>Turns</strong></td>
<td><strong>Speaker</strong></td>
</tr>
<tr>
<td>Adediji</td>
<td>0.561</td>
<td>Adediji</td>
</tr>
<tr>
<td>Adekunle</td>
<td>0.522</td>
<td>–</td>
</tr>
<tr>
<td>Adesanya</td>
<td>0.542</td>
<td>Adesanya</td>
</tr>
<tr>
<td>Owadayo</td>
<td>0.532</td>
<td>Owadayo</td>
</tr>
<tr>
<td>µ</td>
<td><strong>0.539</strong></td>
<td>µ</td>
</tr>
</tbody>
</table>

Speakers with a lower proportion of turns will tend to have more consecutive descents and ascents. In a paradigmatic model of Yorùbá, the number of consecutive descents and ascents would be constrained because there are a limited number of levels. In a three-level system, two consecutive movements up or down one tone level fill the entire extent. This model seems somewhat accurate for the beginning of an utterance. However, once

---

31 Ladd (2008) calls these turning points. With reference to music, Mathesson (1983) calls them pivots.
declination takes control at the end of each utterance, long chains of descents may occur (see Figure 4.3). It is also quite common to have a perturbation of more than a semitone and remain at the same interval. This is addressed by the experimental findings in Chapter 5. This means that many the ups and downs for Yorùbá in Table 4.10 are not changes in tone level. If invariant frequency bands informed tone realization, I would expect an even greater proportion of Yorùbá segments to be turning points because late-phrase declination would be controlled.

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Turns</th>
<th>Same</th>
<th>Up</th>
<th>Down</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yorùbá</td>
<td>0.539</td>
<td>0.119</td>
<td>0.450</td>
<td>0.430</td>
</tr>
<tr>
<td>Nigerian English</td>
<td>0.532</td>
<td>0.112</td>
<td>0.427</td>
<td>0.460</td>
</tr>
<tr>
<td>American English</td>
<td>0.421</td>
<td>0.195</td>
<td>0.367</td>
<td>0.437</td>
</tr>
<tr>
<td>British English</td>
<td>0.423</td>
<td>0.188</td>
<td>0.386</td>
<td>0.424</td>
</tr>
</tbody>
</table>

Table 4.10

The proportion of “turns” is still greater for Yorùbá than in the data for American or British English. The higher incidence of turns in Nigerian English may reflect it is English spoken by L1 (primary language) Yorùbá speakers. Changes in direction are marked. They usually require more effort than staying around the same pitch or continuing in the same direction. In Yorùbá, these turns are motivated phonologically. Nigerian English does not require the turns, but apparently, it is hard to control this learned angularity. A common intonational shape for a single utterance (without a breath) is an arch (Bolinger 1983). This often is articulated by a leap up followed by numerous steps downward, giving way to the same late-phrase declination as Yorùbá. Huron has noted a similar structure for phrases in Western music (1996). The late-phrase downward contour caused by declination is common to many languages because we share the same
physiology and it is a result of the loss of sub-glottal air pressure. It can be counteracted but not without effort. Thus, upturned phrases are less common and have marked meaning. Like questions in English. The difference in Yorùbá (and likely other tone level languages) is in early phrase contour that is markedly angular. These generalizations about the shape of utterances between pitch resets are consistent with the summaries in Table 4.10. The early phrase in Yorùbá is usually angular, so there are more turns. In English, it is often a leap up followed by a gentle decline (unless there are other emphasis that disrupt this trend). Thus, there are more “downs” than “ups.” The distinct path at the beginning of Yorùbá and English utterances converges at the end in physiologically-induced declination. The motivation for pitch contrast in English is merely affective, while in Yorùbá it is both semantic and affective. Hence, English changes pitch less often and has fewer turning points. 32

4.5.3 MAGNITUDE

The previous section dealt with direction of pitch change, often referred to as contour. This section deals with the other component of pitch intervals: magnitude. Together, direction and magnitude form a vector between two adjacent segments. In the two measurements used in this section, coefficient of variation (CV) and normalized Pairwise Variability Index (nPVI), magnitude is considered without direction. For nPVI to be calculated, there must be no zero values in the data, so the “sames” from the previous

32 Here, changing pitch is operationalized as a segment that is at least a semitone higher or lower than the previous segment. Because the precision is a semitone and the values are conformed to the pitch standard (C4=60) it is possible that some of the “ups” and “downs” reflect less than a semitone of difference in mean pitch and some of the “sames” have nearly a semitone of difference.
section are not considered. “Ups” and “downs” are collapsed into one by taking the
absolute value of all intervals. In this way, the pitch interval magnitude is made similar to
duration, which only has magnitude and not direction. Duration is measured linearly. A
measurement from onset to onset is used here. Conversely, pitch is not linear. It can also
go up and down or even remain static between segments. In these magnitude-only
calculations, pitch is treated as if it were uni-dimensional. As a result, the mean and
standard deviations presented in this section are unlike those in previous sections. It is for
interval between segments, not the absolute pitch of individual segments. Pitch height is
not a component, so these calculations are normalized.

Patel suggests the coefficient of variation (standard deviation / mean) as a
measure of intonational variability (2005:63). In addition, I apply nPVI, designed to
model durational variability (Grabe and Low 2002). A normalised pairwise variability
index (1) calculates difference between each pair of successive measurements, (2) takes
the absolute value of the difference and (3) divides it by the mean of the pair. The last
step is the normalization. The mean of all pairwise comparisons is multiplied by 100
(Grabe and Low 2002:3). In the interest of having uniformity in the interpretation of both
onset intervals (durations in seconds) and absolute pitch intervals (magnitudes in
semitones), both CV and nPVI are applied to both dimensions of the signal. Although
nPVI has not been applied to interval series in the past, it is included here because it
reflects contextual information in the ordering CV does not. Consider the series $a$ and $b$:

$$a=[1,1,1,1,1,1,2,2,2,2,2,2,2,3,3,3,3,3,3,2,2,2,2,2,2,2];$$
$$b=[1,2,3,2,1,2,3,2,1,2,3,2,1,2,3,2,1,2,3,2,1,2,3,2,1,2,3,2];$$
They could represent pitch magnitudes (in semitones) or onset intervals (in tenths of seconds). They both have the same multiplicities of 1s, 2s, and 3s, so they produce the same mean ($\mu = 2$), standard deviation ($\sigma = 0.718$), and coefficient of variation (CV=0.359). However, the nPVI returned by each is not the same: $a$ is 4.73 and $b$ is 52.90. This reflects an important difference between $a$ and $b$ that is not captured by CV.

In $a$, an interval of 1 is likely to be followed by 1, while in $b$, this never happens. In $a$, adjacent intervals are mostly the same, while in $b$ adjacent intervals are never the same. CV, like mean and standard deviation, ignore ordering while nPVI makes pairwise comparisons in which the local ordering is important.

Local fluctuations in pitch interval magnitude may not be perceptually salient. Demany, Semal and Pressnitzer (2011) provide evidence that direction of pitch change (which is ignored in the nPVI calculation above) is more salient than magnitude. The diatonic scale has intervals of 1 and 2 semitones but these are treated and to some extent perceived as equivalent steps within the scale. The pairwise variability manifest in series $b$ is important in terms of production and, at the very least, present in the acoustic signal. However, it may or may not be perceived. For tone level languages, I would expect an interval profile more similar to $b$ than $a$, because there is motivation for big and small intervals in close proximity. Moving to an adjacent (M to H) then a non-adjacent tone level (H to L) is common. If frequency bands are present at all, whether fixed or moveable, there should be as much variety locally as globally if the spacing of tone levels is fairly constant within speaker. This makes nPVI a measure with utility that is distinct from CV because it is sensitive to local (pairwise) variation. In short, tone languages should have a high nPVI because of the phonological motivation for local variation in
interval magnitude. In languages with less or little lexical tone presence, there is little motivation for high pairwise variability. There is no constraint against it either.

Neither CV nor nPVI capture the range of interval magnitudes and proportions between them. Doubling every element in a semitone interval series [1,2,3,2,1] produces [2,4,6,4,2]. For both, CV and nPVI are the same. Fortunately, overall magnitude and variability is described by the mean and standard deviation. These four measures (µ, σ, CV, nPVI) are applied to both intonation and timing for the Yorùbá sermons (Table 4.11), Nigerian English sermons (Table 4.12), and all others (Table 4.13).

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Pitch Interval (Semitones)</th>
<th>Onset Interval (Seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>µ</td>
<td>σ</td>
</tr>
<tr>
<td>Adediji</td>
<td>3.433</td>
<td>2.737</td>
</tr>
<tr>
<td>Adekunle</td>
<td>2.584</td>
<td>2.087</td>
</tr>
<tr>
<td>Adesanya</td>
<td>3.402</td>
<td>2.289</td>
</tr>
<tr>
<td>Owadayo</td>
<td>3.191</td>
<td>2.125</td>
</tr>
<tr>
<td>µ</td>
<td>3.349</td>
<td>2.248</td>
</tr>
</tbody>
</table>

Table 4.11: Yorùbá Sermons

Among the Yorùbá sermons, the values for both intonation and timing are similar across speakers. They are not equivalent to the height and span measures, but there are consistencies. In Section 4.5.1, Adekunle had less variability in height and span. He also has a smaller mean interval size (2.584 semitones) and standard deviation from that mean. The range of onset interval nPVI for all Yorùbá sermons is less than 0.5.
Adediji and Adesanya both have more pairwise variability in timing when speaking in English. This is consistent with English as a stress-timed language. Multiple studies have observed larger nPVI values in stress-timed languages (e.g. English and German) than syllable-timed languages like French or Spanish (Grabe and Low 2002; Patel and Daniele 2003; Huron and Ollen 2003). The switch to English does not correlate to a difference in onset interval nPVI for Owadayo (57.2 versus 58.7). Among the three speakers for which I have both Yorùbá and English sermons, pitch interval nPVI is higher for English. The mean interval sizes are also larger for English, consistent with the larger spans (windowed standard deviation) when these speakers switch to English. Akin to the higher proportion of “turns”, this does not reflect a phonological motivation so much as the habit of local pitch variability, which is creative and not contrastive when speaking in English. The ways bilingual speakers manipulate linguistic tone and paralinguistic intonation are to some extent indistinguishable.
Table 4.13 is consistent with other accounts of American and British English: it has high durational variability. The mean calculated here is even higher than the British English onset interval nPVI of ~57 reported by Grabe and Low (2002:6). Their measurement is based on one speaker who was not a professional, so different values may be expected.

The variety in American and British speakers starkly contrasts the close values in Yorùbá for onset interval nPVI (all 57).

Table 4.13 also illustrates a fault with CV for measuring pitch interval variability, to which nPVI is not prone. CV is a normalization of the standard deviation to the mean of all the data ($\sigma/\mu$). The nPVI normalizes each pairwise difference to the pairwise mean. In general, the means and standard deviations of pitch interval are higher for the Yorùbá and Nigerian English recordings, but the CV values are lower. This begs the question: what better describes the role of magnitude within the larger picture of intonational variability? Is it local or global variation? King has the highest CV by a little and the lowest nPVI by a lot. CV is high because his standard deviation in interval magnitude is almost as big as the mean interval magnitude. However, both are relatively low values in themselves ($\mu=2.4$ and $\sigma=1.9$), so the variation represented by CV is quite small. Mean
and standard deviation are more useful for making a comparison across speakers than CV.

The nPVI brings out something that the mean and standard deviation ignore: local (pairwise) variability. Is a pitch change of a semitone likely to be followed by a semitone, or is a pitch change of a semitone likely to be followed by something bigger? Kennedy and King were quite similar in register variability (height and span), and their means and standard deviations are similar again here. However, the pitch interval nPVI tells us that King is more likely to have consecutive intervals of the same size, more likely than any of the other speakers examined. As Patel, Iverson and Rosenberg note, nPVI is a measure of contrast more than variability (2006:3035).

<table>
<thead>
<tr>
<th>Language</th>
<th>Pitch Interval</th>
<th>Onset Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>µ</td>
<td>σ</td>
</tr>
<tr>
<td>Yorúbá</td>
<td>3.349</td>
<td>2.248</td>
</tr>
<tr>
<td>NE</td>
<td>3.744</td>
<td>2.654</td>
</tr>
<tr>
<td>AE</td>
<td>2.950</td>
<td>2.265</td>
</tr>
</tbody>
</table>

Table 4.14: Summaries (British English not included)

The summaries indicate Yorúbá and Nigerian English have more interval variability in terms of mean interval size, standard deviation from mean interval size and pairwise variability. I find the CV calculation suggested by Patel (2005) less telling. American English has a higher CV because it has a smaller mean interval magnitude. My thoughts on the CV trend for pitch interval calculated in this analysis (visualized on the right-side of Figure 4.11), is that it is misleading. It distorts the fact that interval size is considerably

124
larger and more varied in terms of absolute size for Yorùbá and Nigerian English than American English.

![Figure 4.11](image)

The most interesting finding among these speakers is Nigerian English fuses the greater durational variability of American and British English (which is consistent across measurements) with the greater pitch interval variability of Yorùbá. Also of interest, Yorùbá sermons had very little variation across speakers for onset interval nPVI, while American speeches and British sermons ranged widely. Durational contrast is generally higher in native speakers of English, but also quite variable speaker to speaker. This observation is likely exaggerated by the selection of exceptional, and perhaps idiosyncratic, orators.
4.6 Intonational Model

The descriptive statistics for register height and span, and interval direction and magnitude, are not consistent with speaker-specific frequency banding of tone levels. Measurements of professional speakers speaking professionally are not necessarily representative of these very same speakers in more casual conversation, nor of speech in these languages in general. However, I find no evidence to suggest that Yorùbá is prosodically inhibited. Tone levels in Yorùbá are not fixed within a speaker's range any more than they are from speaker to speaker. Tone spacing does not extend to the full vocal range. It uses a narrower span, allowing the register to shift. Intonational shifting of the register may be used as an affective prosodic (paralinguistic) feature without making contrastive lexical tones unintelligible.

Both Ladd (2008a) and Patel suggest pitch does “double duty” in speech (2008:44). However, neither attempts to reconcile this with the range-based scaling model developed by Ladd and addressed by Patel (see Section 3.2). The use of intonation by the priests of All Saints Anglican suggests the range of tone language Speaker A (in Table 4.12) may not always be 100 to 200 Hz. I suggest absolute frequency range need not be constant for any speakers of Niger-Congo tone languages. Instead, it adapts to mood and environment. It may be the scaling (or span) remains constant while height within the speaker’s full range varies. If this is the case, the analysis suggests the span remains constant in log pitch (semitones) not frequency (Hz).

Analyzing Yorùbá sermons led me to a model of paralinguistic intonation in tone level languages (Figure 4.12). The y-axis is log pitch, similar to the plots in Section 4.4. The x-axis is a hypothetical continuum of arousal from low to high. The color trend lines,
blue, green and yellow show a tandem rise of the tone levels with increasing arousal. In this model, span is kept constant (in semitones) but it is also possible for arousal changes to have some effect on span, even in a tone language. It is unlikely however that low energy would allow span to become too small.

![Figure 4.12](image)

In general, span varied less than height. In each of the Yorùbá sermons, the priest started in a subdued mode (with a prayer) and rose to a fever pitch (with a register nearly an octave higher). This represents a move from left to right in the diagram. Some sermons ended with high intonation and energy, staying on the right side of the diagram. In others, the intensity was brought back down for a prayer at the end. A move of right to left in the diagram. A loud, almost shouting voice quality accompanied most of the high intonation in the All Saints sermons as well as American English speeches. In contrast, the British never raised their voices. In Obama’s gun-control speech he uses very high intonation after crying and it is not an aggressive voice, it is plaintive. In English at least, there are
affective cues associated with high intonation other than high arousal cues for assertiveness or anger. These include emotions such as grief or sadness. Though it was not observed in this data, I presume this is the case in Yorùbá. The visual model is too simplified to reflect this, but the presumption is that intonational modulation would adjust tone levels similarly for other emotional valences.

**4.7 Conclusion**

Deutsch, Henthorn and Dolson present findings “in accordance with the hypothesis that speakers of tone language employ absolute pitch as a feature of speech and that they refer to precise and stable absolute pitch templates in enunciating words” (2004:350). These findings were that Mandarin and Vietnamese speakers are more able to speak the same text at the same pitch on different occasions than English speakers. Elsewhere, Deutsch has argued that speaking a tone language, instead of an intonation language, increases the likelihood of perfect pitch among musicians (Deutsch, Henthorn, Marvin, and HongShuai 2006). The findings presented here are exploratory and not experimental, but they support a reconsideration of Deutsch’s findings. Do tone language speakers have more control of pitch than English speakers? Yes. Do they have to constrain intonation to be understood? No. We know that lexical tone paradigms must adapt to different speaker ranges. I find the proposal that lexical tone adapts to synchronous variation between speakers, but not diachronic variation within speaker, preposterous. Furthermore, if intonation is present in tone languages, then Deutsch’s distinction between tone and intonation languages is not accurate.
Ladd has speculated that intonation is present in languages like Yorùbá. Here is evidence consistent with that hypothesis. The question now becomes, how are tones understood with this much fluctuation in register height and span? According to Ladd:

… languages seem not to allow … ambiguities to arise - for example, Yorùbá in general seems not to use paralinguistic modifications of pitch range to signal greater emphasis on a particular word or phrase (see Rowlands 1969: 24ff.), presumably because this would result in too many occasions where lexical identity could be obscured or neutralised. (Ladd 2008:36)

Specific examples of local prosodic emphases were not evaluated in this chapter. I cannot determine what proportion of the pitch contrasts are lexical or not in the contour ornPVI measurements. However, I will speculate on what modifications are allowable or not, partially based on Laniran and Clements (2003) and Figure 4.3 (the “Loruko Jesu” raising and compressing). Laniran and Clements (2003) have already proposed that downstep and high-raising can change the height of tone levels within an utterance. These perturbations can accumulate. There can also be abrupt change on pitch reset. Ultimately, all levels can shift from the bottom octave of a speaker’s range to the top, with no overlap at all between these two registers. Paralinguistic modifications that change the pitch interval magnitude between two adjacent segments are not only possible but frequent. A modification that takes a high tone below an immediately adjacent mid tone is not so likely. My inclination is that these modifications are constrained. For Yorùbá, my theory is contour is determined linguistically while there is more flexibility for magnitude. The point where contour and magnitude intersect is on the small magnitude size: at what point is magnitude so small that a change in tone level is not perceived? Such a just-noticeable difference for speech tones would be larger than for frequency-stable sine waves, but how big? A movement of a semitone along the standardized chromatic scale qualified as a shift in tone level in this analysis. However, that may not reflect actual perception of
tones as same or different. The perception of tone contrast, including the minimum
distance for tone level change is the subject of the next chapter (5).

In this chapter, intonational variability (inclusive of tone) was divided up into
register and interval. Register was further divided into height and span, interval into
direction and magnitude. Measures, including some novel applications, were presented
for each along with rationale. Of these, the descriptive statistics for interval content and
variability were particularly abstract. A high proportion of “turns” indicates the signal is
angular or jagged, but not much more. A high nPVI means a step is likely followed by a
leap or vice versa, but it does not indicate what direction. Measurements for register
height and span (e.g. mean and standard deviation for windows of 100 pitch segments)
have a more concrete connection with the signal. From these, one can make a statement
in the manner of “roughly 50% of the segments fall within this semitone range”.
Including skewness would make that description more accurate. I considered the
possibility that some formula based on windowed mean pitch and standard deviation
could approximate the location of tone levels. This would be somewhat similar to the
visual model presented in the previous section (4.6). Figure 4.13 applies the functions
Mid tone level = \( \mu \), High = \( \mu + \sigma \) and Low = \( \mu - \sigma \) to visualize trends in tone levels for
Owadayo’s second Yorùbá sermon. However, this is more descriptive than predictive.
For one thing, it completely ignores late-phrase declination. If locations of pitch resets
were identified, then a more accurate approximation of tone levels could be made.
Windowed mean and standard deviation describe register height and span fairly well, but
I doubt that they can be used as a predictor of the pitch of a tone (e.g. \( L = \mu - \sigma \)).
The more important question is now: how is lexical tone perceived in the face of all of this intonational variation? If not frequency bands, then what? I have already hinted at my inclination towards a syntagmatic explanation. While this chapter piled up exploratory observations about intonational variability, Chapter 5 reflects an a priori, experimental approach. For now, I offer an answer to the question: Can pitch carry both linguistic and affective cues in Yorùbá? Yes.
5. PERCEPTION OF TONE INTERVALS

Introduction

Long dormant, debate over how to model tone systems has been reinvigorated in recent years (see Dilley 2005, Ladd 2008, Leben 2006 and Clements 2011). According to Dilley, “a syntagmatic tone interval [STI] relates two sequentially-ordered tones” (2005:2). This means that only adjacent pairs of syllables have tonemes, and is called an initializing system because the single syllable is unspecified (Ladd 2008a:190). This is the classical approach proposed by Fant, Halle and Jakobson (1952) and others. Since the 1970s, this perspective has been displaced by autosegmental theory, which requires a paradigmatic representation (after Goldsmith 1976). A paradigmatic tone interval “relates a tone [to] a speaker-specific referent level” (Dilley 2005:2). The referent is often conceived as a frequency band within a speaker’s range, so an alternative terminology is paradigmatic tone level. Because different people have different ranges the tone levels normalize to speaker range (Ladd 2008a:197). In this model, each syllable has a toneme based on where it falls within the frequency-banded range. Based on the findings presented in Chapter 4, I propose that if there are frequency bands in Yorùbá, they are not static within each phrase or over larger time spans. Lexical tone and paralinguistic intonation coexist within the fundamental frequency (f0) domain of Niger-Congo tone language speech, utilizing different temporal scales. Because of the faults with a
paradigmatic model in terms of phonetic analysis, this chapter revisits the syntagmatic model to see what perception has to say.

The two languages studied in this chapter are both Niger-Congo languages but have been distinguished as tone-terracing (Igbô) and discrete tone level (Yorùbá) languages. Dilley distinguishes them another way. Languages that allow paralinguistic modification are non-prototypical and languages that constrain paralinguistic modification are prototypically paradigmatic. In the previous chapter, it was demonstrated that tone levels in Yorùbá vary widely, especially in terms of height and to a lesser extent span. I question whether a distinction between Igbô and Yorùbá based on the presence or lack of paralinguistic intonation is accurate, and whether the distinction between prototypical and non-prototypical languages is meaningful. The constraint I propose for paralinguistic modification is that it cannot distort the pairwise adjacent direction (contour) between lexical tones.

Through a true experiment, the hypothesis that syntagmatic tone intervals form meaningful contrasts is tested. In a large collaborative effort, we asked the question: Is word identification possible based only on Syntagmatic Tone Interval (STI)? Listeners were asked to identify words from a minimal pair without the context necessary for making associative judgments. 1409 Nigerian university students, staff and faculty responded to one of two parallel studies, one for Igbô (n=654) and one for Yorùbá (n=755). My inclination is that tonal contrast is mostly embedded in pitch direction. Clark (1978, 1990) has already argued that Igbô tones are syntagmatic, and Dilley (2005) has to some extent concurred. Yorùbá on the other hand continues to be phonologized in terms of paradigmatic tone levels. The representation of tones as belonging to a level is
not at issue. Non-linguist Yorùbá speakers use the do-re-mi heuristic which is similar. However, there is no way to account for tone perception in all environments based on the conventional model of paradigmatic tone levels as static frequency bands. In the study, perception of tones was tested in the same way for a terracing and a level tone language, Ìgbò and Yorùbá respectively. Positive results demonstrate the perception of tonal contrast in these two languages is quite similar. There is nothing to suggest that lexical tone in one can, and the other cannot, adapt to paralinguistic modification.

**5.1 Minimal Pairs and Pilot Study**

All stimuli were based on disyllables found in scholarly dictionaries, Williamson (1972) for Ìgbò and Abraham (1962) for Yorùbá. These words were treated as two segments (e.g. /i.gba/, /m.ma/) each with one tone. This differs from Bamgbose's approach to Yorùbá phonology which interprets L.H disyllable words as having rising contour on the second syllable, producing three tones: L.LH (Bamgbose 2010). Sloped or contoured tones are not present in Ìgbò (Clark 1990) and because we are interested in cross-language perceptual schema, stimuli for both languages were prepared using the same process.

All possible types of tone-varied disyllable minimal pairs, differing only in the tone level of one syllable, are illustrated in Figure 5.1. The arrangement of the two-tone combinations on a circle is addressed in Chapter 11. In this figure, the black lines, along the circle and across it, indicate contrastive minimal pairs in which one of the two tones
is the same and the other is different. For example, ML and MH have the same first tone, but different tones in the second position.

Figure 5.1: Tone-varied disyllable minimal pairs in Igbo (6 types, left) and Yoruba (18 types, right).\footnote{The formalism behind these diagrams is in Chapter 11.}

Ostensibly, Igbo has only two levels. However, there is a downstepped High tone, represented as either \(+\) or S (for step), which sometimes forms words that are contrastive to homophones of HH or HL.

Initial word lists accounted for as many distinct relationships between minimal pairs as possible. From over 50 minimal pairs for each language, this was culled to 20 pairs based on the results of a pilot study. Because of the absence of word frequency data, testing familiarity of the words in the list was the best option. Words that were ambiguous in natural speech versions in the pilot were deemed in appropriate for speech
resynthesis in the experiment. Piloting also verified functionality and usability of the graphical user interface (GUI). In the end, all possible minimal pair types were included in the Ìgbò list, but only nine types were included in the Yorùbá list (unique tone contrasts are italicized in Table 5.1).

<table>
<thead>
<tr>
<th>Ìgbò Homophones</th>
<th>Tone Contrast</th>
<th>Yorùbá Homophones</th>
<th>Tone Contrast</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.fa</td>
<td>HH-HL</td>
<td>a.ra</td>
<td>MH-MM</td>
</tr>
<tr>
<td>a.ka</td>
<td>HH-HL</td>
<td>a.ro</td>
<td>MH-ML</td>
</tr>
<tr>
<td>a.kwa</td>
<td>HH-HL</td>
<td>ba.ta</td>
<td>LH-LL</td>
</tr>
<tr>
<td>a.kwa</td>
<td>HH-LH</td>
<td>e.ru</td>
<td>MH-ML</td>
</tr>
<tr>
<td>a.lu</td>
<td>HH-HS</td>
<td>i.la</td>
<td>MH-ML</td>
</tr>
<tr>
<td>a.wo</td>
<td>HH-HL</td>
<td>i.re</td>
<td>MH-MM</td>
</tr>
<tr>
<td>e.gbe</td>
<td>HH-HL</td>
<td>i.še</td>
<td>MH-LH</td>
</tr>
<tr>
<td>e.nyí</td>
<td>HH-HL</td>
<td>io.ko</td>
<td>HM-MM</td>
</tr>
<tr>
<td>e.ze</td>
<td>HS-HL</td>
<td>mi.ọ</td>
<td>HH-HM</td>
</tr>
<tr>
<td>i.gwe</td>
<td>HS-HL</td>
<td>mi.ọ</td>
<td>HM-HL</td>
</tr>
<tr>
<td>i.ke</td>
<td>HH-HL</td>
<td>mi.mu</td>
<td>HH-HM</td>
</tr>
<tr>
<td>i.si</td>
<td>HH-HL</td>
<td>o.do</td>
<td>MH-ML</td>
</tr>
<tr>
<td>m.ma</td>
<td>HS-HL</td>
<td>o.gun</td>
<td>MH-LH</td>
</tr>
<tr>
<td>o.du</td>
<td>HH-HS</td>
<td>o.ko</td>
<td>MH-MM</td>
</tr>
<tr>
<td>o.gu</td>
<td>HL-LL</td>
<td>o.kun</td>
<td>MM-ML</td>
</tr>
<tr>
<td>o.ha</td>
<td>LH-LL</td>
<td>o.ri</td>
<td>MH-LH</td>
</tr>
<tr>
<td>o.ke</td>
<td>HH-LH</td>
<td>o.kọ</td>
<td>MM-ML</td>
</tr>
<tr>
<td>o.kpa</td>
<td>HL-LL</td>
<td>pi.pa</td>
<td>HH-HM</td>
</tr>
<tr>
<td>o.nyá</td>
<td>HH-HL</td>
<td>si.ṣu</td>
<td>HH-HM</td>
</tr>
<tr>
<td>u.nyí</td>
<td>HH-HL</td>
<td>si.ṣun</td>
<td>HM-HL</td>
</tr>
</tbody>
</table>

Table 5.1: Homophone pairs used in the study

Not all 18 types in Yorùbá were included largely because the pairs in the initial list corresponding to some types were not familiar to our pilot participants. The types can be divided into two broader categories: pairs contrasted by STI direction (+, 0, -) and by STI magnitude. An example of the latter is HM-HL, in which the direction is descent (−) but the magnitude of change (in terms of level) is greater in the latter. Direction-varied and magnitude-varied pairs are represented in the experimental stimuli for each language.
5.2 Stimuli Preparation and Verification

First, a female and a male speaker of each language recited the list of homophones with both tonal variants and on neutral tone. For each minimal pair, one of two syllables was modulated at semitone increments from one pitch target to another. With a few exceptions, the neutral tone recording was modulated to pitch targets based on the tone-varied recordings. The modulations were made using Celemony’s Melodyne software.\footnote{A drawback to using Melodyne is the cost. Initially, Melodyne Essential was purchased for approximately $50 in 2012. The price has since increased, and more powerful versions of the software are hundreds of dollars. I recently upgraded to Melodyne Editor for $300. Other non-proprietary options for speech processing and synthesis are explained in (Banno et al. 2007), (Chazan et al. 2000), (Gold et al. 2011), (Kawahara et al. 1999) and (Kawahara et al. 2008). The capabilities of Kawahara et al’s TANDEM-STRAIGHT implementation in MATLAB is explored in (Banno et al. 2007) and (Kawahara et al. 2008).} Melodyne has the ability to “modify the pitch center” of segments without altering timing or formants. Its effectiveness is evaluated in the next section.

An example stimuli set is ákwá (crying) modulated to ákwà (cloth) (Figure 5.2 and Figure 5.3).

![Figure 5.2: images for /ʌkwa/ (HH-HL) minimal pair](image)
All modulations of a homophone (both male and female voice) formed the stimuli set for a minimal pair. In all, 452 experimental stimuli were created for Ìgbò and 405 for Yorùbá. The presence of different pitch (f0) trajectories in Yorùbá was not controlled, the preparation of stimuli was the same for both languages.

5.2.1 Pitch-Time Independence

If one is trying to maintain the quality of a vowel (in terms of formant resonance), in a speech or singing application, timbral consistency is essential. When a file is loaded into Melodyne it is automatically segmented. Individual pitch segments (notes) may then be edited, including combining or further segmenting notes. The Direct Note Access feature of Melodyne enables the user to manipulate duration, pitch (f0) and/or amplitude of an individual segment in a signal, while preserving timbral characteristics, in a graphic user interface. Because these elements are interconnected acoustically, to alter one without perceptibly changing another is an impressive accomplishment. The pitch shifting capabilities of Melodyne have been lauded with awards from the recording industry.
In the next section, stimuli sets for the Ìgbò language are examined to confirm that timbre was not significantly altered in the Melodyne-created audio signals. This will support the claim that modifying pitch alone, while keeping other features constant, can change the meaning of homophones. First, I address the techniques that make pitch-time independence possible in audio signal processing.

An historical challenge in effects processing is manipulating time and pitch independently of each other (Zolzer 2011). Tape machines that could manipulate time and pitch independently, such as the Phonogène universel, became available in the mid 1960s. By the 1970s, compact digital machines, such as the Harmonizer, that performed the function of the analog tape machines became available and were commercially successful. Though improved in digital machines, the introduction of artifacts and other distortions to the signal occurs outside of a practical range (usually less than octave in either direction from the original pitch). A plus or minus two-octave transposition produces timbres of a “character that is specific to the harmonizer” (Dutillex et al 2011).

Time-domain algorithms for pitch shifting include time-stretching (or compressing) and resampling to get the desired pitch. Delay-line modulation chunks the signal and then reads the chunks faster for higher pitch and slower for lower pitch. A third option, Pitch Synchronous Overlap and Add (PSOLA), is preferable for “constant timbre transposition”, ideal for applications to speech and singing where distortion of the formants would be the loss of information that is important in the signal (Zolzer 2011). The analysis phase of PSOLA extracts the local filter impulse response and approximates the local spectral envelope. After the sound analysis phase, PSOLA is very efficient computationally and is widely used in speech synthesis. Even in this method, a large
transposition creates some artifacts. These methods are effective for short segments, but for longer signals more advanced methods are needed, such as phase vocoding.

An alternative to the time-segment processes, are time-frequency processes that transform two-dimensional representations of sounds. As opposed to the manipulation of small time segments of sound, these processes work with what is best conceived of as a visual representation. Time-frequency processes follow a basic framework: analysis of a signal producing a representation, a transformation of the representation and a resynthesis of the representation back into sound. The analysis and synthesis scheme is called phase vocoding (Arfib et al. 2011).

Yet another possibility for timbre preservation is separation of source and filter, which uses a model of the source, such as the sound produced in the larynx, and models the filter based on the difference between the output sound and the source. Once the effect of the sound filter is removed, the model of the source signal can be modified (in terms of pitch for example) and the filter can be replaced, with the timbre preserved. Being able to independently manipulate source and filter is inherently artificial because the two interact in vocal production (see Ohala and Eukel 1987 and Whalen and Levitt 1995).

Unfortunately, neither the company Celemony nor the designer, Peter Neubacker, share details about the algorithms behind Melodyne software. It seems unlikely that a source-filter processing is used because the software is intended to deal with signals from a wide variety of sounds from the voice to the piano. However, it is possible that the part of the relatively lengthy analysis process when a file is loaded is approximating a generic source model for the entire signal. Similar to time-frequency processes explained in
DAFx (Zolzer 2011), with Melodyne, there tends to be timbral distortion when pitch-shifting to intervals of more than one octave from the pitch of the model signal.

5.2.2 Verification of Stimuli

Signal processing tools implemented in MATLAB (de Cheveigne and Kawahara 2002; Slaney 1998; Lartillot and Toiviainen 2007) were used to verify that only pitch varied across stimuli in a set, and there was little or no effect on other dimensions in the synthesis. Mel Frequency Cepstral Coefficients were used to measure and compare timbre between 5 to 13 different pitch heights for the transposing syllable in each stimuli set. In all, calculations were made for 39 sets: 20 homophone sets divided by gender with no male version for “unyi”. The evaluation of the resynthesized stimuli is summarized in this section. The transposing syllable was segmented in Audacity and imported in MATLAB, then MIR Toolbox’s “mirmfcc” function was used to produce a single value for coefficients 2 through 13 for the transposing syllable. Then, a difference was calculated between the coefficients of every transposed syllable and the natural syllable that was modified to produce it. A sum of differences were calculated for each stimuli (summing the differences for coefficients 2–13) and also for the same coefficient across all stimuli in the set.

A “Total sum” of differences was calculated by summing the sums of differences for coefficients, and a “Mean sum” of differences was calculated by dividing the total sum by the number of stimuli. The stimuli sets were sorted, and the 10 highest were examined in detail, both in a table of values like Table 5.2 and through listening.
Of the 10 stimuli sets with the lowest mean sum of differences, 9 sets were male voice, indicating that transposing a lower voice is less likely to introduce distortion. There were 5 /a/ vowels, 3 /u/, 1 /o/, and 1 /i/. The /u/ had a smaller range (5 semitones as compared with 12 for most). Of the 10 stimuli sets with the highest mean sum of differences, 8 were female, indicating that transposing a higher voice is more likely to introduce distortion.

The direction of the transposition in these sets was down, not up. A plausible explanation for the distortion in these sets, in addition to the higher female voices, is that none of the transposed syllables had a vowel nucleus of /a/. As an open vowel, /a/ has a higher 1st formant location and thus more flexibility about fundamental and formant interaction.

Among the ten highest mean sums of differences there were 5 /o/, 3 /i/ and 2 /e/, all mid or close vowels with a lower first formant. /u/ would likely have been in this group if the three homophones with /u/ had a larger transposition range. However, the minimal pairs with /u/ on the contrastive syllable are all small range contrasts. For example, a pair with a same tone homophone (high-high) and a homophone with a movement to an adjacent tone-level (high-step or high-mid), not to a nonadjacent tone level (high-low).
The 2nd coefficient often fluctuated but it did not seem to change the quality. Based on Figure 5.4, coefficient 2 may not have as great an impact on vowel quality because it is neither a minimum or maximum for any vowel except for /a/, which was not as strongly effected by transposition as other vowels. Vowel allophones are present in many Ìgbò and Yorùbá words. The presence of an additional and very active contrast (lexical tone) allows for greater variation in production of the vowels. Vowel space is less crowded in Niger-Congo languages than English. Ìgbò has 5–7 phonemic vowels depending on the dialect.

Stimuli sets where there was some concern about distortion were listened to by
the author and a speaker to confirm the fluctuation in timbre did not constitute a phonemic change of vowel. For one set, the female version of /oke/, could be interpreted as /uke/, which does produce nouns when spoken on high-high or low-high pitch. By and large however, Melodyne was found to preserve the timbre within the evaluation metric, producing a standard deviation value of less than .2 coefficient values in each set, for all but four sets of stimuli. These four sets were all modified recordings of female voice.

Pitch-shifting using Melodyne's Direct Note Access (DNA) feature is quite effective and easy to use right out of the box. While it does not seem to distort timbre significantly, another question is how it handles signals with continuous pitch between its automatically segmented “notes”. A method of evaluating pitch trajectory using DCT coefficients is proposed by Devaney et al (2011). One variation among Niger-Congo tone languages is whether moving between tone levels is connected through a sloping of pitch or is disconnected through abrupt changes to pitch. The major drawback to Melodyne is the lack of user control, including over pitch trajectory. In this study, pitch trajectory was not intentionally modified.

5.3 Experimental Design and Results

5.3.1 Task and Modules

The task was the same in the primer (5 iterations) and experimental module (20 iterations): (1) hear an audio stimulus (single word), (2) see two images representing a tone-varied minimal pair, (3) hear the audio stimulus again, and (4) make a selection from the two images.
A “forced-choice” format was used, wherein participants choose from one of two images (representing words of the minimal pair) with no alternative. In the primer module, the audio stimulus for each image pair was always the same and was a natural speech, not resynthesized. The purpose of the primer was two fold: to verify participants understood the task and had at least modest fluency. There were “correct” responses for the primer module, based on pilot study data, but there were not correct responses for the experimental module. For this module, most of the stimuli were resynthesized, but natural speech was also included for each set. In each set there were 5 to 13 transposed stimuli and 2 natural stimuli for each gender. The stimuli sets ranged in size from 14 to 30, with stimuli for both genders included in a single set for each homophone. Each participant heard only one version from the 20 experimental stimuli sets (grouped by homophone and image pair), thus many participants were needed to gather sufficient data. To avoid inter-stimulus effects, bias towards image side, and other undesirable outcomes, a complex randomization was used. The following parameters were randomized: order of image pairs / stimuli sets, image side, and selection of a stimulus from each set.
5.3.2 Participants

The pilot study was completed at Lagos State University in May of 2014. From May through July the experiment was conducted in locations throughout southern Nigeria, including University of Nigeria-Nsukka, Imo State University, University of Port Harcourt, University of Ilorin and University of Lagos. At times, two to four laptops with the GUI were set-up and monitored in public areas and volunteer participation was solicited. On other occasions, we were allowed to use an office and students within a department were asked to participate. Participants received no compensation but generally reported enjoying the task, which took 2–3 minutes. Often, participants returned with friends who also wished to perform the task. When groups of friends participated in succession, there was a tendency to crowd around a laptop. Study participation was not distraction-free in all cases. However, the instructions given to participants were not to carefully evaluate a stimulus. Participants were instructed to choose the image closest to their hearing of each audio stimulus, and it was emphasized that there was no “correct” response for the experimental stimuli. There may be some noise in the data as a result of the very public environments in which the study was conducted. If anything, I believe this will have weakened the effect, not strengthened it, creating false positive results. Responses from the primer (which did have “correct” values) were used to cull data from participants that either did not understand the task or were not fluent.
5.3.3 RESULTS

The results for female-voice versions of the homophone group /a.kwa/ are shown in Figure 5.6. Mean response rates for each experimental stimuli in the set are in colored dots. The x-axis indicates the semitone difference between the mean pitch of the two syllable segments. A positive value means it is an ascent (as found in the left plot). A negative value means it is a descent. /a.kwa/ is the only homophone with two minimal pairs included for the Ìgbò study (HH-HL and HH-LH), /mi.ọ/ is the only Yorùbá homophone with two pairs included in the study. The left plot in Figure 5.6 modulates HH to LH and the right plot HL to HH. In the left plot the first syllable is lowered, in the right plot the second syllable is raised.

Figure 5.6: Results for /a.kwa/ HH to LH (left) and HH to HL (right) for female voice
The y-axis indicates the response rate. For the left plot in Figure 5.6, 100% LH means all participants selected the left side of Figure 5.7, 100% HH means all participants selected the right side of Figure 5.7.

For the right plot in Figure 5.6, 100% HH means all participants selected the left side of Figure 5.2 (which is also the right side of Figure 5.7). 100% HL means all participants selected the right side of Figure 5.2. The color indicates the number of responses to the stimuli. Large stimuli sets (e.g. 13 female-voice experimental stimuli for /a.kwa/ HH/LH) had approximately 15–25 responses per experimental stimulus. Sets with fewer experimental stimuli had more responses per stimulus. Results for natural stimuli are also included with the same x- and y-axis positioning.

One way to evaluate the data is to apply a significance test for each experimental stimulus. With over 400 for each language, it is hard to synthesize the data into a meaningful result in this way. However, I have prepared such an interpretation for the data visualized in Figure 5.6. The data in the left plot of Figure 5.6 corresponds to Table 5.3, the right plot to Table 5.4.
<table>
<thead>
<tr>
<th>Interval (semitones)</th>
<th>Responses (n)</th>
<th>Mean (µ)</th>
<th>HH (µ=.973)</th>
<th>LH (µ =0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>19</td>
<td>0.947</td>
<td>0.632410</td>
<td>&lt;0.0000001</td>
</tr>
<tr>
<td>1</td>
<td>21</td>
<td>0.714</td>
<td>0.003117</td>
<td>&lt;0.0000001</td>
</tr>
<tr>
<td>2</td>
<td>23</td>
<td>0.000</td>
<td>&lt;0.0000001</td>
<td>NaN</td>
</tr>
<tr>
<td>3</td>
<td>22</td>
<td>0.045</td>
<td>&lt;0.0000001</td>
<td>0.346624</td>
</tr>
<tr>
<td>4</td>
<td>21</td>
<td>0.048</td>
<td>&lt;0.0000001</td>
<td>0.335420</td>
</tr>
<tr>
<td>5</td>
<td>17</td>
<td>0.000</td>
<td>&lt;0.0000001</td>
<td>NaN</td>
</tr>
<tr>
<td>6</td>
<td>14</td>
<td>0.143</td>
<td>&lt;0.0000001</td>
<td>0.284293</td>
</tr>
<tr>
<td>7</td>
<td>17</td>
<td>0.059</td>
<td>&lt;0.0000001</td>
<td>0.284293</td>
</tr>
<tr>
<td>8</td>
<td>17</td>
<td>0.059</td>
<td>&lt;0.0000001</td>
<td>0.284293</td>
</tr>
<tr>
<td>9</td>
<td>14</td>
<td>0.143</td>
<td>&lt;0.0000001</td>
<td>0.086047</td>
</tr>
<tr>
<td>10</td>
<td>14</td>
<td>0.143</td>
<td>&lt;0.0000001</td>
<td>0.086047</td>
</tr>
<tr>
<td>11</td>
<td>17</td>
<td>0.059</td>
<td>&lt;0.0000001</td>
<td>0.284293</td>
</tr>
<tr>
<td>12</td>
<td>21</td>
<td>0.143</td>
<td>&lt;0.0000001</td>
<td>0.082760</td>
</tr>
</tbody>
</table>

Table 5.3: Welch’s t-tests for stimuli in the set: /a.kwa/ HH-LH female-voice

In these tables, two t-tests are applied to responses to each stimulus, comparing them to the responses for each natural stimulus associated with the pair. The encoding from the Python GUI for the key responses was used, 1 for Image A, 0 for Image B responses, with side randomization counteracted. A mean of 1 indicates all responses were for Image A, a mean of 0 indicates all responses were for Image B. Welch’s t-test was used because it is less sensitive to variance than the Student’s t-test or z-test. The z-test does not return a value for data without any deviation (such as the natural stimulus for /a.kwa/ LH). The notion here is that if the p-value is less than .01, then the experimental stimulus is not the same word as the natural stimulus. In Table 5.3, the experimental stimulus with an interval of 0 returned p-values of .632 when compared to responses to natural HH and <.000001 when compared to natural LH. This indicates that it may be the same word as the natural HH stimulus, but it is nothing like the natural LH stimulus. The NaN values in Table 5.3 are produced by a lack of variance in either of the data sets compared, experimental stimuli with intervals of 2, 5 and 6 semitones and the natural stimulus LH had 100% Image B responses (a mean of 0 and standard deviation of 0). This indicates
they may be the same word. Conversely, intervals 2–12 all produced p-values of <.000001 in comparison with responses to natural HH.

<table>
<thead>
<tr>
<th>Interval (semitones)</th>
<th>Responses (n)</th>
<th>Mean (µ)</th>
<th>HH (µ=0.973)</th>
<th>HL (µ=0.046)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-12</td>
<td>21</td>
<td>0.048</td>
<td>&lt;0.000001</td>
<td>0.973916</td>
</tr>
<tr>
<td>-11</td>
<td>15</td>
<td>0.200</td>
<td>&lt;0.000001</td>
<td>0.144963</td>
</tr>
<tr>
<td>-10</td>
<td>16</td>
<td>0.125</td>
<td>&lt;0.000001</td>
<td>0.382945</td>
</tr>
<tr>
<td>-9</td>
<td>17</td>
<td>0.118</td>
<td>&lt;0.000001</td>
<td>0.414853</td>
</tr>
<tr>
<td>-8</td>
<td>20</td>
<td>0.500</td>
<td>0.0000003</td>
<td>0.0000465</td>
</tr>
<tr>
<td>-7</td>
<td>17</td>
<td>0.118</td>
<td>&lt;0.000001</td>
<td>0.414853</td>
</tr>
<tr>
<td>-6</td>
<td>12</td>
<td>0.667</td>
<td>0.001780</td>
<td>0.000013</td>
</tr>
<tr>
<td>-5</td>
<td>17</td>
<td>0.471</td>
<td>0.000002</td>
<td>0.001179</td>
</tr>
<tr>
<td>-4</td>
<td>29</td>
<td>0.621</td>
<td>0.000132</td>
<td>0.000005</td>
</tr>
<tr>
<td>-3</td>
<td>14</td>
<td>0.929</td>
<td>0.475967</td>
<td>&lt;0.000001</td>
</tr>
<tr>
<td>-2</td>
<td>20</td>
<td>0.950</td>
<td>0.659733</td>
<td>&lt;0.000001</td>
</tr>
<tr>
<td>-1</td>
<td>12</td>
<td>1.000</td>
<td>0.574461</td>
<td>&lt;0.000001</td>
</tr>
<tr>
<td>0</td>
<td>23</td>
<td>1.000</td>
<td>0.435142</td>
<td>&lt;0.000001</td>
</tr>
</tbody>
</table>

Table 5.4: Welch’s t-tests for stimuli in the set: /a.kwa/ HH-HL female-voice

Only one experimental stimulus in the /a.kwa/ HH-LH set had responses significantly different from either natural stimulus: the interval of 1 semitone (see Table 5.3). The values in Table 5.4 demonstrate that experimental stimuli with intervals of -8, -6, -5 and -4 were incongruent with both natural stimuli. The threshold between H and L tone levels is closer when ascending than descending. This suggests the presence of a tone level in-between H and L when descending. There is the down-stepped high tone or step tone (S), which does not have full status as a tone level in Ìgbò tonologies, but is sometimes lexically contrastive (such /a.lu/ HH-HS). [á.kwá] (HH) and [á.kwá] (HS) are sometimes allotonic in fluid speech (as in the tongue twister in Chapter 1), but not here. Regardless of the status of step tone, there is an ambiguous range of intervals between HH and HL suggested by responses to this stimuli set.
An anonymous reviewer of a short conference paper version of this chapter expressed the concern that with so much data “any results might show up as statistically significant.” In actual fact, there is not so much data per stimulus, as one can see in the tables above. These tables demonstrate the same neutral tone stimulus\(^{35}\) can be modulated to have three different meanings, and this is statistically significant. The p-values for individual stimuli are significant not because of huge numbers of responses, but a modest number of responses and a very strong effect. Other stimuli sets produced similarly significant results, with the exception of /o.gun/ in Yorùbá and /o.ha/ in ìgbò.

In considering more, and indeed the rest, of the data, it would be pedantic to demonstrate over and over again that the results are significant. As the reviewer indicated, it does not take much for a regression of 2500 datapoints to be statistically significant. It is however pertinent to consider effect size, which is done here through a coefficient of determination (r\(^2\)).

Before considering the data on the large scale, first let us consider the /a.kwa/ plots in Figure 5.6 from the effect size perspective. A visualization of the audio data appears in Figure 5.8 and Figure 5.9 respectively. The fundamental frequency plot in the top of each figure shows the pitch contour of each female-voice stimulus in the set, and the MFCC plot in the bottom of each figure verifies that timbre remains largely intact through the modulations. Note that the first stimulus in both figures is the same. In Figure 5.8, the first syllable is modulated in each subsequent resynthesis, and in Figure 5.9, the second syllable is modulated.

\(^{35}\) The same neutral recording was used for all female-voice /a.kwa/ experimental stimuli.
A sigmoid curve (or *s-curve*) was fit to the data in both plots of Figure 5.6, produced $r^2$ values of 0.947 and 0.854, respectively. These coefficients speak to the strength of the phenomenon triggered by transposing a single syllable. An inflection point in the sigmoid curve between predominantly HH and predominantly LH responses is calculated at 1.1 semitones for the stimuli visualized in Figure 5.8. This means that after the first stimulus in the f0 plot, the rest tended to be interpreted as LH. To the left of the first dotted red line in Figure 5.8 participants predominantly chose the “crying” image, and to the right they tended to choose the “egg”.

Figure 5.8: Fundamental frequency plot (using YIN) and MFCCs (using MIR toolbox) of the entire stimuli set for /akwa/ modulated from HH to LH
An inflection point of -5.5 semitones was found for the stimuli in Figure 5.9. It is useful to consider the individual stimuli means and p-values here and recall that intervals of –8 through –4 were unlike both natural stimuli. The s-curve regression does not model the data for /a.kwa/ HH-HL as it does for HH-LH. This and other minimal pairs that skip over a tone level have ambiguous stimuli in the middle that are not accommodated by the s-curve. /a.kwa/ HH-LH has a higher effect size (.974) and steeper slope at the point of inflection because there is no “step” tone (S) in ascent.

HH-HL minimal pairs are common in Ìgbò. These two tone sequences provide a very clear contrast between homophones through in their syntagmatic tone interval (STI). The difference in STI direction (0 v. –) is reinforced with a decisive drop from the high tone (H) past what could be interpreted as a down-step (S) to the low tone level (L). This is a stronger contrast than found in HH-HS or HS-HL minimal pairs because it makes use of both STI direction change and large STI magnitude. However, because HH-HL crosses an adjacent tone-level to a non-adjacent tone-level it is excluded in Figure 5.10, which
only displays results for minimal pairs that are contrasted by tone alternation between adjacent levels.

Figure 5.10: Sigmoid-curve regressions and inflection points for Ìgbò data

Figure 5.10 shows s-curve regressions and inflection points for aggregate data by minimal pair types. This means that data for all minimal pairs composed of the same tone sequences (e.g. HS-HL had three: /e.ze/, /i.gwe/, /m.ma/) were aligned by interval size for both female and male voice stimuli. Means for each stimulus were calculated, then means calculated for all stimuli of the same interval size across words and genders. The regressions were calculated from these values. Inflection points and $r^2$ values are included for each regression, the effect size is quite high for each aggregate set. The plot of all aggregate regressions is normalized to the same minimum and maximum values (the maximum response rate for Tone A/Image A is the minimum).

In both Figure 5.10 (Ìgbò data) and Figure 5.11 (Yorùbá data), the sigmoid curves for distinct syntagmatic direction (e.g. HM to HH), have a sharp slope, but the sigmoid curves for distinct syntagmatic magnitude (e.g. HL to HM), do not. This indicates the
threshold between the perceptual categories HH and HM, for example, is quite crisp, while the threshold between HM and HL is fuzzier. Generally, STI direction is a stronger contrast than STI magnitude alone.

Regardless of the crispness of the perceptual categories for STI direction and STI magnitude, a rather astonishing result is the similarity between the location of the inflection points for Ìgbò (Figure 5.10) and Yorùbá (Figure 5.11). Ìgbò is a terraced two-tone language with down-step, while Yorùbá has three discrete tone levels. Hence, there is no corresponding inflection point in Ìgbò for Yorùbá’s MH-LH. A worthy post-hoc criticism of the experimental design by D. Robert Ladd is that it discounts the effect of segment slope (contoured tones) in Yorùbá. This is a weakness with regard to understanding Yorùbá tonology specifically. However, the approach of synthesizing stimuli with the same process for both languages strengthened the argument for a cross-linguistic perceptual schema. Despite very different accepted phonological models and to some extent distinct phonetic realization of tones by Ìgbò and Yorùbá speakers, the
findings indicate syntagmatic direction may be used to differentiate common tone-varied homophones. Additional paradigmatic features such as slope of a tone may make such judgments more secure, but direction alone, without greater context is enough for most of the homophones studied.

5.4 Evidence for Syntagmatic Direction

For me, the significance of these findings is not in the t-tests for responses to single stimuli, or even the effect size, but in the consistency of the findings across words, different voice ranges (gendered voices), and most importantly, across languages. In both Îgbò and Yorùbá, word identification is closely tied to the direction between adjacent tones (+, 0, -). Our study had weaker results for distinguishing homophones that only differ in magnitude, not direction. A model of the constrastive strength of minimal pair types is found in Figure 5.12. An adaptation of the minimal pair diagram Figure 5.1, Figure 5.12 displays the strongly contrastive minimal pairs (differing in STI direction) in green and the weaker contrasts (differing in STI magnitude) in yellow.
5.4.1 TONE-LEVEL FISSION AND THE PITCH PROXIMITY PRINCIPLE

HH, MM and LL disyllables are not entirely flat, there is some wiggle room. While crisp thresholds (with a sharp slope in the regression) were not found for STIs of +1 and +2 or −1 and −2 tone levels, magnitude of change does play a role in differentiating STIs of 0, +1 and −1. There is evidence for a cross-language minimum magnitude of change of +2 or −3 semitones to confidently leave a tone level. This forms an asymmetrical pitch-height window around the first tone that determines the STI direction, allowing for some fluctuation in pitch while remaining at the same tone level. Because this window is larger than the range of f0 perturbations attributed to intrinsic pitch of vowel (IPV) effects, it follows that IPV is not constrained in Niger-Congo A tone languages. This corresponds to the concept of a “Trill Threshold” introduced by Miller and Heise (1950), later refined as a “fission boundary” van Noorden (1975) and also...
much work by Bregman. Van Noorden found a fission boundary of 2 semitones or less for tones under 700 ms in duration. Huron coalesces this research into a pitch proximity principle for voice-leading in homophonic music (2001:21). The similarity between itinerant tone levels and voice leading is explored in Chapter 11. More immediately, the reliability of syntagmatic direction as a contrast and the minimum interval (+2/–3 semitones) that constitutes a change of tone level form the basis for an initialization of tone levels in the next chapter.

5.4.2 SYNTAGMATIC MAGNITUDE

Ìgbò and Yorùbá speakers can judge magnitude but it may take longer and the threshold between small and large magnitude perceptual categories varies from person to person. In a psychoacoustic study (not specific to tone language perception), Demany, Pressnitzer and Semal (2011) found the ability to judge change of direction is nearly automatic and is an ability most people have. Conversely, judging magnitude of pitch change has a greater cognitive burden and is generally not as accurate. In general, making perceptual judgments of magnitude is hard in all domains, not just pitch (Donkin, Rae, Heathcote and Brown 2015). An even more recent study by Bidelman and Chung (2015) suggests that tone language speakers (Mandarin speakers were studied) have hemipherization of contour (here direction) and interval (here magnitude) processing. All of these studies, and our own, suggest that the brain processes direction and magnitude differently, and I takes this as justification for doing so in a formal linguistic model. Because magnitude alone was not as strongly contrastive as direction, some paradigmatic
cues, in pitch (f0) trajectory or non-formant timbre (such as aperiodicity in the low range) may help listeners to distinguish homophones that only differ in magnitude of change, such as HM from HL. I cannot evaluate the role of such cues because our experimental stimuli do not reflect these effects. Another possibility developed in the rest of this dissertation, is that contour theory (of music) which compares pitches beyond immediate adjacency, may provide an alternative to a strictly paradigmatic or simple-adjacency syntagmatic model of tone perception.

![Figure 5.13: /igba/ homophone collection on minimal pair circle with ascent triangle](image)

A better understanding of the role of syntagmatic relationships in tone perception will contribute to a more developed theory of generative tonology and constraints on Niger-Congo A tone languages. Figure 5.13 takes the same minimal pair circle for Yorùbá from Figure 5.1, but with a complete homophone collection imprinted. All high-level initial positions are blank because there is a phonetic constraint against high-tone on the first syllable of vowel-initial words. If redundant positions are ignored (MM and LL), only
one two-tone combination does not surface for the homophone /i.gba/. The lack of LM within an otherwise populous homophonic space can be explained through constraints on syntagmatic tone intervals and the relationships they form between homophones. First, no homophone collection includes all three possible ascents (connected by the yellow triangle in Figure 5.13). Furthermore, no homophone collection includes both LM and MH, which are identical in terms of syntagmatic tone interval. In general, MH is more common than LM likely because it is easier to produce a movement from an unmarked tone to a marked tone than vice-versa. So, although LM and MH are equivalent in terms of syntagmatic perception, they are productively distinct, which likely leads to secondary paradigmatic cues.

By stripping away all paradigmatic context, we found that disyllables could be identified based on syntagmatic direction, and to a lesser extent syntagmatic magnitude. While the role of non-adjacent syntagms are developed in the coming chapters, this says nothing about how single syllable words are perceived--words that as Ladd points, cannot have a syntagm. I propose that paradigmatic cues other than frequency banding of pitch height can be used to identify these words in the event that segmental phonemes are not enough. As Ladd has pointed out in correspondence, speech uses many different features and it is naïve to think a model restricted to one feature is fully explanatory (2015). Thus paradigmatic cues, such as pitch trajectory and registral effects, that strengthen the contrast between same direction minimal pairs (LM and LH) may also help to differentiate single tone words, on L, M or H. Furthermore, such non-pitch height cues could initialize or re-initialize tone levels. Although, this is not elaborated to a great
extent in the coming chapters, it is assumed that these features help out especially where
a model based exclusively on syntagmatic tone intervals has its limits.

Common Practice Tonality uses both syntagmatics and paradigmatics: in a
monophonic melody, key is established through syntagmatic intervals between pitches,
then pitches become paradigmatic scale degrees. Establishing a key initializes the scale
degrees which then become paradigmatic and take on a quality. This process can be
strengthened or undermined by adding harmony. Although the diatonic scale is built on a
network of superparticular ratios and I am not suggesting that African tone level systems
are, is it possible that the initializing-to-normalizing model of tonality applies?
I first interpreted the do-re-mi heuristic as an indication that tone was relative pitch, but
more recently, I have thought tone could be initialized through syntagmatic relationships
and then enter a paradigm that is sustained in gentle shifting, but can also be disrupted
and reinitialized. Instead of an absolute frequency band within the speaker’s range or a
refined contour relationship to the proceeding tone, I think the referent could be a
network of adjacent and proximal non-adjacent tones, similar to traditional contour
theory but without the requirement of prior segmentation into groups of pitches. This
model is consistent with the evidence that phonological equivalence is not established
through absolute pitch height across voices, or even within a single speaker's range.
6. INITIALIZATION AND TEMPORAL SCALES

Introduction

Evidence from the last two chapters supports two claims about tone level languages. First, the pitch dimension carries both linguistic signals and paralinguistic cues (Chapter 4). Second, changing pitch direction can change the meaning of words regardless of absolute pitch height and without context (Chapter 5). These statements are complementary. Placement of tones within the range can vary significantly so long as directional relationships within words are maintained. To have interval direction we must have a window around a syntagmatic referent (antecedent realized tone) which helps to categorize adjacent pitches as the same or different. Based on findings presented in Chapter 5, the “fission boundary” (after van Noorden 1975) for Igbo and Yoruba tones is roughly the same. This makes sense because they are geographically-proximal and genetically-related cultures. In these two languages, if a consequent segment is two semitones above or three semitones below its antecedent (+2/–3), the tone level has changed. The up, down or same (+,0,–) ternary contour categories are sufficient for most homophonic minimal pairs, but not all. Homophone groups with a HM and HL variant are found in the Yoruba lexicon. Instead of using refined contour categories to differentiate such words (e.g. +2, +1, 0, –1, –2), I propose a model based on initializing tone levels and keeping track of tone level orientation, similar to auto-correlation.
6.1 Tone Level Initialization

I propose that adjacent and non-adjacent pitch-height comparisons contribute to the formation of perceptual categories. This has a basis in musical contour theory, which is elaborated in the next chapter. A distinction between directional (contour) and magnitude (interval) components of pitch-height vectors is supported by behavioral and neurophysiological studies from the 1970s to the present, including classic studies by Dowling and Fujitani (1971) and Edworthy (1982) as well as recent findings by Demany, Semal and Pressnitzer (2011) and Bidelman and Chung (2015). If the role of syntagmatic direction is clear, but role of magnitude in question, how can three or more categories of tone levels be perceived? Welmers suggests that three tone levels may be realized as pitches of a C-Major triad and produce intelligible speech (1973:81). This would be quite artificial for speech and uninteresting for a musical setting, but it is consistent with the fission boundary found in the last chapter (+2/–3 semitones). While a mapping to a C-Major triad would not distort tone levels, there are a rich variety of realizations found in speech and song that do not resemble the frequency bands implied by Welmers’ description.

With each breath, at the very least speakers reset pitch height and span. In the Yorùbá sermons analyzed in Chapter 4, a new height and span will be introduced with a new utterance different from the last. This suggests tone levels may be initialized with every new utterance. When a listener knows the tendencies of a specific speaker, this process is undoubtedly faster. Listeners may habituate to a speaker’s use of interval magnitude, but there is across speaker and within speaker variation. Having a flexible framework for perceiving syntagmatic magnitude allows listeners to adapt to different
speakers and to intonation changes. If a listener does not know a speaker’s tendencies, they become familiar with them through statistical learning. Furthermore, the interval magnitude between two levels is not constant nor is the location of the levels in pitch space, so there is shifting that takes place almost as soon as initialization has occurred.

Syntagmatic direction between adjacent tones can only account for tone intervals in two-level systems: + is a Low to High (LH), – is a High to Low (HL) and 0 is staying at the same tone level (LL or HH). However, Ìgbò is a two-level system, but is confounded by a downstepped high tone (!H or S) that is sometimes tonemic. An alternative to having categories of magnitude size for all speakers is to use non-adjacent contour. Musical contour theory is elaborated starting in the next chapter, but a basic tenet is apparent enough to start our deployment here: directional relationships are formed between adjacent and non-adjacent pitches. We can apply two height labels to two pitches that are not the same: one is higher and one is lower. If two pitches are at the same perceived height it is impossible to speak of relative height. If two pitches at different heights are joined by a third at a different height from the other two, we can apply three labels about relative height: one pitch is the highest, one the lowest and one is in the middle. In theory, any ordering of three tones at distinct levels can initialize a three-level system: all ascents (LMH), all descents (HML), ascent-descent (MHL or LHM), or descent-ascent (HLM or MLH).

To be able to say that pitches are at different heights, we must have some criteria for sameness versus difference. In musical contour theory, this is typically where notes fall on a musical scale (often the chromatic scale). Instead of a musical scale with stable frequency bands, e.g. a note that at 440 Hz plus or minus 50 cents is A4, I propose a local
window of sameness. This is based on the strong cross-language perceptual category of sameness found in the study in the previous chapter $<+2, -3>$. The assymetrical window of sameness reflects a general tendency to lose height even among phonologically equivalent tones. Once a threshold between same and different is established, comparisons of adjacent and non-adjacent events can be made. Having further comparisons helps to categorize pitch events into tone levels in systems with more than two levels. To illustrate this concept, I will use Marvin and Laprade's CSEGs (contour segment$^{36}$ classes) from their 1987 article. First consider the CSEG $<1 0>$. The integers represent contour pitches numbered from low (0) to high (the cardinality of the segment minus one (n–1)). $<0 1>$ could be MH or LM in a three-level system such as Yorùbá. If we combine it with another dyadic contour (either $<0 1>$ or $<1 0>$ ) with a single overlapping nexus point and the combined contour is decentric (does not return to the initial tone level) then we have established three levels:

$$<1 0> + <1 0> = <2 1 0>$$

In the resultant CSEG, $<2 1 0>$, the first two segments are at distinct tone levels, both of which are higher than the level of the third. Friedmann (1985) suggests the notion of a contour interval (similar to refined contour). The contour interval between the first and last elements of the CSG $<2 1 0>$ is $–2$, making them respectively, high and low tones in a three-level system.

Combining $<1 0>$ with $<0 1>$ with a single nexus point yields two possibilities.

$$<1 0> + <0 1> = <1 0 2> \text{ (or } <2 0 1>)$$

---

$^{36}$ From this point forward, segment is used to refer to groupings of pitch events instead of single pitch events.
The first is \(<1\ 0\ 2>\) in which the last element is the highest. The second is \(<2\ 0\ 1>\) in which the first element is the highest. Based on these tenets of contour theory and fission boundaries of +2/–3 semitones, an acoustic model is shown in Figure 6.1.

Three plots are included, each representing a stage in the initialization process. In plot 1, one tone with an absolute pitch height of 59 semitones (where 60=C4) appears. The fission boundaries are indicated as dotted lines. In this model, a single tone is unspecified, it is impossible to tell whether a tone is at the High, Mid or Low level in isolation. If the next tone falls within the dotted lines, it will be perceived at the same (still unspecified) level. If a disyllable homophone is formed and there is only one same level homophone in the group (e.g. MM), this is sufficient to correctly identify the word (as demonstrated in Chapter 5). In plot 2, a second tone appears that is three semitones below the first (56 semitones), producing a CSEG of \(<10>\). These two pitches are different, but it is unclear where they fit within a three-level context. Is it HM or ML?

With two different levels, the pitch space is now primed for full initialization, dependent on the relative height of a third tone. If the third tone is at or below the bottom dotted line
in the middle plot of Figure 6.1, all tones are now specified, forming a sequence of HML. Going forward, a rise of <+2> will enter the mid tone level, and a rise of <+5> will enter the high tone level. All three outcomes are illustrated in Figure 6.2, with full initialization reflected in the color coding of the tones and boundaries (blue for high, green for mid and yellow for low.

Comparing local adjacencies using boundaries of (+2/–3) suggests an interval sequence of <-3, -3> is reliably HML, and is similar to Welmers’ description covered at the beginning of Chapter 4. However, these are minimum distances, so HML could also be <-4,–3>, <-3,–4>, <-5,–5> and so on. The minimum downward magnitude is larger than the do-re-mi heuristic implies. HML (plot 1) is not the only outcome for the priming in Figure 6.1. An alternative is plot 2 of Figure 6.2 in which the third tone is 2 semitones higher than the first. The interval series <-3,+5> produces the tone level sequence MLH. This alternative initialization has produced a different tonological identity for tone 1 and tone 2 although they have the same absolute pitch in both the first and second plot of Figure 6.2. All three tones in these two initializations are at different levels according to local contour comparisons. In the first plot the last tone is the lowest. In the second the
last tone is the highest. In the third plot of Figure 6.2, the last tone is in the middle. This
does not produce HLM because it is within the window of equivalence for the first tone.
The interval series <-3,+2> is underspecified, it could be either MLM or HMH. The
sequence is still primed for a third level different level from the previous two. A tone
above the top dotted line or below the bottom dotted line would fully contextualize the
sequence.

Because of the close spacing of the first two tones, it was not possible for a third
tone to appear at a pitch between them and be at a distinct level. Mid tone can be the final
step in the initialization if the distance between the first two tones is substantial enough.
Figure 6.3 shows the process for HLM initialization with an interval series of <-5,+2>.

In this model, an interval series of <-5,+3> is underspecified because the first tone and
the last tone are within the same window (Figure 6.3, plot 1). An initial interval of <-5>
can also be part of a HML (plot 2) sequence or MLH (plot 3) if the third tone is a new
low or new high.
In plot 2 and plot 3, tone level dispersion is not uniform. In both cases the third tone is far enough above (+2) or below (–3) all prior tones to assert itself as a distinct level. The distance between the first two tones was wide enough to be two levels apart, but because the third tone initialized a new high or low level and the system is constrained to three levels, the potential for a level between the first two tones is lost.

The asymmetrical window of equivalence around a tone (+2/–3) means a smaller magnitude is necessary to initialize tone levels in ascent than descent. An interval series of <+2,+2> produces the tone sequence LMH. In this case, the minimum magnitude for ascent is identical to the do-re-mi heuristic. Do-Re-Mi is [0,2,4] in normalized pitch space, so is the LMH initialization in Figure 6.5.
After this initialization, an interval of –1 or –2 will still be in the high level, –3 or –4 will indicate a movement to mid, and –5 or more will be at low tone level. The threshold between high and mid (the green dotted line) is based on the window around the current tone (High). The threshold between mid and low (the yellow dotted line) is based on the window around the immediate prior (and now specified) mid tone. A rise of any magnitude after the last tone plotted in Figure 6.5 will remain within the high tone level. Likewise, anything beyond the mid to low threshold will be low tone. This implies that mid tone level is more constrained than high and low. This suggests high tone level may be raised and low tone level lowered freely, similar to observations in Laniran and Clements (2003). Laniran and Clements also note that a sequence like LHH may be realized with the first H halfway between the L and second H in pitch height. This brings forth the possibility that the last frame in Figure 6.5 is not LMH at all, but LHH.

6.2 Lexical Priority

Speech analyses are consistent with the framework in the previous section early in an utterance, but as semantic momentum grows there is less of a burden on tone. All forms of contrast are more important at the beginning of an utterance than at the end. Tone contrast is particularly obscured by declination right before pitch reset. Intralexical contours (within word boundaries) remain largely intact, but there are often interlexical breaks (between words) in the orientation to paradigmatic tone levels. Tone fluctuates in terms of importance. The tone level paradigm may become obfuscated, but it can be brought back on track at any point through lexical induction. Lexically-informed tone
level orientation, based on holistic word identification, takes precedent over the deductive process outlined in the previous section. The same process can also override the current orientation within the paradigm. Word identification based on phonic content and syntagmatic direction may force the paradigm to shift. This may be interpreted as an error by a listener, but is not so ambiguous as to be unintelligible.

Lexical priority and an inductive process of tone level orientation can be illustrated with the LMH initialization first presented in Figure 6.5. According to the initialization process presented thus far, a fourth tone <+2> higher than the third tone would also be at the high level (plot 1 of Figure 6.6).

![Figure 6.6](image)

If the first two tones form a LM word, the second to third tones a MH word, or the first three form a LMH word, all is well. However, a different paradigm may be induced if the first two tones are part of an LH word (skipping the mid level as in the middle plot) or the second and third tones are part of an LM lexical tone (as in the third plot). When there is only one direction and multiple magnitudes associated with the phonic content of a word (e.g. HM and HL), a speaker must be more careful about magnitude and following the existing paradigmatic orientation is helpful. In situations where there is only one ascent
or descent associated with a disyllable homophone, any magnitude in the right direction will be correctly identified, so the speaker may be more liberal with magnitude. This was observed in the experimental data. Stimuli shifted past the natural speech model for change of one tone level continued to be identified based on syntagmatic direction even if the magnitude was as great or greater than what may be a span of two tone levels in natural speech. Because of the binary choice between a minimal pair, with no third option, it may be the speech sounded unnatural and the participants simply chose the best option. However, I propose listeners would make a similar choice based on the same available information in conversation.

Perceptual data indicates syntagmatic direction within a word is crucial to word identification. I have no experimental findings about inter-lexical relationships. Phonetic analysis indicates syntagms formed by the tone levels conventionally ascribed to words (by dictionaries, etc) remain largely intact, but there is also a lot of deviation (see Figure 6.7). Some of this may be described with complex syntactic rules, but it could also be
intra-lexical tone is more flexible. It does form syntagmatic relationships based on
underlying paradigmatic associations, but these are less determinant. Following the
initialization process outlined in Section 6.2 and continuing to adapt the thresholding to
each new tone, one could perpetually deduce orientation within a three-level tone system
(or two or four-level system). This is the closest approximation of frequency bands that
are appropriate to these languages. However, at any point level-tracking undermined by
lexical tone with a paradigmatic assignment consistent with the articulated syntagm but
inconsistent with the paradigmatic orientation prior to its appearance. This is the
prerogative of lexical tone.

6.3 Lexical-Order Comparisons

Lexical tone often fits within a larger framework, despite the very real possibility that it
may not. Otherwise, there would be no empirical support for paradigmatic tone levels,
and there is some. First and foremost, speakers of Yorùbá acknowledge three levels (do-
re-mi) and different speakers ascribe the same words the same tone labels (using
solmization syllables). Three levels are not a formal contrivance. It is fairly certain that a
MH word followed by another MH word will have a descending inter-lexical syntagm,
just not as sure as the lexical syntagm for both words will be ascending. Furthermore, a
distinction between lexical and inter-lexical tone is not always appropriate because of the
agglutinative nature of Niger-Congo tone languages. Compound words with elisions are
common and clitic tones may be considered somewhere between lexical and inter-lexical
tones (see Akinlabi 2000). The window of equivalence applied in Section 6.1 is much
like a frequency band, but it may shift as syntagmatic magnitude varies. Lexical tone has more of a motivation to be specific, but clitic and inter-lexical tones may also form tonemic syntagms, even after initialization. The notion that listeners are constantly segmenting the signal into words is very litero-centric. All of the participants in our study were literate in English, but Yorùbá and Ìgbò literacy is a specialized skill. A distinction between morphemes, words and compound words is sometimes difficult, and may not always be cogent. For a model of tone that can be applied more readily, without segmentation into words, the fission boundaries can be applied within a lexical-order time window, as opposed to drawing a distinction between lexical and non-lexical tones. By lexical-order, I mean local comparisons that do not extend beyond the size of words. If words are generally two to three syllables long, then a lexical-order window of comparison encompasses up to two or three segments, whether or not they are within a word or not. Hence the distinction between a lexical window of comparison (restricted to syntagms formed within the word) and lexical-order window of comparison which is the same size regardless of whether it crosses a word boundary or not.
Figure 6.8 is an illustration of how a frequency band that is limited to a lexical-order allows shifting of tone levels. Tones 1, 3 and 5 are all at the same level, but tone 5 is not within Tone 1’s frequency band. However, because Tone 3 is within Tone 1’s lexical-order window and Tone 5 is within Tone 3’s lexical-order window, the congruence is transferred. The transference of phonological equivalence is visualized as a complex of interlocking circles (Venn diagrams) in Figure 6.9.
All four tones are phonologically equivalent so long as adjacent tones share a frequency band. By restricting frequency banding to a lexical-order time window it is possible for there to be large changes in register height. Of course, people breathe, so this effect is in addition to pitch reset and reinitialization at phrase boundaries. The aggregate effect makes the astonishing intonation changes in the Yorùbá sermons in Chapter 4 conceivable.

6.4 Segmentation and Orders of Windowing

A lexical-order window is not an exact size, it is about the size of a word. A word is smaller than a phrase and a phrase is smaller than larger possible segmentations like a section, which is some significant part of the whole. Likewise, the lexical-order is smaller than the phrase-order, and the phrase-order is smaller than a section-order which is smaller than the whole. A segmentation into words is needed to distinguish lexical tone from interlexical tone. Just like the segmentation into pitches, a lexical segmentation is sensitive to the content of the signal, a lexical-order window is insensitive.

There are two primary uses of windowing into these orders. The first is to make a summary of the data within a window. Chapter 4 already saw a lot of use of summaries of phrase-order windows and section-order windows of pitch segments, including mean pitch and standard deviation within 10-segment and 100-segment windows. The second use of the temporal scales (lexical, phrase, section) is to compare $n$-order summaries within an $n+1$–order window. Possibilities include: segments compared within a lexical-order window (as in the initialization process), words compared within a phrase-order
window (e.g. Figure 6.10), and phrases compared within section-order window (e.g. Figure 6.11).

![Figure 6.10: Phrase-Order Lexical Comparisons](image1)

![Figure 6.11: Section-Order Phrase Comparisons](image2)

If I make a lexical-order comparison between two segments that are phonologically equivalent (e.g. two low tones), I expect them to share a frequency band. If I make a phrase-order comparison of two low tones, I do not expect them to share a frequency band. The segment itself is the only autonomous unit, and is therefore not an order. To go smaller than the segment is to enter the world of perceptually continuous data. The pitch segment is needed for all larger orders to exist. The segment is prototypically
paradigmatic, syntagmatically it does not exist. All of the higher orders are syntagmatic arising from the absolute values of individual segments.

6.5 Scale-Dimension Temporalities

A single dimension of sound (pitch) can have multiple roles in speech production and perception by using different temporal scales. This is sympathetic to the notion of auditory streams theorized by Albert Bregman, but is a further segregation not an analagous one. It is a stratification of activity within a single monophonic signal, which is in itself an auditory stream. I do not propose that tone and intonation are different auditory streams of the signal. Nor do I propose that tone levels are different auditory streams per se. The coming chapters offer one approach to formally distinguishing small structures, such as lexical tone, from larger structures, like paralinguistic intonation. The temporal scales introduced in this chapter with reference to pitch can be applied to other dimensions of monophonic sound, to almost any sound from a single source: a human voice, a solo cello or a lead dundún (talking drum). It is not elaborated in this dissertation, but the application of temporal scales to other dimensions of sound has the potential to synthesize two significant bodies of work within the theoretical humanities: (linguistic) autosegmental phonology (e.g. Goldsmith 1976) and (musical) contour analysis (e.g. Marvin and Laprade 1987). A tenet of autosegmental theory is that speech has multiple tiers with different phonological rates of change: the segmental tier, stress tier, timing tier and tone tier. One tier does not necessarily maintain a one-to-one correspondence with another tier, making cognitive processing of the sound non-linear.
and multi-temporal. The tiers represent four modes of linguistic contrast (phonic, timing, stress and tone) roughly corresponding to a four-dimensional paradigm of music (pitch, rhythm, dynamics and timbre), with the pitch-tone and timbre-phonic dimensions (or tiers) exchanged in the respective listings, which to some extent reflect prioritization in linguistic and music theory respectively.

A few conjectures comprise the nascent theory of stratifications based on multiple sound tiers and temporal scales expressed here. The first is monophonic sound has a distinct temporality for each dimension (e.g. musical pitch, rhythm, dynamic, timbre) and temporal scale (e.g. linguistic segment, word, phrase, section, complete signal). This implies a schema of twenty dimension-scale temporalities. However, I do not suggest all are carefully articulated or perceptually salient.

<table>
<thead>
<tr>
<th></th>
<th>Identity</th>
<th>Lexical</th>
<th>Phrase</th>
<th>Section</th>
<th>Global</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Duration</strong></td>
<td>DI</td>
<td>DL</td>
<td>DP</td>
<td>DS</td>
<td>DG</td>
</tr>
<tr>
<td><strong>Pitch</strong></td>
<td>PI</td>
<td>PL</td>
<td>PP</td>
<td>PS</td>
<td>PG</td>
</tr>
<tr>
<td><strong>Amplitude</strong></td>
<td>AI</td>
<td>AL</td>
<td>AP</td>
<td>AS</td>
<td>AG</td>
</tr>
<tr>
<td><strong>Timbre</strong></td>
<td>TI</td>
<td>TL</td>
<td>TP</td>
<td>TS</td>
<td>TG</td>
</tr>
</tbody>
</table>

A second conjecture is that relative levels of activity within each temporality reflect the focus of the sender and what an acculturated listener (of music or speech) is likely to attend to. I take an agnostic view of the salience of each dimension, temporal scale, and distinct temporality across socio-cognitive realms. Following a weak Sapir-Whorf Hypothesis (linguistic relativism), I suggest focus on specific dimension-scale temporalities characterizes a language or musical practice. One mode may have dominance, such as the phonic-segment temporality in Indo-European languages, or the
tone-segment temporality in other languages. If a dimension is not in focus within a temporal scale, it may be altered without much consequence. An example is the pitch dimension on small temporal scales (segment, word) in English. One may use a variety of contours for the same words without altering its identity. Conversely, comparison of the same phrase in different dialects of the Ìgbò language cluster in Nigeria—e.g. “They didn’t buy anything”: Fá égóró ífé (Ọnjìcà dialect); Há ázúghí íhyé (Central dialect) (Clark 1990:15)—shows significant allophonic variation can occur without distorting meaning. Speakers of different Ìgbò dialects share focus on the tone-segment temporality and this mode may be more salient than others.

The final conjecture involves comparing different analyses of the same signal: one that follows conjecture two (the sender’s focus is implied by the signal) and one that reflects general characteristics of the receiver’s primary language (which may not be the same language as the signal). Stark contrast between signal-specific (inductive) and receiver-based (deductive) analyses reflects ineffective channel coding (failed communication) between people that do not speak the same language or like the same music.

Autosegmental phonology is rich in terms of its acknowledgement of different temporal scales between phonological tiers (e.g. phones, tones), however it could be improved by acknowledging different temporal scales within a single tier.
6.6 Phonological Equivalence in Music

This chapter concludes my exegesis of phonology of African tone systems. In the rest of the dissertation, the idea of phonological equivalence is extended to music, both literally and metaphorically. In Chapters 7–11, the concept of phonological equivalence informs a modified approach to contour analysis and is applied to instrumental music of Europe and America as illustrations of its utility. Chapters 12–16 are a more literal investigation of phonological equivalence in poetic and musical realizations of linguistic tone. Artistic realization of underlying tone spans the gamut from obfuscation to fidelity that exceeds that found in most speech. In Nigerian music, the boundary between language and music is fluid. In a culture with “talking” drums, it not only makes sense but is necessary to look for phonological equivalence to understand the music. But cross-domain mappings of tone-to-tune are not at the limit of usefulness for this concept. Music and language share the same sensory pathways and many of the same cognitive processes, so it is perhaps unsurprising that the application of a phonological equivalence model to music of cultures without lexical tone, though metaphoric, is very fruitful.
§

II. Tune
7. **Contour Theory and Pitch Perception**

**Introduction**

Contour theory and analysis has a multi-disciplinary basis in music research with significant contributions from ethnomusicology,\(^\text{37}\) cognitive science,\(^\text{38}\) composition theory,\(^\text{39}\) music analysis,\(^\text{40}\) and music information retrieval.\(^\text{41}\) The contour literature on melody is unified by agreement that rises and falls in pitch are salient, but the type of comparisons made (local, global or all possible) varies widely. The promise of contour analysis is comparing melodies that are aurally gestalt but differ in interval content.

![Figure 7.1: an excerpted phrase from Schoenberg (Op. 19, No. 4) identified by Morris (1993), and a COM-matrix after Morris (1987) produced from the excerpt](image)

---


\(^{39}\) See Morris 1987.


One of the most useful tools for describing contour and measuring similarity is the *COM-matrix* introduced and formalized by Robert Morris in his 1987 book, *Composition with Pitch-Classes* (see Figure 7.1). A COM-matrix is produced by making comparisons between all pitches in the segment up to the possible *degrees of adjacency* (from 1 to $n-1$). Two segments of the same cardinality ($n$), whether found or generated, are considered equivalent if they produce the same matrix and sum to the same CSEG. If not, a calculation of similarity can be made, called *CSIM* by Marvin and Laprade (1987). The COM-matrix, CSEG classes and CSIM have limited utility in computational analysis because of the need for manual segmentation. A computer must be told where to look and what cardinality to use. The COM-matrix also has a rigid standard for equivalence (hence the need for CSIM) and comparing segments of different lengths is not accommodated. Furthermore, these tools are intended for highly varied pitch content and struggle with recurring pitches within a melodic segment, likely to be found in tonal and non-Western music. These issues, (1) segmentation, (2) equivalence, (3) cardinality sensitivity, and (4) pitch multiplicity, have motivated revisions and alternative formulations by Robert Morris (who devised the COM-matrix), David Huron, Larry Polansky, Ian Quinn and others.

### 7.1 Origins of Contour Theory

In the mid-twentieth century, two intellectual movements began to influence music analysis: Gestalt psychology and information theory. These two paradigms share an interest in pattern recognition and necessitate the formation of categories. For information
theory, the categories are very small and break up a whole into discrete features that can be polarized into one bit of information: on or off, present or absent. Conversely, Gestalt psychology involves the recognition of an entire object and fitting it into categories such as 'dog' or 'cat', 'sports car' or 'sedan', ‘fugue’ or ‘sonata’. One challenge is getting from the low-level discretized features to high-level categories. This requires figuring out which features are contrastive, and contribute most significantly to the formation of categories (such as words) and the human judgment of similarity and difference.

Intellectual pioneers include Leonard Meyer (1957), who applied information theory to music, and James Tenney (1961), who applied Gestalt psychology. Such contributions, while novel and fitting to the emerging zeitgeist of the information age, were not without precedent. Music has been recognized as an artform using modes of imitation since Aristotle. Melodic imitation is addressed as a device for expanding a musical idea in renaissance treatises, such as Zarlino’s *Istitutione harmoniche*, and is present in much of the world's music. In polyphony, voices or instruments are often mimetic, as in a fugue or an antiphonal folk chorus. In general, people create and enjoy music with considerable repetition and imitation. There are many other gestalts in music, but it is melodic gestalts that have been of primary interest to Gestalt psychologists and music theorists studying melodic variation. Innumerable transformations can be applied to pitch content, such as transposition, retrogrades, inversions, M-functions, Q-relations and so on. It is not trivial that listeners recognize a diatonically or chromatically transposed melody as similar. This was a point of interest for early Gestalt psychologists (Clampitt 1999:62–69). For music theorists, transpositional equivalence has become somewhat of a given because our standard formal models, particularly scales and
intervals, explain it. The extent to which other transformations, whether formalized or not, yield gestalts (for listeners) is less clear.

What degree of similarity (and difference) is tolerated within a category? Setting a threshold for invariance is contextual, even arbitrary or subjective. A low bar is sometimes desirable, and other times not. Self-similarity within a musical work brings cohesion and is aesthetic. If it were not, then Johann Sebastian Bach might not be so celebrated. Similarity to other musical works is a thornier issue. In oral culture interrelatedness is expected, but in literate and recorded culture intellectual property has become the primary means of capitalizing music and other scholarly and creative work. Even as the Internet challenges this status quo and the creative commons counters the notion of intellectual property, musicians generally strive to avoid being deemed derivative, while sticking to a fairly narrow aesthetic range: similar enough for people to like it, but not too similar. The legal community has adopted the term “substantial similarity” for violations of one artist's copyright by another. This is (usually) a fairly high bar. Lower than transposition, for instance, but bigger than “fair use.” When comparing two different works by different authors to make a judgment on whether one is derivative of another and there is a financial stake, our definition of invariance must be nuanced. In such a situation, melodic imitation should exclude many melodic transformations that would be considered evidence of a composer's ingenuity in other cases. In addition, other aspects of the music should be considered for similarity, such as rhythm, harmony, timbre and form. Making a judgment on melodic similarity for the purpose of establishing or clarifying intellectual property rights is highly problematic because there is so much melodic similarity in the world. Are two melodies with an arch-
shaped contour substantially similar? If so, the courts would be bogged down with copyright litigation. Huron suggests speech and musical phrases often exhibit a pattern of early-phrase rise followed by declination (an arch shape) (1996). Both “substantial similarity” and “fair use” are determined on a case-by-case basis, which is considerably less methodical than much music analysis. However, the concept of substantial similarity is a form of recognizing equivalence between things that are not empirically the same.

When analyzing or performing a piece, one generally wants to be as cognizant of self-similarity as possible. Patterns reflect cogency in the mind of the creator. In an interpretation of a score, whether analytical or performative, looking for all manner of transformations is productive.

The emergence of contour theory in the late-Twentieth century is a development of what Tenney and Polansky (1980) call temporal gestalt units (TGs). TGs for music are differentiated from spatial gestalt units (SGs) for visual art because music is a temporal art. The musical score is designed to communicate information for performance, but is also celebrated as a visual artwork in itself. Famous examples some of Baude Cordier’s Chantilly Codex compositions of the Fourteenth century and works by George Crumb in the late Twentieth century. Much music analysis has focused on the score, and extracts information visually, not aurally from listening to the musical work. It is much easier to make comparisons in this way. The recognition of melodic contours in a musical score is often a visual search, conflating Temporal Gestalts and Spatial Gestalts. However, research on perception by Dowling (1971, 1978) and Edworthy (1983, 1985) suggest formal contour similarity is a source for gestalt perception that is to some extent independent form absolute pitch height. Findings by Demany (et al 2011) suggest
detecting direction of frequency shift is autonomic aspect of aural perception. As noted in previous chapters, the musical staff has also been used to model tone levels, an application of notation that does not reflect pitch equivalence, but phonological equivalence. Adegbite’s dissertation (1978) analyzing tones of Yorùbá poetry using a three-line staff. Two tones placed on the same staff line do not necessarily have the same absolute pitch height or the same interval above tones at another level, but belong to the same phonological category. Contour theory reflects a similar leveling of pitch space so that melodic segments (of multiple pitch events) may be deemed similar or equivalent based on relative pitch relationships within each melodic segment and irregardless of absolute pitch content.

Two pitch events temporally proximal but diachronically ordered form a relationship. This relationship can be measured as a melodic interval with direction and magnitude. Deciding how much involvement magnitude should have in a theory of contour is one of the main discrepancies between scholars. A recent participant study highlights the differences between perceptual and cognitive processes that evaluate pitch relationships. According to Demany, Semal and Pressnitzer (2011), perception of direction is implicit (low-level) while perception of magnitude is explicit (high-level). The frequency-shift detectors (FSDs) for direction are an autonomic function of bottom-up perception and have fine-grained sensitivity to small changes (e.g. a just-noticeable difference, or JND). Processing of magnitude has a greater cognitive burden and is much more variable. Sensitivity of magnitude judgments are nowhere near the JND for direction, around 3 Hz below 500 Hz and 10 cents above 500 Hz (Kollmeier, Brand and Meyer 2008). Magnitude errors are on the order of semitones or even larger. Even among
musicians with similar backgrounds, the categorization of melodic intervals based on magnitude is not uniform. It is an ability that often takes many years of training to develop. In Chapter 2, different heuristics for pitch were discussed. Instead of *high and low*, the Suya of the Amazon basin speak of *young and old* (Seeger 2004) and I proposed that tension is another heuristic in a culture where variable-tension instruments are prevalent. Regardless of the cognate for pitch height, frequency-shift detection of direction of pitch change (higher or lower, greater or lesser, older or younger), seems to be a normative perceptual mechanism.

A trend in musicology coincided with the emergence of information theory and spread of Gestalt psychology: growth in the study of non-Western music and the comparison of different musical traditions to each other. One potential of contour theory is the analysis of music that is not scored, this was particularly useful before computer-assisted analysis was widely available. Charles Seeger can be credited with the first methodology for contour analysis, although he did not call it contour. In “On the Moods of a Music-logic” (1960), Seeger’s goal is to find evidence of “intrinsic order” in music be it “folk, popular or art.” Seeger draws on Meyer's work on information theory and music (1957) and historical thought, referencing Aristotle's *Poetics* and using terminology associated with early chant notation (neumes). Seeger focuses on direction and not magnitude. His moods are small units of two sizes: binary (three elements with two relationships) and ternary (four elements with three relationships). The comparison between two events is between their respective tensions. This is a relationship of greater or smaller (increase or decrease), which can be extended to pitch or duration. An increase in tension for pitch is an increase in frequency. An increase in tension for duration is an
increase in duration. The relationship to neumes is through the use of the terminology *torculus* (for an increase, then decrease in tension) and *porrectus* (for a decrease, then increase of tension). The plus sign $[+]$ is an increase in tension, the minus sign $[-]$ is detension and invariance, termed tonicity, is represented by the $[\equiv]$ equal sign. Like many later explications of contour, Seeger ignores repeated pitches (or durations) in his “moods”. There are 4 binary moods ($2^2$) and 8 ternary moods ($2^3$) for pitch and duration. Combining ternary pitch and duration moods creates 64 ($8^2$) aggregate moods. Another important concept in Seeger’s article is that of centricity: a mood that returns to its starting point is centric, one that goes away from the starting point is decentric. A layer of abstraction is added with the variables delta $[\delta]$ and omicron $[\omega]$. Delta is assigned to the first tension relationship $[+\ or\ -]$ and the omicron the opposite tension $[-\ or\ +]$ and they are kept constant within a mood. This framework makes it possible to recognize inversions and some retrogrades as equivalent. A binary torculus $[+\ -]$ and, its inversion, a binary porrectus $[-\ +]$ have the same reduction $[\delta\ \omega]$, as are their retrogrades, $[\omega\ \delta]$ would become $[\delta\ \omega]$. Likewise, a ternary torculus-porrectus $[+\ -\ +]$ and, its inversion, a porrectus-torculus $[-\ +\ -]$ are the same $[\delta\ \omega\ \delta]$, as are their retrogrades. Both are pallindromes. The binary interpretations of both tension ($+/\ -$) and appearance order within the mood ($\delta/\ \omega$) reflect the information theory paradigm.

Kolinski (1965) has a similar goal and follows a similar methodology. Kolinski also wants an analytical method that can be applied to a diverse corpus of music. His units are larger, extending beyond a cardinality of four, and the units are still combinatorial. For example, Kolinski calls the strict alternation of $[+]$ and $[-]$ a pendulum, and differentiates between returning and progressing types (similar to Seeger's
centric and decentric moods). While Seeger explored the combination of a pitch mood with duration mood, Kolinski restricts his analytical method to the combination of melodic structures through nexuses that are direct (overlapping) and indirect (contiguous but not overlapping). A founding member of the Society for Ethnomusicology, Kolinski spent a good deal of his career transcribing field recordings gathered by himself and others. In his article, he compares a melody from Benin (a Dahomean song) and from British Columbia (a Kwakiutl song).

The range of the song is the same as that of the preceding one; however, within this similar framework the melodic line of the Dahomean song is bold and ragged, contrary to the smoothness of the Kwakiutl song... the melody descends from the highest tone a minor tenth down to the lowest one and then rises again an octave to the second highest tone (Kolinski 1965:116).
In Figure 7.2, Kolinski’s structural units do not convey magnitude information in themselves. The similar range of the two pieces is gleaned from the pitch values of the y-axis. Much like the higher proportion of turning points (or pivots) observed in Yorùbá in Chapter 4, Kolinski notes a prevalence of “non-recurrent one-member” movements in the Dahomean song. Or, in Seeger's terms, alternations between delta and omicron. His description betrays a European sensibility. Angularity is “bold and ragged“ in comparison to the smoothness of the Kwakiutl song. What is lacking in the discussion is consideration that the language of the Kingdom of Dahomey is Fon, a two-level tone language with tone levels of high and low. A reasonable conjecture is that the angularity in the music reflects angularity in speech, either through a mapping or aesthetic affinity. A tone-to-tune mapping cannot be assessed because the text is not present in Kolinski's analysis. A general aesthetic of angular melodies may reflect angular speech, much like the high pairwise variability (nPVI) in duration found for English speech and music by Patel and Daniele (2003). Similarity of speech and music profiles in Niger-Congo cultures will be addressed further in Chapters 12–16.

Cognitive and behavioral scientists were the next to take up the issue of melodic contour, producing a lively stream of articles form 1971 to 1985. W. J. Dowling is among the first and most prolific investigators of the perception of melodic contour similarity. His research questions were motivated by interest in a variety of music, from the similarity of subjects and answers in Bach fugues, to tune recognition of folk songs. Two important aspects of Dowling's methodology were the definition of contour as a series of ups and downs (directional relationships without magnitude) and the use of forced-choice (binary or polar response) experiments. For example, participants might be asked to
evaluate a two melodies as the same or different. Initially, Dowling asked participants to compare melodies based on folksongs. Later studies by himself and others used novel melodies because of concern long-term familiarity might effect results. In the work of Dowling, his colleagues and others, controlling variables to distinguish effects of direction, magnitude, chroma and tonality is central to forming conclusions. The results of Dowling’s initial experiments at UCLA in 1971 have largely been reinforced by later experiments: in the absence of an established key, people evaluate similarity and difference based on contour (the series of ups and downs in a melody). Some additional nuance to the perception of melodic contours is added through experiments by Judy Edworthy (1982, 1985). Edworthy found that in longer sequences, once a tonal framework is established, the interval information (magnitude of change) becomes more important to judgments of invariance. Edworthy writes that “contour can be encoded independently of tonal extent“ (1985:375), a finding consistent with both Dowling’s articles on the subject and with Demany, Semal and Pressnitzer (2011).

In 1985, the psychomusicology community largely rested on investigating contour perception and music theory became engaged. Friedmann’s first article on the subject (1985) is highly disciplinarian and does not cite any of the articles from comparative musicology or music psychology, which accounts for some differences about the importance of magnitude (interval) in contour space. His basic framework is similar though developed in apparent isolation. This speaks to the intuitiveness of contour analysis. Many scholars have come to it independently. Friedmann's contour adjacency series (CAS) is a series of adjacent pairwise comparisons of pitch height, consistent with Seeger’s tensions and Dowling’s concept of contour. Friedmann does not have an interest
in comparing a diverse range of music, but adding a complementary layer of analysis to
dense post-tonal music. In doing so, Friedmann introduces a perceptually salient aspect to
the analysis of music that is largely bewildering to the ear.

Unlike Friedmann, Marvin and Laprade’s version of contour analysis published
two years later (1987), actively engages with several of Dowling’s articles. The same
year, Friedmann’s response (1987) outlines the similarities and differences between
Marvin and Laprade’s methodology and his own. Marvin and Laprade’s CSEG and
Friedmann’s Contour class (CC) are identical equivalence classes. Both maintain the
time sequence of a contour and number relative pitch heights from 0 to n-1, where n is
cardinality and 0 is the lowest pitch. In <03421>, n=5, the first pitch event is the lowest
and the third (medial) pitch event is the highest. Both methods explore aggregate
properties and subset relations beyond the injection function that produces a CSEG or
CC. To calculate similarity for segments that are not equivalent as CSEGs or CCs, ML
use the COM-matrix (comparison matrix after Morris 1987). Friedmann uses the CIA
(Contour Interval Array) and summations of it: CCV I (Contour Class Vector I) and CCV
II (Contour Class Vector II). The COM-matrix used by Marvin and Laprade compares all
INTs. INT1 is equivalent to Friedmann's Contour Adjacency Series (CAS). Further INTs
compare non-adjacent events: INT2 compares events with one intervening pitch, INT3
events separated by two events, and so forth. The primary conceptual difference is F's
emphasis of the contour interval. The contour interval measures contour space in terms of
vertical adjacency (in order of pitch heights from lowest to highest in a contour). The
maximum contour interval within a contour of cardinality n would be n-1. According to
Friedmann, Marvin and Laprade reject the contour interval because of a “counter-
intuitive contradiction of pitch-space“ (1987:270). Friedmann acknowledges that a large contour interval could correspond to a small pitch interval and vice-versa.\(^{42}\) This is similar to the concept of a contradiction between species of generic intervals in Scale Theory. This distortion of pitch space in contour space (the “pitch-interval flattening power”) is valuable to Friedmann and he relates it to a “renaissance-like concept of line and mode” (1987:270).

Both Morris (1987) and Marvin and Laprade (1987) go beyond adjacent comparisons into complex contour, venturing into theoretical territory that does not have empirical evidence from behavioral studies to support it. To say a contour pitch has a value of 2 because it is above the minimum (0) and above one other pitch (1) in the segment, is not so different from formulating a contour interval of +2. So, Marvin and Laprade are not so far off from Friedmann even though they reject the contour interval as a construct. And, contrary to Friedmann’s distinction between the two methodologies, both have flattened pitch-space. The distance between vertically adjacent c-pitches in c-space is unspecified. Friedmann layers abstraction with contour interval vectors (maintaining kinship with set theory), and Marvin and Laprade restrict themselves to calculations using the COM-matrix (1987:228), but Schmuckler finds that both methodologies produce similar results:

Although rigorous and systematic, this focus on the interval content or contour subset information neglects a crucial aspect of melodic contour namely, the global shape of the contour. Although global contour shape can be discerned from these representations (imagine the CC or CSEG as line drawings), neither model adequately characterizes this information. (1999:300)

\(^{42}\) See Marvin and Laprade 1987:228 for their discussion of contour intervals.
Schmuckler tested both Friedmann's CIVI and CIVII and Marvin and Laprade’s CSIM and CSEMB reductions and found that neither significantly predicted perceived similarity. Although they may indeed inspire composers (especially after Morris’s treatise (1987)), these abstracted measurements of the contour do not seem to reflect features that influence perception. In line with the results of this study, Schmuckler and others (e.g. Eerola et al 2007) have focused on global measures of contour such as mean and pitch extremes (maxima and minima), following methods such as Morris (1993) and Huron (1996).

Morris’s contour reduction algorithm (1993) provides a much less rigid form of equivalence and fully accommodates comparison of different cardinalities of contour segments. Through pruning all but time and height extreme contour pitches, CSEGs are reduced to a prime of 2, 3 or 4 elements. However, there is still dependence on segmentation and segments reduced to primes lose characteristics that may be important. Complexity is tracked through the assignment of a depth based on the number of passes necessary to reach prime form. Despite the critical stance in this dissertation, Morris’s algorithm continues to be a mainstay in music theory. It is central to recent dissertations by Schultz (2009), Bor (2009) and Sekula (2014) and applied to rubato by Ohriner (2012).

A concurrent body of work on contour comes from computational (or systematic) musicology, and more recently music information retrieval (MIR). Huron offers operational concepts of similarity, not restricted to contour, that make it possible to “characterize degrees of resemblance” beyond “absolute matches” (2002, 21). The Humdrum Toolkit includes two commands that can be used to calculate similarity between melodic contour segments, correl and simil. The user has the prerogative to set
constraints on pattern matching, including assigning a penalty to deletions and insertions that make one object more like another (1994). Another method implemented by Huron in Humdrum, is a reduction based on initial pitch, final pitch and the mean pitch inbetween, producing nine varieties of three-point gross shapes (1996).\footnote{These are: ascending, descending, concave, convex, horizontal-ascending, horizontal-descending, ascending-horizontal, descending-horizontal, horizontal (Huron 1996, 9).} The reduction of contour slices introduced in Section 7 of this article is more sympathetic to calculating an edit distance between segments of cardinality 10 and 11 (as in the simil command in Humdrum) than reducing all segments under consideration to primes (Morris 1993) or gross shapes (Huron 1996).

Polansky’s “Morphological Metrics” (1996) is prolific in its formalization of distance (or similarity) measures, and responds to the challenge of comparing different length “morphs”. Two ideas from Polansky are particularly relevant to this article. First, the notion of “memory decay” motivates a weighting of combinatorial direction (1996:336), discussed in Section 7.2. Polansky also applies a moving window to an unsegmented pitch series to measure contour, an innovation sustained in many computational approaches to contour, including this one. Aside from important formal differences, I differ from Polansky in what I interpret as an agnostic view of segmentation, specifically “avoidance of the concept of inclusion” (1996:293). Whether elements from an entire structure (melody) also belong to substructures (melodic segments) is an important question.

Another thread in contour theory, and an interesting twist on the COM-matrix, appears in Quinn (1997, 1999). Quinn opts for binary “C+” contour ((1) ascending or (0)
non-ascending) instead of ternary contour [+ 0 -] for filling matrices. This modification allows Quinn to calculate fuzzy values for cells. Quinn writes:

To find the essence of contour is tricky because there are so many ways of notating contour. Pictures, contour-pitches, and COM (comparison) matrices come immediately to mind as candidates. None of these modes of representation, however, captures the essence of contour as simply and elegantly as does one simple relation: ascent (Quinn 1997, 248).

In this dissertation, binary ascent (and by extension the C+ matrix) is also adopted for simplicity and elegance, but not primarily for the purpose of averaging contour comparison matrices to create fuzzy categories of belonging as Quinn ultimately does in his 1997 article. Here the binary categories make the techniques developed here for symbolic music (MIDI data) extensible to recorded music, where qualifying what is and is not the same pitch is more challenging.44

Quinn offers two methods for representing variation between contours of the same cardinality. One is the fuzzy C+ matrix, which makes it possible to represent contours with similar but not equivalent matrices with a single matrix. The other is C+SIM. Unlike Marvin and Laprade’s CSIM, which uses only the upper-right corner, Quinn uses the whole matrix (save the central diagonal which neither methods uses). This is necessary because of the use of binary instead of ternary contour. The number of comparisons is doubled and so are the number similarities and differences, so it produces the same percentage in the end, as Quinn demonstrates.45

44 This avoids declaring and substantiating a just-noticeable difference to form three categories in audio recordings.
45 He suggests that calculating C-SIM (note the -) would yield the same results.
Like Schmuckler’s test of various models (1999), Quinn tested C+SIM as a predictor of participants’ judgment of similarity, but found that it “may not tell the whole story“ (1999:454). Although not as grounded in empirical evidence as adjacency, Quinn’s emphasis on the salience of non-adjacency is important.

7.2 Degrees of Adjacency and Perception

Within the music theory literature on contour, there are generally two factors in making pairwise comparisons between notes: (1) which to compare and (2) how to compare them. The contour adjacency series is a note-to-note model (comparing only adjacent notes) and using ternary direction [+,-]. This is the model addressed by a number of perceptual studies (e.g. Dowling and Fujitani 1971, Edworthy 1982). Morris (1987) and Marvin and Laprade (1987) formalize an exhaustive combinatorial model, in which every note is compared to every other note, also using ternary direction (see Figure 1). For detailed discussion of both models see Quinn (1999).
second factor, how to compare pitch height, begs the question, what is contour? Is it direction? Or, is it direction and magnitude? Polansky and Bassein introduce the concept of non-ternary (n-ary) contour values, exemplified by the quintary categories: “a lot less than, less than, equal to, greater than, and a lot greater than” (1992:277–8). Similar meta-intervals (with both direction and magnitude) have been proposed for tone languages and used by the refined contour search parameter of Huron’s ThemeFinder algorithm (1999). Though the magnitude component of refined (or n-ary) contour is generic, a case can be made for avoiding magnitude altogether in pitch height comparisons. Frequency ratio and tonal context strongly influence the perception of magnitude, but not direction. Demany, Semal and Pressnitzer (2011) provide experimental evidence that judgment of direction (as in ternary contour) is nearly autonomic while judgment of magnitude is a higher-order cognitive process. In general, beyond pitch height comparisons, unidimensional judgment of magnitude is difficult (Donkin et al 2015). In Marvin and Laprade (1987) contour intervals arise from multiple degrees of adjacency, not incorporating generic magnitude. Ian Quinn (1997) takes n-ary contour in the opposite (smaller) direction, with binary C+ ascent, but ternary categories can still be deduced by making pairwise comparisons in both directions (see Section 4). Because the judgment of magnitude is highly contextual, variable across cultures and between people, my position is that magnitude should be left out of pairwise contour comparisons. Choosing which notes to compare, the degrees of adjacency, is a harder decision.

My interest lies in identifying melodic patterns intentionally developed by musicians and/or salient to attentive listeners. Where the boundary lies between what is aurally apparent (e.g. a melodic sequence) and what is beneath the surface (e.g. an urlinie) is not known. Perceptual studies from the 1970s and 1980s suggest contour is the most prominent and memorable aspect of novel or transposed melodies outside of or prior to the establishment of a tonal paradigm. Performance of perceptual judgment tasks decreases as the length of melody increases from three to seven pitches and rapidly erodes beyond lengths of nine pitches (Edworthy 1985:383). Fewer empirical studies have addressed perception of non-adjacent pairwise comparisons. A study by Quinn suggests melodies with adjacent and non-adjacent equivalence are more likely to be categorized as similar by listeners than melodies that only share a contour adjacency series (1999:454). Edworthy’s participants struggled to retain contours of cardinality nine, so it seems reasonable to argue that eight degrees of adjacency greater, as found in a COM-matrix for a nine-pitch segment, are beyond the cognitive grasp of most listeners.

The maximum number of degrees of adjacency for a pitch series of cardinality \( n \) is \( n-1 \). For COM-matrices to be applied in analysis, phrase boundaries are needed to break up entire pitch series into segments. Using \( n-1 \) degrees of adjacency for an entire piece is unwieldy, even if it is short. Furthermore, the COM-matrix sets a high bar for formal equivalence that fails to recognize highly similar contours, such as the subject and tonal answer of Bach’s C-minor Fugue (BWV 847), as the same. The alternative

48 A COM-matrix models complex adjacency because it involves both adjacent and non-adjacent comparisons.

49 In an application of the segmentation algorithm to Bach’s C-minor Fugue (BWV 847), it was found that the subject and its answers are only equivalent up to 10 degrees of adjacency (adjacency radius of 5). The cardinality is 20, but all possible degrees of adjacency (19) are not relevant to determining that the subject
presented here is to commence contour analysis without any prior segmentation based on
independent criteria, and lower the bar for formal equivalence by restricting degrees of
adjacency to a constant window size. If analyzing a piece with a melody of 100
consecutive pitches, potentially 99 degrees of adjacency can be used. The fact the melody
starts on a higher note than it ends may be an important and curated detail of a musical
work that many listeners catch. However, what about the third note and the ninty-sixth
note (a comparison at the ninety-third degree of adjacency)? To apply contour analysis to
un-segmented music, some limits on degrees of adjacency must be imposed, even if at
first, they are somewhat arbitrary.

Quinn (1999:454) found participants were more likely to judge contours as the
same if they had equivalence beyond note-to-note comparisons. However, the effect on
similarity judgments was less impressive than one might expect and it is not known to
what degree of adjacency listeners attend. Our sensitivity is likely less than \( n - 1 \) (the
maximum for any cardinality \( n \)) in most contexts. Without prior knowledge, a listener
does not know what length of segment to listen for, unless clearly punctuated in time. A
COM-matrix for a 12–pitch segment holds 66 contour comparisons:

\[
\frac{n^2 - n}{2}
\]

The extreme level of adjacency (\( n - 1 \)) may hold special prominence if heard in isolation
from the rest of the piece, but it is doubtful that all degrees of adjacency hold the same
sway over perception. Polansky (1996:337) suggests a weighting for each degree of

and its answers have highly similar, even equivalent, contours. Findings on contour recursion in tonal
music (including imitative polyphony and jazz improvisation) are addressed in my dissertation and will be
published in separate articles.
adjacency. Quinn’s study results are consistent with this (1999:453), but did not operationalize a test for the weighting of degrees.

The decrease in performance beyond a cardinality of nine found by Edworthy (1985:383), bears some resemblance to Miller’s rule for mental processing of information: working memory can hold “seven plus or minus two” objects (1956). The heuristic can only be applied loosely here because the same listener may remember pitch information differently depending on context. For a non-AP listener acculturated to a tonal system, once a tonal context is in effect, she likely uses the same referent (tonic) for all pitches in a series. In the absence of a tonal paradigm or before it has been established, this cannot be the case. Whether holding nine notes in terms of a tonal referent or eight directions between nine pitches in a contour adjacency series, a cardinality of nine notes is close to Miller’s upper extent of nine objects if each note corresponds to one memory bin. The extent to which further degrees of adjacency make a contour model more or less like echoic memory of melody can only be conjectured here. Using ternary contour categories [+ , 0 , –], a segment of cardinality nine has 15 pairwise comparisons within two degrees of adjacency (n-1 + n-2) and 21 for three degrees of adjacency (n-1 + n-2 + n-3).50 If each cell of a COM-matrix is an object, then Miller’s upper limit of nine memory bins is reached rather quickly (see Figure 7.4).

50 This can be generalized as: (n-1)+(n-2)+…+(n-k) for some k, 1<k<n.
The number of unique comparisons for \( n-1 \) degrees of adjacency reaches 10 at a cardinality of \( n=5 \). However, it is not the pairwise comparisons that correspond to Miller’s memory bins. The contour comparisons (though ternary in the COM-matrix) are closer to bit values that encode a description of the object, in this case, the pitch.\(^{51}\) Miller uses information theory to address auditory perception (pitch and loudness) in addition to other modes of perception including vision. For pitch, Miller interprets results from Pollack’s study of pitch memory (1952), wherein participants were asked to remember a collection of numbered pitches and then respond to pitch stimuli with the corresponding number. Participants tended to make identification errors when primed with collections of six or more pitches. Based on this, Pollack made a calculation of human “channel capacity” for pitch objects: a bit depth of 2.3 (1952:748).\(^{52}\) While both Pollack and Miller define it as an absolute judgment task\(^ {53}\), it is more likely a judgment of relative pitch for most listeners. Participants were primed for the task in Pollack’s study by hearing a pitch collection in series from low to high, so the ordinal number of the pitch is also its relative height within the series.

\(^{51}\) Both Miller (1956) and Pollack (1952) use the term *tone* instead of *pitch*.
\(^{52}\) A bit depth of 2 would yield four objects and 3, nine objects. Bit depth is converted to unique values by taking 2 to the power of the bit depth (e.g. \( 2^x \), where \( x \) is the bit depth).
\(^{53}\) In current psychology literature, this is referred to as absolute identification.
Let us consider the cognitive process of Pollack’s participants, however speculative. First, the primer needs to be memorized and, at first, let us presume that is done in chronological (and ascending) order. Figure 2b leaves the channel coding open, but reflects the upper limit of 6 pitches. Non-AP listeners would contextualize the series based on pitch relationships therein and not recognizing absolute frequency. Out of a tonal context, recognizing super-particular ratios would be useful for trained musicians, but not for others, and such ratios were not present in the series (an equal logarithmic spacing of pitches ranging from 100 to 8000 Hz). In a tonal context, scale degrees or solmization would be quite effective, but that is not possible with this task either. A series of six pitches evenly distributed across this range is quite spread out. Each pitch would produce a distinct sensation because of the wide spacing and dispersion of corresponding sensitivity on the basilar membrane. Two pitches spread across this range are easily categorized as high and low, a distinction that perhaps can be based purely on sensation. For four, we could add categories of mid-high and mid-low (categories of tone level often used in Autosegmental Phonology).

54 Here channel coding refers to the categorization of perceived pitch in echoic memory, and bypasses the rather complex question of how a physical signal is converted to a mental impulse. See de Cheveigne (2005) for models of pitch perception from Helmholtz to more recent scholarship. Clearly, the bit depth of 2 to 3 of Pollack (reiterated by Miller) is not intended to encapsulate the complexity of pitch perception, but simply the information necessary to differentiate like sounds that differ only in terms of pitch height.
At some cardinality, added memory bins are no longer discrete and the objects within them tend to be confused. Pollack found this effect at six pitches. However, Pollack found inaccuracy did not tend to occur in judgments of the highest pitch stimuli, as it appears in Figure 7.5. Once the cardinality exceeded five, middle pitches were associated with greater error than the extremes (low and high) of the series.\textsuperscript{55} The underspecification and low-salience of medial pitch events is also noted in tone language speech (Connell 2012). Presumably, the bit depth of 2.3 represents a conversion of the decimal integer six to a binary integer 110 (8 integers requires a bit depth of 3, 4 integers a bit depth of 2).

\textsuperscript{55} Another possible cause for error in the absolute identification task is the number of stimuli between the prime and current stimulus. Recent studies (Matthews and Stewart 2009; Donkin et al. 2011) have explored sequential effects.
This would imply a loss of information at six and above. However, the loss is not at the extremes, it is in the middle. Hence, the first bit, should differentiate the lowest pitch and the highest pitch, and added bits can specify relative pitch heights in-between.

Figure 7.6 revises Figure 7.5 to reflect an encoding of the pitches in the series based on relative pitch height and a merging of the middle bins into a single bin. In the bit encoding, the first logical represents low (0) or high (1), a second binary value is added if the pitch is below (0) or above (1) the extreme represented by the first value. We could continue adding binary values as we move further and further towards the middle, but
already by the third, we have surpassed Pollack’s proposed bit depth of 2.3. In the study, response time was unrestricted, so a possible cognitive process is to compare the stimulus with memory of the primer. As a participant, I would compare the stimulus to the primed low-to-high pitch series to find a match, then reconfirm the placement of that pitch in the series by comparing it to the others, e.g. “this sounds like the note that was immediately higher than the lowest note, which places it second in a series from low-to-high.”

Matching the stimulus to a pitch from the primer may be an absolute frequency identification using associative memory, but assigning the stimulus a number from the series is a relative frequency comparison using syntagmatic features, and most likely does not involve magnitude (interval size) so much as direction (contour).56

Alternatives to Miller’s magic number have been proposed, usually smaller. Points on a line, color, pitch, loudness are all categorized as objects on a uni-dimensional continuum in Miller’s paper, but they are not uniform in terms of neural pathways or processes. This may explain some of the variation in channel capacity between the modes presented in Miller’s article. For instance, a bit depth of 3.2 is calculated for points on a line (which is right around nine objects). If channel coding has relevance to pitch perception, binary comparisons of pitch height (1 for higher, 0 for equal or lower), as suggested by Quinn (1997), are a possibility for encoding pitch syntagmatically. For channel coding of fully combinatorial contour, as in a C+ matrix, each degree of adjacency requires a bit (see Figure 7.7).

56 Though not in reference to Pollack’s study or a discussion non-adjacent comparisons, Lartillot (2004, 58–60) provides a similar account of listening strategies.
A C+ matrix for three notes requires two bits, four notes requires three bits, five notes four bits, and so on. Using C+ ascent with n-1 degrees of adjacency is not as economical in terms of bits as the encoding shown in Figure 7.6, but it is more robust (the encoding in Figure 7.6 is intended for a distinctly non-musical pitch series from low to high). If Pollack’s and Miller’s application of information theory has any bearing on short-term memory of contour, and if Polansky’s notion of “memory decay” is valid (as Quinn’s study results suggest): in continuous (un-segmented) music, local degrees of adjacency beyond three may not be salient and therefore may not be relevant to an analytical technique based on normative perception (as contour analysis is).57

57 Marvin and Laprade (1987) cite psychological research extensively.
8. CONTOUR LEVELS

Introduction

Contour level analysis is an extension of prior contour theory (e.g. Morris, Marvin and Laprade, Polansky, and Quinn) not challenged by pitch multiplicity and cardinality sensitivity. Based on the phonological equivalence of tone levels, a contour level is a measure of relative height within a moving window of comparison, making it possible to handle diatonic and non-diatonic music, and see equivalence in a new way. Summing binary values for pairwise comparisons of pitch height within the window produces a quasi-paradigmatic value for every element in a pitch series. The level of abstraction produced is fundamentally different from contour space. Contour levels are not contour pitches (see Marvin and Laprade 1987). Contour pitches order all pitch content from low to high. If a melody 20 notes long is under consideration and every pitch is unique, they will be numbered 0 to 19. If the same melody is reduced to contour levels using four degrees of adjacency, each pitch will be assigned to one of five levels, 0 to 4 (lowest to highest). The compression of pitch space is determined by the number of degrees of adjacency used and is unaffected by pitch uniqueness or multiplicity. Whether there are tens of elements or thousands, whether it spans octaves or three adjacent notes in a scale.

58 In formal mathematical language the “pitch series” is an ordered multiset. “Ordered” in time and “multi-” because multiplicities are allowed. Each item does not need to be unique for the analysis to work. This is key to its effectiveness for tonal music.
(e.g. do-re-mi), if there are four degrees in the window, there will be five levels. If there are two degrees, then three levels (the number of degrees plus one). The analyst has much more control than in the full combinatorial model. Because this is a revision of contour analysis based on Niger-Congo tone systems, in the spirit of Kofi Agawu’s 2015 keynote address to the Society for Music Theory, this is a rethinking of music theory with African aid.

An attempt, however futile, has been made to be cognizant of the breadth of empirical and theoretical work on contour. Contour levels have more detail than a contour adjacency series and more flexibility than a COM-matrix. The method is equally effective for tonal and post-tonal music, Western and non-Western music. One can formally recognize melodic segments of different cardinalities and different numbers of unique pitches as similar, and without using fuzzy values, so one can also speak of equivalence. Unlike conventional contour matrices, the windowing enables computational approaches that can applied to an entire signal without manual segmentation (see Chapter 9). In this way, contour levels are also a theory of musical organization and form, not just melodic gestalts. Contour level space is both an abstraction of pitch space and a distortion of it. This method can produce strong analyses that balance descriptive and explanatory adequacy. It can also produce weak analyses with arbitrary information that obfuscates analytical goals. This chapter is a primer on how to achieve the former and avoid the latter.

To illustrate the utility of this system for music analysis, in addition to continued reference to tone language speech, it is applied to two pieces previously addressed in the literature: Schoenberg’s Opus 19, No. 4 (1913, analyzed in Morris 1993) and Johann
Sebastian Bach’s C-minor Fugue (BWV 847) from the Well-Tempered Clavier Book I (1722). First, I turn to the subject of the fugue addressed in Dowling and Fujitani (1971).

Figure 8.1 includes a contour adjacency series (CAS) for the first presentation of the C-minor fugue subject. A CAS makes only pairwise comparisons, there are no non-adjacent comparisons. For the 20 notes in this melody, there are 19 (n-1) contour comparisons. First formalized by Friedmann (1985), the CAS has antecedents in Seeger (1960), Kolinski (1965) and Dowling’s work in the 1970s. Citing Tovey (1957), Dowling and Fujitani (1971) point out that the relationship between the subject and answer is not exact transposition. The tonal answer is an “adaptation” of the subject.

In terms of adjacent pairwise comparisons of direction (i.e. a CAS) the two are equivalent, but they suggest the similarity goes beyond mere pairwise direction though it does not extend to precise distances in steps or semitones, as in transposition. For the first five notes of subject and answer, Dowling and Fujitani use pairwise comparisons of
interval magnitude for a more detailed description, not unlike the application of nPVI to pitch interval magnitude in Chapter 4.

<table>
<thead>
<tr>
<th></th>
<th>Subject</th>
<th>Tonal Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Semitone Intervals</td>
<td>–1 , +1 , –5 , +1</td>
<td>–1 , +1 , –7 , +3</td>
</tr>
<tr>
<td>Directions</td>
<td>– , + , – , +</td>
<td>– , + , – , +</td>
</tr>
<tr>
<td>Magnitudes</td>
<td>1 , 1 , 5 , 1</td>
<td>1 , 1 , 7 , 3</td>
</tr>
<tr>
<td>Pairwise comparisons</td>
<td>= , &lt; , &gt;</td>
<td>= , &lt; , &gt;</td>
</tr>
</tbody>
</table>

Table 8.1: Analysis of subject sug-segment by Dowling and Fujitani (1971:XX)

The adjacency interval series of the subject is not shared by the tonal answer (at the top of Table 8.1). However, the comparison of adjacent interval magnitudes yields the same result. Equivalence relations between adjacent interval magnitudes have not been extended to other analyses, but they nicely capture the balance between too much and too little information.

<table>
<thead>
<tr>
<th></th>
<th>Subject</th>
<th>Tonal Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generic Intervals</td>
<td>–1 , +1 , –3 , +1</td>
<td>–1 , +1 , –4 , +2</td>
</tr>
<tr>
<td>Pairwise comparisons</td>
<td>= , &lt; , &gt;</td>
<td>= , &lt; , &gt;</td>
</tr>
</tbody>
</table>

Table 8.2: Interval series in generic intervals

Dowling's comparison of semitone distance confirms it is not a chromatic transposition. It is the magnitude of the third and fourth intervals in generical intervals that make the answer in Figure 8.2 “tonal” (adapted) instead of “real” (diatonically transposed, see Table 8.2). Dowling is correct that the discrepancy occurs in the first five notes, however, it is not until we consider the entire subject and entire tonal answer (cardinality 20) that this contour poses a problem to COM-matrices. The CSEGs for the first five notes of each are exactly the same (83201). However, the complete subject and tonal answer do
not belong to the same CSEG class. The contours are not equivalent using conventional contour analysis because the tonal answer has a C5 instead of a D5 for its fourth note. This chapter is about that note.

The interest of Morris (1987) and Marvin and Laprade (1987) is in post-tonal music in which recurring pitches are not found in close proximity. The subject of the C-minor fugue includes lots of repetition of pitches, as is characteristic of common practice music. This causes some problems for the COM-matrix and its summation, the CSEG, which will be discussed further. But the greater issue in this case is that the fugue subject and the tonal answer do not produce the same COM-matrix, precisely because of the variation in interval magnitude in the first five notes that Dowling and Fujitani highlight. No informed analyst or listener would question that these are very similar melodies, and that similarity resides in the contour. It would be preferable to have a more detailed analysis object than the contour adjacency series that describes both the Fugue subject and its answer. There must be some way to express this similarity. Dowling and Fujitani's pairwise comparisons of magnitude offer this, but the elegance of contour theory is that it is constrained to directional comparisons. There is much more evidence for the perceptual salience of direction than magnitude. Quinn’s C+ matrix with fuzzy values (1997) makes it possible to produce a single analysis object to describe similar contours that do not form identical COM-matrices as in the fugue subject and answer. However, the fuzzy values confound the summation of the matrix into a CSEG.

Much like speech contours in tone level languages, musical contours are treated with transformations other than transposition (see Chapter 10). The tonal answer is one such instance. The difference between the tonal answer and a diatonic transposition is so
nuanced: the third interval is down four steps (a fifth) instead of down three steps (a fourth). This in turn alters the next interval. Thereafter, all intervals are true to a diatonic transposition. Contour level analysis addresses this problem by restricting the degrees of adjacency.

### 8.1 Windowing Contour Matrices

Imagine a gigantic matrix for all of Owadayo’s second Yorùbá sermon (analyzed extensively in Chapter 4). The matrix includes contour comparisons for over 1700 segments and is produced using Quinn’s binary C+ contour, wherein if a pitch is higher than the referent, the value is 1 and if it is equal or lower, the value is 0. In all, there would be approximately three million pairwise comparisons.
The central diagonal is the identity. The diagonals adjacent to the identity are similar to the contour adjacency series (or Marvin and Laprade’s Int+1 and Int–1), but must be used in tandem to have the complete information about direction (because binary C+ contour is used). All pairwise comparisons to the initial pitch (as referrent) are found in the top row and to the final pitch (as referrent) on the bottom row. The global maximum and minimum are found by summing entire columns. If there are no recurring pitches (which is not the case with the semitone values used here), the maximum contour pitch will be the cardinality minus one, and there will be one minimum with a sum of zero. The full matrix can be divided into smaller blocks. Figure 8.3 shows an initial block and a final block. Figure 8.5 magnifies the initial block.
As long as a block is aligned to the central diagonal, it is a complete COM-matrix in itself. The central diagonal is still the identity and initial, final, maxima and minima within the block are found as described above. The initial section block in Figure 8.4 encompasses the first three phrases and the first 53 segments of the sermon. The average size of the phrases in this section of the sermon is 17.67 segments (53/3), this is a phrase-order temporal scale. The first phrase has five words spread across 11 segments. The average size of the words within the first phrase is 2.2 segments (11/5), this is a lexical-order temporal scale. In Chapter 6, I proposed tone levels adhere to frequency bands within a lexical-order temporal scale (2 to 3 segments), but this is not so for a phrase-order temporal scale (10+ segments). Contour levels are a variation on this theme. The rationale for contour levels is that the most salient contour comparisons are within a lexical-order window of adjacency. Instead of comparing every element to every other element (n-1 degrees of adjacency), each element (a focus) is compared to other but not all elements (referents) using the same window of degrees of adjacency.
Going back to the huge matrix of the entire sermon. Possible window sizes range from 1 to the cardinality of the pitch series minus one (n-1). A small window of just a few degrees of adjacency corresponds to the lexical-order and produces contour levels that approximate the phonological equivalence of tone levels with a formal mathematical construct. A slightly bigger window, around 10 degrees of adjacency, is a phrase-order comparison of segments. The section order is the most flexible, it could include a few phrases or as many as appropriate. Finally, a conventional matrix is global with n–1 comparisons for each event. No two elements are compared using the same hindsight or foresight (degrees of adjacency before or after). Instead of hopping segment by segment, the window adjusts in orientation to each element. Restricting the window size well below n–1 degrees of adjacency (e.g. 2 to 4 degrees) and keeping a constant orientation
between the element in focus and its referents exemplifies contour level analysis as it is envisioned here.

### 8.2 Continuous C+ Matrices (CONTCOM)

Within the moving window, the degrees of adjacency are constant not the referent pitches, so a square matrix is no longer appropriate. An alternative organization of contour comparisons in a matrix is illustrated with Schoenberg’s Opus 19, No 4.

Figure 8.6: Second phrase of Schoenberg, Op. 19, No. 4, with 4–degree window around third pitch

The second phrase of Schoenberg’s piece (as identified by Morris 1993) is 12–notes in length. A COM-matrix for this melodic segment is 12–by-12, so is a C+ matrix. Instead of ternary contour [+ 0 –], the C+ matrix uses binary values, wherein a value of 1 signifies the column pitch is greater than the row pitch and 0 signifies the column pitch is equal to or less than the row pitch. Figure 8.6 shows a 4–degree window (dashed ellipse) around the third pitch (dotted circle) of the phrase. In the rightmost matrix in Figure 8.7, values within a 4–degree window around the central diagonal (the identity) are shown in bold.
By extracting a window of four degrees around the central diagonal, I get a series of contour slices. These can be arranged side by side with the rows now representing a degree of adjacency instead of a referent pitch. A Continuous C+ matrix (CONTCOM) can be made for any monophonic voice in a score (encoded as MIDI or XML) or a segmented recording of speech, and so on. The melodic foreground of the Schoenberg piece is short, only 47 notes long, so a CONTCOM for the entire piece can be visualized. While a CONTCOM for all of Bach’s C-minor Fugue or Owadayo’s sermon cannot be visualized on a page, the CONTCOM is apt for computational analysis of these larger forms.
Figure 8.8 shows the relationship between the windowed comparisons within phrase 2 extracted from the $C^+$ matrix in Figure 8.7 and the CONTCOM for the entire melody in Schoenberg's piece. Two degrees of adjacency above and below the main diagonal\(^{59}\) can be squashed into a four-row column if the identity is removed. In Figure 8.8, values for the third pitch are sliced from the windowed $C^+$ matrix, and arranged side-by-side with other slices in a CONTCOM for phrase 2. The placement of the phrase 2 CONTCOM is shown within the CONTCOM for the entire piece. MIDI pitches ($C4=60$) are shown above the full CONTCOM and notation for the melodic foreground is shown below.

Backward comparisons at the beginning and forward comparisons at the end are impossible, so these entries are blacked out. With a radius of two degrees around the focus, the first two and last two contour slices will have insufficient context. This is

\(^{59}\) In Morris's COM matrix, the upper left and lower right portions are inversions of each other, in the $C^+$ matrix they are not because the binary relationship is asymmetrical (1 means higher, 0 means equal or lower).
similar to the initialization problem highlighted by Ladd (2008a) for the syntagmatic approach to tone. Because it effects both the beginning and ending slices, I call this problem undercontextualization. This will be addressed later in the chapter.

A continuous C+ Matrix (CONTCOM) lies somewhere in between the note-to-note (pairwise) model and the full combinatorial model of contour (n-1 degrees of adjacency). The note-to-note model has been addressed frequently by perceptual studies and is equivalent to Friedmann’s Contour Adjacency Series. In general, music theorists have considered degrees of adjacency beyond immediate neighbors, but as emphasized in the previous chapter, using all degrees of adjacency may not be necessary. To be created, a continuous C+ matrix (CONTCOM) requires a limit on degrees of adjacency, avoiding an all-or-nothing approach to super-adjacency.

8.2.1 Contour Slices and Contour Levels

Each column of the matrix at the bottom of Figure 8.8 is a contour slice. This is a discretization of relative pitch height which compares the pitch in focus to other pitches within the adjacency radius. For a window size of four degrees, there are four comparisons, for six degrees there are six comparisons and so on. In the Schoenberg example, the windowing is equal distributed among pre- and post-adjacent degrees, but this is not a requirement. CONTCOM easily adapts to any orientation around the focused pitch, symmetrical or asymmetrical. Comparisons that are proximal in CONTCOM with a symmetrical radius are to some extent dependent. The +1 comparison of one slice is the complement of the –1 comparison in the next slice, unless there is an adjacent repeated
pitch. Ignoring the partial redundancy,\textsuperscript{60} there are 7 independent comparisons in two adjacent slices of a CONTCOM with a symmetrical window of four degrees (a two-degree radius around the focus). While super-adjacency is useful, piling on degrees quickly becomes overly descriptive (as will be demonstrated with the Bach Fugue subject). Smaller window sizes grant flexibility to recognize similar but inexact contours within a CONTCOM and to model contours with levels, the parallel to contour pitches in this revised approach.

In the \textit{cseg}, pitches are mapped to contour pitches, ranked from low (0) to high (n-1) in contour space (see Marvin and Laprade 1987). With windowing, height in contour space is fundamentally different. In fact, it is no longer contour space in the same sense. Pitches are not assigned a unique contour pitch, but given a value for relative height within the window. Notes with different pitch may be assigned to the same level, and notes with the same pitch may be assigned to different levels. Contour levels are simply the sum of the contour slice and range from 0 to the degrees of comparison. For a window of four, the maximum level is 4. Because 0 is a level, the total number of levels is the number of degrees plus one. The next few figures illustrate the most important use of CONTCOM: searching for similarity. This may be done using contour slices or contour levels. Contour slices retain their syntagmatic context, but contour levels do not. Contour levels are quasi-paradigmatic, categorical relative pitch height similar to the tone levels addressed in the first part of this dissertation.

\textsuperscript{60} I say partial because it is necessary to have both to know that it is a repeated pitch. A repeated pitch is indicated in the CONTCOM by the absence of ascent (a 1 value) in both the +1 degree of a one slice and –1 degree of the next slice.
<table>
<thead>
<tr>
<th>Contour Slice</th>
<th>Multiplicity</th>
</tr>
</thead>
<tbody>
<tr>
<td>[0;0;0;0]</td>
<td>12</td>
</tr>
<tr>
<td>[1;1;1;1]</td>
<td>7</td>
</tr>
<tr>
<td>[1;0;1;1]</td>
<td>4</td>
</tr>
<tr>
<td>[0;1;0;1]</td>
<td>3</td>
</tr>
<tr>
<td>[1;1;0;0]</td>
<td>3</td>
</tr>
<tr>
<td>[1;1;1;0]</td>
<td>3</td>
</tr>
<tr>
<td>[0;0;1;1]</td>
<td>2</td>
</tr>
<tr>
<td>[0;1;0;0]</td>
<td>2</td>
</tr>
<tr>
<td>[1;1;0;1]</td>
<td>2</td>
</tr>
<tr>
<td>[0;0;1;0]</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 8.3: A ranking of contour slices in Schoenberg by multiplicity

We start by finding recurrences of the same contour slice within the Schoenberg CONTCOM. Recursion of larger patterns is addressed in the next chapter. The most common slice within a window of four degrees is the local minima. There are slices that are exactly the same as other slices (Figure 8.10) and other slices which are unique within the series (Figure 8.11).
There is exact similarity between slices, and then there is similarity through contour level, the sum of the contour slice. The most common level in the Schoenberg piece is 0. This corresponds to the most common slice \([0;0;0;0]\). The second most common level is 3 (the sub-maxima), the sum of any slice with three ascents (1s).

<table>
<thead>
<tr>
<th>Level</th>
<th>Multiplicity</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 (1.0)</td>
<td>7</td>
</tr>
<tr>
<td>3 (.75)</td>
<td>11</td>
</tr>
<tr>
<td>2 (.50)</td>
<td>10</td>
</tr>
<tr>
<td>1 (.25)</td>
<td>3</td>
</tr>
<tr>
<td>0 (0.0)</td>
<td>12</td>
</tr>
</tbody>
</table>

Table 8.4: A ranking of contour slices by contour level (windowed relative pitch height)

Figure 8.12 shows the CONTCOM with an added row for the sum of each column. This is the contour level for each pitch. Level 3 events are in bold in Figure 8.12 and boxed in Figure 8.13.
Note that slices 27 and 28 (adjacent slices in bold outline) are submaximal to different local maxima. Slice 27 is submaximal to 28, but 28 is submaximal to 30. With a window of six degrees, neither would be a sub-maximum.

8.2.2 CHOOSING WINDOW SIZES

Beyond the phonetic framework of the lexical-order (see Chapter 6 or Section 8.1) and theories of perception and memory discussed in Chapter 7, there are practical considerations for setting a window size to make a CONTCOM. First, one must consider what the minimum cardinality of segments should be. If one is interested in comparing melodic segments of cardinality 10 or greater, then the window should be no larger than 9 degrees and probably much less. The total number of degrees should not exceed the minimum cardinality of interest. Second, one should consider how many unique pitches will be in melodic segments. In the Bach C-minor fugue, the subject has six unique pitches, so it does not make sense to have a window greater with more than five degrees of adjacency. As will be shown, this artificially elaborates the signal. A third consideration is how rigid one wants to be about contour similarity in an analysis. The level of detail in the continuous matrix increases with the number of degrees of adjacency included. The lower the degrees, the more flexibility there is to recognize segments as similar that are not strict transpositions. In the CONTCOM for the Schoenberg piece (Figure 8.8), two degrees of pre- and post-adjacency are used for each pitch in the series, producing a window of four degrees symmetrically-oriented around the focused event. Alternatively, the orientation does not have to be split equally into hindsight and

226
foresight. The degrees could be all pre-adjacent or all post-adjacent. To model real-time perception of speech, it makes sense to orient the window towards pre-adjacency. Hearing into the future is not so concrete as comparing a pitch to the pitches before it. However, different possible outcomes in post-adjacency might be useful to model expectation. For analysis of a musical composition, pre- and post-adjacent degrees are certainly appropriate.

CONTCOM is not entirely without precedent. Marvin and Laprade (1987) call the diagonals above the central diagonal INT$_1$ and so on to INT$_{n-1}$. The diagonal INTs of the COM-matrix correspond to the rows of CONTCOM. Because we are using binary ascent categories of comparison, it is preferable to include what might be termed INT$_{-1}$ and INT$_{-2}$. As Quinn emphasizes in his 1997 and 1999 articles, C+ comparisons do not differentiate between the “0” and “-” categories of the ternary comparisons, so it is necessary to use the entire C+ matrix (excluding the central diagonal) to calculate similarity (C+SIM).\(^{61}\)

Another note on previous work is that windowing has been applied in contour theory before, but with very different motivation and application. In his 1993 article, Morris formalizes a contour reduction algorithm, which reduces the contour of a melodic segment to prototypes with only initial, final, maximum and minimum elements. Two recent dissertations, Schultz (2008) and Bor (2009), introduce windowing to Morris’s contour reduction algorithm. These however are windows for pruning medial pitches and not windows of comparison. A commonality is that Bor shows that different window

---

\(^{61}\) Likewise it is necessary to include pre- and post-adjacency comparisons to know that a pitch is repeated in CONTCOM (see Section 8.2.1).
sizes produce different results in the reduction algorithm. Likewise, different window sizes produce different results in contour level analysis. Bor calls these different window sizes 3–depth, 5–depth, and so on (2009:8, 55–77). Because the goal of windowing in this case is to create the CONTCOM, not to prune pitches in a reduction algorithm, Bor’s 3–depth corresponds to a Window of 2 comparisons. Window size of 2 produces two comparisons in a contour slice and excludes the identity. A window size of two will however produces three contour levels (explained in the next section).

8.2.3 GLOBAL COMPARISON MATRIX (GLOBCOM)

In addition to local adjacency, the relationship of a pitch to prominent events in a piece may be perceptually salient. Comparison of every pitch to every other pitch in a piece is impractical, and is not suggested by conventional contour analysis, which requires manual segmentation. If one imagines a COM-matrix for a large (un-segmented) pitch series (as in Section 8.1), there are two types of comparisons that may be particularly salient:

1) *Local contour* including more than one (but not too many) degrees of adjacency, as in CONTCOM; and

2) *Global contour*, wherein each pitch is compared with:

a) the initial pitch (1),

b) the global maximum (+),

c) the global minimum (-), and

d) the final of the series (End or n for cardinality).
Figure 8.14 shows a continuous Global C+ Matrix (GLOBCOM). The binary comparison categories [0,1] are not uniform here: for the 2\textsuperscript{nd} row (global max comparison) and the 3\textsuperscript{rd} row (global minimum comparison) they represent location in time (1 is the extrema or any event after) and for rows 1 and 4, [0,1] are the same as for the CONTCOM (i.e. Quinn’s C+ ascent binary). The first pitch compared with itself is 0.

![Figure 8.14: a Global Comparison Matrix (GLOBCOM) for Schoenberg Op. 19, No. 4. The bolded slices are the locations of the Global Maximum and Global Minimum](image)

While GLOBCOM is not useful for identifying segments based on local contour patterns (as the CONTCOM is), it is apt for placing an event or contour segment within the larger context of a piece. Three sections of a complete series can be identified with GLOBCOM if the initial or final pitch is not also a pitch extrema. These are: 1) before the first global extreme, 2) in-between the extremes, 3) after the last global extreme. These do not necessarily correspond to segmentation based on contour recursion, in fact, it is predicted there is no such correspondence and the global maxima and minima would tend to occur within and not at the onsets or offsets of recursive contour segments. If the sections are roughly equivalent (and they are not here), the “in-between” section should be the most dynamic in terms of pitch and therefore contour because it spans the extremes in an equal number of events. GLOBCOM’s potential is not addressed further in this dissertation, it is mentioned as alternative form of continuous matrix deserving further exploration.
8.3 Tonal Motivations for Contour Levels

Schoenberg's Opus 19, No. 4, is of an ideal length to visualize an entire CONTCOM, but post-tonal music is not the primary domain of contour level analysis. It can also be applied to tonal music, and with better effect than prior methods because non-adjacent recurring pitches do not strain the system. Figure 8.15 is a representation of Bach's C-minor Fugue (BWV 847) with the three voices arranged in order of appearance as blocks within a gigantic C+ matrix. Iterations of the Fugue subject appear as smaller phrase blocks (in blue) along the central diagonal. The subject has cardinality 20 and, through its eight iterations, accounts for over 20% of the pitch content of the piece (160/750).

The next chapter details a pattern-finding algorithm that can identify all appearances of the subject within a CONTCOM for all three voices of the fugue. In this section,
windowing is demonstrated as a solution to the problem posed in the introduction to this chapter. We know all iterations of the subject have similar contour, but how can we express this formally?

Measure 1, Voice 1:          Measure 3, Voice 2:
Measure 7, Voice 3:          Measure 11, Voice 2:
Measure 15, Voice 1:         Measure 20, Voice 2:
Measure 26, Voice 3:         Measure 29, Voice 2:

Figure 8.16: All iterations of the C-minor Fugue subject

To define the first presentation of the subject (the model) in traditional contour space, I will use Quinn’s C+ matrix (on the left side of Figure 8.17). The sum of each column in the matrix is given below, but it does not form a cseg of consecutive integers because of several recurring pitches. To get a contour space reduction for a melody with repeated pitches, we must apply a function on the row of sums called consecutize. Consecutization removes all the gaps (between 1 and 3, 3 and 7 and so on) revealing a cseg with 7 contour pitches (0 to 6). There are seven unique pitches in the subject, so there should be seven contour pitches in the cseg of the subject. This C+ matrix and cseg on the left of Figure 8.17 describes the model and all of its diatonic transpositions (real answers), but it does not describe the tonal answer. The only machinery for modeling the subject and its tonal
answer in a matrix, though they only differ by a pair of diatonic intervals, is using fuzzy matrix (introduced by Quinn 1997).

The matrix on the right of Figure 8.17 is a fuzzy C+ matrix. Each cell is an average value for all iterations of the subject. The consistency across iterations of the subject, including the tonal answers, is very high. The lowest C+SIM for any two iterations of the subject is .987. There are only 5 fuzzy comparisons. Summing fuzzy values into a cseg is not addressed by Quinn (1997 or 1999). An issue exposed in Figure 8.17 is summed fuzzy values lead to fuzzy sums. When we apply cons() to the sum (below the fuzzy matrix) we enter artificial contour space with a greater number of contour pitches (9) than is found in either the subject or its tonal answer (7). There is no cseg, with the same number of pitches as contour pitches, that can describe all appearances of the subject. The contour segment shared by the subject and its iterations cannot be modeled in conventional contour space. Individual realizations of the subject can be described with a cseg but not
all with a single cseg. Conventional contour analysis fails to describe (and explain) patently similar contours in tonal music, unless proportions of similarity (CSIM or fuzzy values) are applied. Contour level analysis provides an alternative approach.

Figure 8.18 shows the fuzzy C+ matrix with a consistent window of 10 degrees of adjacency for each column. This is the largest window size that can be maintained without including fuzzy values. By windowing in such a way, we get all crisp 0 and 1 values. As in Section 8.2, the extracted contour slices can be arranged side-by-side with rows aligned to a specific degree of adjacency. Instead of maintaining n-1 degrees of comparison for each pitch, there are 10 degrees of comparison. Whether those degrees are all forward, all backward, or somewhere in-between fluctuates based on proximity to the boundaries of the segment. Allowing the window to shift in its orientation around the focused event is one possible solution to the undercontextualization problem of a
CONTCOM. This is not necessary for pattern-finding within an unsegmented series (as in the next chapter), but necessary for summing slices into contour levels.

The modified matrix on the right of Figure 8.18 represents all of the recursion of the subject without fuzzy values, so it can be summed into contour levels below each column. Although a window of ten is consistent across all iterations, this creates some artificial detail in the analysis. A window of 10 degrees produces 11 levels. This enters an artificial space, much like the summed and consecutized fuzzy values. A \( cseg \) has a number of unique values equivalent to the number of unique values in the melodic segment. There are seven unique pitches in the model, so as advised in Section 8.2.2, there should be less than seven levels which requires a window size of less than six degrees. An even number of degrees is preferable to form a symmetrical radius around the focus, so a window size of four degrees, as applied to the Schoenberg piece, is logical. Figure 8.19 (in the next section) shows a matrix for the subject and its iterations like the CONTCOM for Schoenberg's piece in 8.2, but with the exception that the 4–degree window is allowed to shift it its orientation around the focus.

### 8.4 Staff Notation of Contour Levels

Conveniently, five contour levels can be visualized on the musical staff. This parallels the staff mapping of tone levels proposed by Yorùbá music scholars (see Chapter 2). When staff lines represent contour levels (as in Figure 8.19), they do not indicate a constant pitch or frequency band, reinforced by the lack of a clef. The levels indicate contour equivalence, not pitch equivalence. In conventional contour space, contour intervals are
generic, but contour pitches themselves still retain pitch equivalence within the contour segment. The absolute pitch of a contour level changes over time, much like a tone level. This is observed by comparing standard notation for the Fugue subject with the contour level mapping (on the staff) for the Fugue subject in Figure 8.19.

There are seven pitches in the model segment at the top of Figure 8.19. The local minima, which are very important to this subject, are clearly represented in the staff mapping. There are seven pitches in the tonal answer (at the bottom of Figure 8.19), but they are used differently. The fourth note in the tonal answer is a fifth below the third note. This one aberration confounds a cseg but poses no problem for a contour level series. The mapping has the same cardinality as the object, so there is no undercontextualization. It is

---

62 Another important note in each iteration is the fifth note (Ab4 in the subject and Eb4 in the answer). This is emphasized aurally through its metric placement.
much more sensitive than a contour adjacency series, but can still recognize what Covey calls an adaptation: transformations other than strict mod12 or mod7 transposition.

8.5 Features of Contour Levels

Because contour levels shift in pitch height, they have special properties both conducive to analysis and counter-intuitive to other musical logic. In essence, the levels correspond to local minima, maxima and one or more mediants within the local window. With a two-degree window there is one medial level, with a four-degree window there are three medial levels, and so on. The aberrant behavior of contour levels is exemplified by a two-degree window symmetrically-oriented around the focus, but other window sizes and orientations have their own peculiarities. There are two types of anomalies that occur: when different pitches are assigned to the same level, and when the same pitch is assigned to different levels. The former is desirable and makes the productive reductions, as applied to the Bach subject and its answer, possible. However, the latter is perplexing.

Figure 8.20: Same-pitch different-level anomalies

Figure 8.20 shows two instances of recurring pitches being assigned to different levels. On the left, a repeated pitch loses and regains height. On the right, the same pitch (D) has a level of 1 when it is in the middle of an ascent or descent, but a level of 0 when it is a
lower neighbor. If the orientation were shifted to all pre-adjacent degrees, the second E
(on the left side) would be at level 1.

The inverse phenomena, wherein the different pitches are reduced to the same
level, is useful not only to show similarity between contours with intervallic variation,
but also as a means of reduction, ultimately making it possible to compare contours of
different cardinalities (see Chapter 11). Morris (1993) introduces an algorithm for
reducing contours to time and pitch extrema: the first, last, highest and lowest pitches in a
segment. In contour level analysis, reduction is also used to compare segments, but the
reduction is fundamentally different. Redundant Contour Slices (RCSs) are adjacent
columns in a CONTCOM with the same simple-adjacency relationships (+/− 1). An
ascending chromatic scale from C4 (MIDI pitch 60) to C5 (MIDI pitch 72), can be
reduced to three slices. All but the first and last slices are reduced to a single intermediate
slice in an ascent.

Likewise, an ascending C-major arpeggio reduces to three slices. Both Figure 8.21 and
Figure 8.22 reduce to the same contour slices, but the mean pitch for the reduced slice is
not the same because the reduced pitches are not distributed evenly between the first and
last note in the arpeggio. If there are super-adjacent comparisons (i.e. +2 or −2), they will
need to be recalculated after a mean pitch is calculated for the reduced slice.
Morris's reduction would only leave two elements in the reduced contour of the ascending scale or arpeggio. The initial and final would double as the maximum and minimum respectively. There is no information for the relationship between the first pitch to a prior pitch and or the final pitch to later pitches, making them distinct from the intermediate notes. The first pitch and the last pitch of a finite contour series have special prominence, but in the following figures, one can also imagine these are segments from a larger contour stream and that what is beyond the edges may involve a change of direction. In a larger CONTCOM with multiple turns, reduction of ascents and descents is broken by local maxima and minima (turns). Local minima and maxima are never eliminated through this process because they cannot be redundant. Furthermore, one slice is retained among the redundant slices. Contour slice reduction is considerably less drastic than reducing all but time and pitch extremes.

There is also some precedent for the elimination of repeated pitches in ethnomusicological studies (Polansky 1996:261). While reducing local contour recursion may eliminate some repeated pitches from the matrix, it does not delete them all and is therefore not equivalent to deleting repeated pitches.
Figure 8.23 shows the process of reduction for a contour segment of the same length as the chromatic scale but many repeated pitches in the center (instead of stepwise motion). This produces a longer reduced contour segment than the chromatic scale because the first and last repeated pitch have different slices. Hence, eliminating repeated pitches and eliminating redundant contour slices are distinct operations.

8.6 Five-Level Reduction of Recursive Contours in Bach

The richness of a contour segment lies in the number of turns it contains. The subject of Bach’s C-minor Fugue is irreducible because there are no redundant slices at a depth of four degrees.\(^{63}\)

![Figure 8.24](image)

This is not the case for all of the recursive contours in the Fugue, including the first counter-subject.

\(^{63}\) There is only one at two degrees: the penultimate slice.
In contrast to the angular subject, the first counter-subject includes two scalar descents. All iterations of this counter-subject produce the same contour slices and levels.

The first descent has more than five elements, so the contour levels plateau from events 3 to 6. If the window had two degrees of comparison, 2 through 7 and 10 through 12 would be reduced.
With five levels, only descents with cardinalities greater than five reduce (as in Figure 8.27).

The episodes of the C-minor Fugue also include a recursive contour segment. Iterations of the episode material are diatonic transpositions, save an inversion in Measure 13, and an tonal adaptation in Measure 24.

All iterations may be represented with the same contour slices, contour levels and five-level reduction. The first three notes are spread across the five levels. Two long descents both reduce significantly.
Iterations of the second counter-subject pose challenges to any form of contour analysis.

The use of octave displacement and partial transformations (elaborated in Section 9.4) defy any measure or model for contour similarity, including contour levels.
9. CONTOUR RECURSION AND AUTO-SEGMENTATION

Introduction

In this chapter, *contour recursion* is proposed as a basis for melodic segmentation of musical works and a computational method. I posit there are optimal segmentations that reflect composer intentions and listener capacities, and these can be found through computation. Formally, this methodology is embedded in contour matrices (Morris 1987), summations of matrices (Marvin and Laprade 1987) and binary contour comparisons (Quinn 1997) along with the widely used signal processing techniques of windowing (applied to contour previously by Polansky 1996). Here contour recursion refers to any pattern of ups and downs found at multiple indices in an ordinal pitch series. Non-adjacent pairwise comparisons are made, but these are within a constant window of degrees for which cardinality is a consideration but not determinant. At its most basic level, the segmentation algorithm seeks to describe a large portion of a monophonic signal with a small number of repeating patterns. The process can be refined by using a larger or smaller window of comparison, reducing local recursion, allowing or prohibiting overlapping iterations, setting a minimum cardinality, excluding segments that span gaps in the series (restarts) or not, and so on. Reliance on information that may not be available depending on the source (symbolic or recorded), such as meter, articulation and dynamics, is intentionally avoided. Consideration of non-pitch features would generally improve an analysis based on the methods in this article, and to a limited extent
durational information (gaps between offsets and onsets) is used. In Marvin (1991) analysis of melodic contour and durational contour are complementary, and that would be a first step in expanding this method. It is expected that segmentation based on contour recursion is relevant to other dimensions of sound. However, because the intent is to present a carefully formalized and applied theory, limiting the discussion to one dimension reduces the number of variables and functions that need to be defined. All that is needed to apply this method in its current form is a series of pitch values ordered in time with onsets and offsets. Pitch can be measured in semitones (C4=60) or frequency (Hz), but should not be reduced by octave equivalence or to any particular tuning system. If a finer resolution than semitones is needed to distinguish non-equivalent pitches, then frequency should be used. Time information can be in beats or seconds, it is simply used as a chronological index. The robustness of the segmentation algorithm hinges on its ability to identify recursive patterns, recognize transformations, and compare segments of different cardinalities. The latter is accomplished through reducing contour recursion locally, wherein an ascending scale and an ascending arpeggio are both reduced to a cardinality of three (see Figure 9.10 and Figure 9.11). All of the component functions of the algorithm are applied to a uniformly encoded contour matrix that lies between the note-to-note model of a contour adjacency series (Friedmann 1985) and the full combinatoriality of a COM-matrix.
9.1 Manual Segmentation

Morris (1993) presents an analysis of the melodic foreground of a Schoenberg piano miniature (Op. 19, No. 4). He introduces an algorithm that reduces phrase segments to the relationship between time and pitch extrema (the first, the last, the highest and the lowest). Unlike the COM-matrix, the contour reduction algorithm is not sensitive to cardinality. Like the COM-matrix, it cannot be applied meaningfully until the piece is segmented. If anything, segmentation is more crucial, without it, an entire piece is reduced to a prime of (at most) four elements. Morris’s extraction of the melodic voice from the full texture of Op. 19, No. 4 (which has block chords interspersed) is uncomplicated, but the segmentation into phrases deserves more exploration. The careful elaboration of the contour reduction algorithm starkly contrasts this single sentence used to describe the entire process of segmenting the piece into phrases: “Phrase boundaries conform to traditional criteria: slurs and other forms of articulation, punctuating gaps, shape, and referential affinity” (Morris 1993:209). Two of Morris’s criteria for segmentation sound a lot like contour: “shape, and referential affinity”. His description indicates he segmented it by visually examining a score. So, contour was used liberally as a visual gestalt in segmenting the work, and then very methodically to reduce each segmented phrase.

![Figure 9.1: Schoenberg, Op. 19 No.4, segmented into phrases (from Morris 1993, 214)](image-url)
In examining Figure 3a at sight, without fine detail to pitch, we can see that phrase 1 and phrase 4 have similar contours, in a different range and with different durations. The variation in interval magnitude means it is not an exact transposition. Phrase 4 has two sets of recurring pitches (F4–F4 and Bb3–Bb3) within the segment, so the COM-matrices for the two are similar but not equivalent. Based on phrases 1 and 4, and ignoring the others, we could make a generalization that where there are recurrences of a similar contour segment of cardinality (n) there is a phrase of cardinality (n).

\[ n_{\text{phr}} = n_{\text{seg}} \]

However, this breaks down when it is extended to the rest of the piece because Morris’s phrase 2 includes two iterations of a similar contour with CSIM of .80 (12 out 15 comparisons are the same, see Figure 9.2).

An expert human analyst does not always see a one-to-one correspondence between the boundaries of recurring contour segments and phrases. We can revise our heuristic as so: where there is one or more recurring contour segments of cardinality \( n \) there may be one or more phrases of cardinality \( n^i \), where \( i \) is a positive integer up to the number of contour segments found in a cluster.

\[ n_{\text{phr}} = n_{\text{seg}}^i \]

246
The generalization that a phrase may be formed out of an isolated recurring contour or a cluster of recurring contours describes half the phrases identified by Morris. The relief left between the segments that form phrase 2 and phrase 4 forms Morris’s phrase 3, leaving just one boundary between phrases 5 and 6 (addressed in Section 9.4). Contour patterns can be used to both segment a pitch series and compare melodic segments to each other.

9.2 Melodic Pattern-Finding using CONTCOM

There are other useful interpretations of pitch data using a CONTCOM (introduced in Chapter 9), but the primary motivation here is finding optimal segmentations. First, I implemented a simple algorithm in MATLAB to search for the most common melodic segments of a single cardinality. The algorithm approaches the search with no information about the piece except an ordinal pitch series. The figures in this section are unique applications of the search algorithm to six different cardinalities starting with two and augmenting segment size until there are diminishing returns, which for this piece is cardinality seven. Greater sizes could easily be searched for as well, but beyond a certain cardinality, there is no recursion at all.

The algorithm excludes cells that compare the focus to referents outside of the segment, so the dyad does not take advantage of the non-adjacent degrees included in \text{CONTCOM}_4 (see Figure 9.3).
All dyads in the Schoenberg miniature are either ascents or descents, there are no horizontal dyads. In Figure 6b, the most common melodic triad is shown as the search algorithm sees it. In contrast to the contour slice, this jagged excerpt of the CONTCOM gives complete contour information (as in a C+ matrix) about the relationships between three adjacent elements but nothing about their relationship to outside pitches.
Dyads and triads were identified as basic building blocks of melodic contour by Seeger (1960) and Kolinski (1965). The sub-segmental nature of these lower cardinalities is reflected in the decomposition of CSEGs into CSUBSEGs by Marvin and Laprade (1987). Figure 9.3 and Figure 9.4 demonstrate the over-segmentation that occurs when the algorithm searches for cardinalities two and three.

As the cardinality increases, the amount of recursion of a single segment decreases (see Table 9.1). For cardinality four, the most common contour segment has six instances, as compared to 10 for triads. At four, contours more characteristic to the piece begin to appear. At six, a meaningful analysis emerges. The high-frequency hextad has the same start points as the high-frequency quintad, but neatly closes the gap between the first and second occurrence, filling out Morris’s phrase 2. A constraint on the search algorithm could be added to exclude contour segments that span rests. As Tenney and Polansky note, temporal separation has a segregative effect on a monophonic succession of elements (1980:208). If such a constraint were added, all the identified hextads would fall within Morris’s phrase boundaries (Figure 9.1). Cardinality six provides the optimal segmentation using a single pattern. Extending the search algorithm to cardinality seven reduces the number of items returned by the most common segment, so there is less coverage ($7\times3 < 6\times4$), and two of Morris’s phrase boundaries are crossed.
Figure 9.5: CONTCOM with the most common melodic quadrad in bold.

Figure 9.6: score and CONTCOM with the most common melodic quintad boxed.

Figure 9.7: score and CONTCOM with the most common melodic hextad boxed.

Figure 9.8: score and CONTCOM with the most common melodic septad boxed.
A trend emerges from increasing the cardinality of the search algorithm. In this piece and all others I have studied, a point is reached beyond which every segment is unique and there is no recursion. In the Schoenberg miniature, the *cardinality saturation point* is nine. At this and greater segment lengths, the total number of segments equals the number of unique segments, and both decrease until the cardinality of the entire series is reached (see Table 9.1).

<table>
<thead>
<tr>
<th>Cardinality</th>
<th>Total Segments</th>
<th>Unique Segments</th>
<th>Maximum Iterations</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>46</td>
<td>2</td>
<td>24</td>
</tr>
<tr>
<td>3</td>
<td>45</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>44</td>
<td>17</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>43</td>
<td>25</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>42</td>
<td>30</td>
<td>4</td>
</tr>
<tr>
<td>7</td>
<td>41</td>
<td>34</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>40</td>
<td>37</td>
<td>2</td>
</tr>
<tr>
<td><strong>9</strong></td>
<td><strong>39</strong></td>
<td><strong>39</strong></td>
<td><strong>1</strong></td>
</tr>
<tr>
<td>10</td>
<td>38</td>
<td>38</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>37</td>
<td>37</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>36</td>
<td>36</td>
<td>1</td>
</tr>
<tr>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
</tr>
<tr>
<td>47 (n)</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 9.1: The cardinality saturation point for the Schoenberg miniature is 9

Beyond the saturation point, there is no recursion. In the lower cardinalities the number of unique segments is at or near the possible number of permutations. For a piece with considerable contour recursion, such as this Schoenberg miniature, the number of unique segments does not keep pace with the number of possible permutations as cardinality increases.
9.3 Segmentation Algorithm

Based in part on Morris’s phrase boundaries, a ground truth for automatic segmentation appears in Figure 9.9. This serves as an empirical standard that, if successful, the algorithm will replicate without any information specific to the piece except for the pitch values, onsets and offsets.

Figure 9.9: A ground truth for the contour search algorithm, with two recursive segments (one circled and one boxed)

Over 83% (39 out of 47 events) of the pitch series can be accounted for with two model segments. The circle motive is a septad and the boxed motive is in most cases a hextad. The third appearance of the boxed motive is extended into a septad by a redundant contour slice (see 9.3.1), the fourth appearance is a similar contour in retrograde. Because the ground truth uses multiple cardinalities, the segmentation algorithm must search and pick segments of multiple cardinalities to succeed. The constraints and methods in Table 7a are also added to improve the algorithm. As input parameters, they can be turned “on” or “off” or adjusted to be appropriate to the musical object being analyzed. The parameters are restricted to pitch information with the exception of number 8, SEEGAP. Of the parameters, 1–3 effect creating and pre-processing the CONTCOM, 4 is a constraint on cardinality, 5 augments the search algorithm by implementing a secondary...
search for transformations of recursive segments within the segment pool (SEGPOOL), and 6–9 filter or evaluate candidates in the SEGPOOL.

### Input Parameters for the Segmentation Algorithm

<table>
<thead>
<tr>
<th>Input Parameters for the Segmentation Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. (Adjustable) <em>Windowed Degrees of Adjacency</em> (DEGWIN): the number of degrees around a focus for comparison.</td>
</tr>
<tr>
<td>2. (Optional) <em>Delete Redundant Contour Slices</em> (REDRCS): this ensures that contour segments expanded by a recursive contour event are recognized as similar to segments without this type of expansion (see Section 4.2). Note: RCS that are adjacent to a time gap (rest) are not reduced. (Default: on)</td>
</tr>
<tr>
<td>3. (Optional) <em>Delete Repeated Pitches</em> (DELREP): distinct from REDRCS (see Section XX) and comes after pre-processing to avoid the creation of new RCSs). (Default: off)</td>
</tr>
<tr>
<td>4. (Adjustable) <em>Minimum Cardinality</em> (MINCARD): the minimum cardinality for the search module, the default is four because two and three over-segment the series.</td>
</tr>
<tr>
<td>5. (Adjustable) <em>Recognize Similar Segments and Transformations</em> (MINSIM): similar segments and transformations such as inversion, retrograde and retrograde inversion, should be recognized as equivalent above a threshold of C+SIM (e.g. 0.8). The windowed C+SIM pertinent to CONTCOM has (2n - 2\Sigma(1,\ldots,r)), of elements to compare if (n&gt;=r) (where (r) is the adjacency radius). The number of comparisons in windowed C+SIM and full C+SIM are compared in Table 7b.</td>
</tr>
<tr>
<td>6. (Adjustable) <em>Minimum Recursion</em> (MINREC): a segment must be recursive to be part of the segmentation, hence cardinalities above the saturation point (where all possible segmentations are unique) are ignored. This can be adjusted to require iterations greater than the default of two.</td>
</tr>
<tr>
<td>7. (Adjustable) <em>Allowable Overlap</em> (OVERLAP): although it is conceivable that contour segments may overlap to create larger segments (in turn becoming subsegments), the emphasis here is on discrete segments. By adjusting the OVERLAP parameter from the default of value of zero, to some positive integer ((n)), iterations of the recursive contour are allowed to overlap by (n) events.</td>
</tr>
<tr>
<td>8. (Optional) <em>See and avoid time gaps</em> (SEEGAP): according to Morris’s phrase boundaries, rests often indicate a boundary, so contour segments that span breaks between a MIDI offset and onset will not be included.</td>
</tr>
</tbody>
</table>

Table 9.2: Input Parameters for the Segmentation Algorithm

### 9.3.1 Reduction of Redundant Contour Slices

Of these parameters, one in particular deserves elaboration, REDRCS, the reduction of redundant contour slices. It is a solution to the cardinality sensitivity problem made possible through the formalization of contour slices (a similar reduction can be applied to contour levels). Through the reduction, melodic segments of different lengths may form equivalent contour segments. Unlike other methods of contour segment reduction,
including the contour level reduction in Section 8.6, the reduction is applied to the entire series before segments are identified. There are no cardinal primes because in contour level space all maxima and minima are equal (level $n$ and level 0 respectively). Slices for two or more consecutive unaccented notes are reduced to one, specifically passing or repeated tones, not notes at a change of direction, what Thomassen calls pivots (1982). Redundant contour slices (RCSs) are defined here as consecutive columns in CONTCOM that hold the same adjacent relationships ($+/−1$ degrees) as the previous column. In its application to ascending or descending motions, the reduction mirrors the fusing of middle memory bins in Figure 2c. For example, an ascending chromatic scale from C4 (MIDI pitch 60) to C5 (MIDI pitch 72) is reduced to the first slice, the last slice and one intermediate slice (see Figure 9.10). Likewise, an ascending C-major arpeggio is reduced to three slices (see Figure 9.11).

The reduction is the same in terms of adjacent degrees (as pictured). However, cells for non-adjacent comparisons will need to be recalculated based on a new mean pitch. There
is no information for the relationship between the first pitch to a prior pitch or the final pitch to later pitches, making them distinct from the intermediate notes in that they are under-contextualized. As Pollack (1952) found, the first pitch and the last pitch of a finite pitch series have special prominence. In addition to initials and finals, local minima and maxima are never eliminated through this process, and although it has similarities to Morris’s algorithm, it is considerably less drastic than reduction to a prime. All pivots and at least one medial event between pivots (if there is one) are kept. REDRCS may reduce some slices for repeated pitches, but it does not remove them all. Figure 9.12 shows the reduction for a segment of the same length as the chromatic scale in Figure 9.10, but many repeated pitches in the center instead of stepwise ascent. This produces a longer reduced contour because the first event at 66 (F#4) produces a different slice than all the others. Deleting repeated pitches (DELREP) and reducing redundant contour slices (REDRCS) are distinct operations on the contour series. Contour slices adjacent to time gaps are not reduced if SEEGAP is turned on. They are more likely to be structural than embellishing. The function REDRCS embodies a perspective on pitch salience (or accent) and underlying contour. The chromatic scale and ascending arpeggio are equivalent when reduced, but the segment with many repeated pitches in the middle is not. Only two redundant slices are found in the CONTCOM for the Schoenberg piece because it is highly angular (see Figure 9.13).
Figure 9.13: Schoenberg, Op. 19, No.4, includes two redundant contour slices and no consecutive repeated pitches

After preparing the CONTCOM (with or without reduction), the search module returns indices in a segment pool (SEGPOOL) for occurrences of each unique segment of CONTCOM within the constraints of the input parameters. For each cardinality, the most common segment is presented to the segment evaluation module (EVALSEG) as a candidate to be the primary segment. Multiple ranked evaluation criteria are used to select from the full range of cardinalities in the SEGPOOL (see Table 9.3).

<table>
<thead>
<tr>
<th>Segment Evaluation Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>(or 2) COVERAGE: is the number of occurrences (without overlap) multiplied by cardinality. The candidate from the SEGPOOL with the maximum coverage is selected. A drawback to this criterion for the primary segment is that it will almost always choose a lower cardinality which breaks up the melody too much, potentially leading to a porous segmentation with many small gaps below the minimum cardinality.</td>
</tr>
<tr>
<td>If maximum coverage is produced by more than one candidate, further EVAL criterion are used:</td>
</tr>
<tr>
<td>(or 1) POINTS are awarded as follows: +1 for each contiguous segment offset to segment onset, -1 if an iteration of the segment crosses another iteration beyond the overlap setting (for which the default is 0, allowing no overlap). Using this criterion favors analyses without gaps.</td>
</tr>
<tr>
<td>If there are multiple segments with positive points, take the greatest cardinality, if zero is the greatest, use the next EVAL criterion:</td>
</tr>
<tr>
<td>SIM-COUNT: largest cardinality of same maximum number of iterations as lower cardinalities.</td>
</tr>
<tr>
<td>If no candidate meets this criteria, go on:</td>
</tr>
<tr>
<td>LAST-COUNT: largest cardinality with minimum amount of recursion (e.g. 2 iterations).</td>
</tr>
<tr>
<td>These criteria are reused to evaluate candidates as secondary segments. The loop of EVALSEG stops when no candidates in the SEGPOOL meet this criterion.</td>
</tr>
</tbody>
</table>

Table 9.3: Segment evaluation criteria
An input parameter COVRPNTS, determines the first criterion for evaluating candidates in the SEGPOOL. Points is intended for tightly-constructed music with close recursion (such as this composition by Schoenberg), and coverage is intended for music with frequent paradigmatic cadences, and all else. Selecting the primary segment (the first parent segment) is the most crucial step. After this point, only segments that fit into the relief left by the primary segment are used.

9.3.2 THE ROLE OF MODIFIED C+SIM

C+SIM, as described in Quinn (1997), is calculated by comparing each cell (save those in the central diagonal) in two C+ matrices of the same size and then dividing the count of identical comparisons by n^2–n (subtracting the central diagonal). Depending on the cardinality of the segment and the degrees of comparison in the CONTCOM, the CONTCOM segment may have substantially fewer elements to compare to other CONTCOM segments of the same cardinality. A modified C+SIM for CONTCOM segments has 2m - 2Σ(1,…,r) elements to compare if n≥r (where r is the radius of degrees around the focus).
Table 9.4: Number of comparisons in mod C+SIM (degree limited) and C+SIM (n-1) for various cardinalities and degrees of adjacency around the focus.

<table>
<thead>
<tr>
<th>Cardinality</th>
<th>Degrees</th>
<th>Windowed C+SIM</th>
<th>Full C+SIM</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>4</td>
<td>10</td>
<td>12</td>
<td>0.83</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>12</td>
<td>12</td>
<td>1.00</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>12</td>
<td>12</td>
<td>1.00</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>14</td>
<td>20</td>
<td>0.70</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>18</td>
<td>20</td>
<td>0.90</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>20</td>
<td>20</td>
<td>1.00</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>18</td>
<td>30</td>
<td>0.60</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>24</td>
<td>30</td>
<td>0.80</td>
</tr>
<tr>
<td>6</td>
<td>8</td>
<td>28</td>
<td>30</td>
<td>0.93</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>22</td>
<td>42</td>
<td>0.52</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>30</td>
<td>42</td>
<td>0.71</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>36</td>
<td>42</td>
<td>0.86</td>
</tr>
</tbody>
</table>

C+ SIM is used to look for segments similar to those selected from the SEGPOOL. The input parameter MINSIM sets a threshold for adoption of similar and transformed segments as child segments, forming a segment family.

9.3.3 WORKFLOW

The workflow in Figure 9.14 produces a single analysis. This process may be run repeatedly, varying a single parameter at a time within a range, before a final evaluation of multiple analyses. In the SEGCOMP phase, if the indices of a superior parent segment's child segments (similar and transformed iterations) correspond with the indices of an inferior parent segment, the resulting analysis for one set of parameters may be reduced before it enters the final evaluation. A key metric for ranking analyses is the number of recursive segments used (optimum is 1). So, any reduction improves the performance of an analysis in the final evaluation. Reduction occurs when the indices of the child segments of a superior parent segment are contained by (or contain) the indices
of an inferior parent segment. The inferior parent along with its children (if any) join the family of the superior segment. Even if adopted segments are not the same cardinality as the parent segment, they are still accepted. Like the reduction of redundant slices, but through a different means (that can be compounded with REDRCS), this brings the possibility of recognizing contours of different cardinalities as related.

Figure 9.14: Workflow for the segmentation algorithm

9.4 Discussion of Schoenberg Analyses

All of the candidate analyses are ranked in a final evaluation metric, using one or more of the following criteria:

   EVAL1: Number of contour segment families;
EVAL2: Number of pitches left out of the segmentation;

EVAL3: Total number of segments divided by the number of leftover pitches.

For all, lower values are preferred. The last criterion gives preference to larger cardinalities. Any single criteria can be used, or they can be multiplied creating an evaluation product that is ranked higher if it is of a lower value. Table 9.5 is a ranked list of analyses with EVAL products less than 5.

<table>
<thead>
<tr>
<th>DEGWIN</th>
<th>MINCARD</th>
<th>SEEGAP</th>
<th>Eval 1</th>
<th>Eval 2</th>
<th>Eval 3</th>
<th>Product</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>0.16</td>
<td>1.11</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>0.16</td>
<td>1.11</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>0.16</td>
<td>1.11</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>9</td>
<td>0.17</td>
<td>1.50</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>9</td>
<td>0.17</td>
<td>1.50</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>0</td>
<td>1</td>
<td>9</td>
<td>0.17</td>
<td>1.50</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>0</td>
<td>1</td>
<td>18</td>
<td>0.11</td>
<td>2.00</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>13</td>
<td>0.19</td>
<td>2.44</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>18</td>
<td>0.11</td>
<td>2.00</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>13</td>
<td>0.19</td>
<td>2.44</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>13</td>
<td>0.19</td>
<td>2.44</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>13</td>
<td>0.19</td>
<td>2.44</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>13</td>
<td>0.19</td>
<td>2.44</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>0</td>
<td>1</td>
<td>21</td>
<td>0.13</td>
<td>2.63</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>21</td>
<td>0.13</td>
<td>2.63</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>0</td>
<td>1</td>
<td>21</td>
<td>0.13</td>
<td>2.63</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>21</td>
<td>0.13</td>
<td>2.63</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>13</td>
<td>0.25</td>
<td>3.25</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>13</td>
<td>0.25</td>
<td>3.25</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>0</td>
<td>1</td>
<td>24</td>
<td>0.14</td>
<td>3.43</td>
</tr>
<tr>
<td>8</td>
<td>7</td>
<td>0</td>
<td>1</td>
<td>24</td>
<td>0.14</td>
<td>3.43</td>
</tr>
<tr>
<td>10</td>
<td>7</td>
<td>0</td>
<td>1</td>
<td>24</td>
<td>0.14</td>
<td>3.43</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>0</td>
<td>1</td>
<td>25</td>
<td>0.15</td>
<td>3.75</td>
</tr>
<tr>
<td>8</td>
<td>6</td>
<td>0</td>
<td>1</td>
<td>25</td>
<td>0.15</td>
<td>3.75</td>
</tr>
<tr>
<td>10</td>
<td>6</td>
<td>0</td>
<td>1</td>
<td>25</td>
<td>0.15</td>
<td>3.75</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>6</td>
<td>0.33</td>
<td>4.00</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>6</td>
<td>0.33</td>
<td>4.00</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>6</td>
<td>0.33</td>
<td>4.00</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>6</td>
<td>0.33</td>
<td>4.00</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>6</td>
<td>0.33</td>
<td>4.00</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>17</td>
<td>0.25</td>
<td>4.25</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>17</td>
<td>0.25</td>
<td>4.25</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>31</td>
<td>0.14</td>
<td>4.43</td>
</tr>
<tr>
<td>8</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>31</td>
<td>0.14</td>
<td>4.43</td>
</tr>
<tr>
<td>10</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>31</td>
<td>0.14</td>
<td>4.43</td>
</tr>
<tr>
<td>12</td>
<td>7</td>
<td>0</td>
<td>1</td>
<td>31</td>
<td>0.14</td>
<td>4.43</td>
</tr>
<tr>
<td>12</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>31</td>
<td>0.14</td>
<td>4.43</td>
</tr>
<tr>
<td>14</td>
<td>7</td>
<td>0</td>
<td>1</td>
<td>31</td>
<td>0.14</td>
<td>4.43</td>
</tr>
<tr>
<td>14</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>31</td>
<td>0.14</td>
<td>4.43</td>
</tr>
</tbody>
</table>

Table 9.5: Input parameters and evaluation of resulting analyses, ranked lowest to highest in terms of evaluation product (last column).

260
Analyses corresponding to the three boxed portions of the table are visualized in Figure 9.15 (Analysis 1), Figure 9.16 (Analysis 2), and Figure 9.17 (Analysis 3). The same analysis may be returned by similar input parameters. Analysis 1 is closest to Morris’s segmentation. Analysis 2 allows segments to cross time gaps (rests). Analysis 3 is included to illustrate over-segmentation resulting from a minimum cardinality of less than 4. For each analysis: (1) the input parameters and output evaluation criteria values are listed; (2) the CONTCOM\textsubscript{4} is shown for the entire piece (Analysis 1 and 2 also show the score); (3) a table of the segment indices is provided; (4) the model for each segment is shown and labeled according to its rank; and (5) if there is more than one segment in a family, a weighted fuzzy summary is calculated based on the model segments and number of iterations for each model. All C+SIM values are restricted to the windowed degrees of adjacency. If the child segment(s) is larger than the parent, these additional cells are shaded gray. The fuzzy values follow Quinn (1997) with the number of iterations for each segment in the family used as a weighting in the calculation. A fuzzy matrix is just one way to summarize similarity between related segments produced by the algorithm, whether iterations of the same model or within the same family. Once one has segments, there are many, many ways to calculate degrees of similarity between them. Options for working with different cardinalities include Morris’s contour reduction algorithm (1993), Huron’s sim\textit{i}l command in the Humdrum Toolkit (1994) or gross shapes (1996), Polansky’s Combinatorial or Linear Direction metrics (1996), phase spectra from Fourier analysis (Schmuckler 2010) or various statistical methods.

Analysis 1 and 2 are the highest-ranking analyses by evaluation product. Rows 1–3 of Table 9.5 correspond to Analysis 1 (Figure 9.15) and rows 4–6 to Analysis 2 (Figure
Keeping the window size constant and varying the minimum cardinality (MINCARD) from 4 to 6 produced the same analysis. The difference between Analysis 1 and 2 in terms of input parameters is the setting for SEEGAP: Analysis 1 has it “on” and Analysis 2 has it “off.” “Punctuations” were part of Morris’s manual segmentation. Excluding segments than cross gaps from the SEGPOOL produces an analysis identical to the ground truth. Analysis 3 is the highest-ranking among the analyses produced with the minimum cardinality set below four. Setting a minimum cardinality of four returns segments that are characteristic to the piece, at cardinalities of six and seven, instead of over-segmenting the piece with triads that are the basic building blocks of all melodic contour (see Seeger 1960, Kolinski 1965). Using a minimum cardinality (MINCARD) of four may be a good general practice for segmentation based on contour recursion.
Figure 9.15: Analysis 1 is highest-ranked by EVAL Product, SEEGAP is turned on (locations indicated on CONTCOM by arrows)
DEGWIN  |  MINCARD  |  SEEGAP  |  Eval 1  |  Eval 2  |  Eval 3  |  Product  \\
--- | --- | --- | --- | --- | --- | --- \\
4  |  4–6  |  0 (off)  |  1  |  9  |  0.17  |  1.50  \\

Figure 9.16: Analysis 2: similar to Analysis 1 but with SEEPGAP turned off, allowing segments to span rests (as from event 35 to 36)
Analysis 1 (Figure 9.15) demonstrates that Morris’s segmentation can largely be recreated with an automated process of searching for and evaluating contour recursion.

That is, with two exceptions: (1) leftover pitches not included in the recursive segment collection, and (2) the combination of recursive segments into a single segment as in Morris’s phrase 2. Exception 1 can be overcome without durational information, but exception 2 cannot. In a post-segmentation module, leftover pitches could be joined to a recursive segment by adding a conditional statement: if there are leftover pitches in clusters of less than the minimum cardinality, they should be joined with the closest segment not separated by a offset-to-onset gap (rest). Any larger cluster should form a non-recursive segment of its own. This produces Morris’s phrase 3 and phrase 5.

Regarding exception 2, Morris’s identification of phrase 2 seems based on uniformity in
duration (a string of sixteenth notes), perhaps a form of “referential affinity”. Because the segmentation algorithm works with only pitch and not durational information, it does not see the larger rhythmic grouping observed by Morris. The fusion of these segments into one phrase cannot be accomplished without considering durations.

9.4.1 Recursive Segment Models (CLSEGs)

CONTCOM is very robust for computation, but less than ideal for visualization. In classic contour theory (e.g. Morris 1987, Marvin and Laprade 1987), the CSEG class is useful for nominalizing a contour matrix into a visually and verbally digestable format, e.g. <0 2 1 > : “zero-two-one.” In this modified contour space, contour pitches have been replaced with contour levels. By extension, the analog to CSEGs are CLSEGs. For the Schoenberg piece, the highest-ranked analysis used CONTCOM_4, so any CLSEG produced from it will be a CLSEG_4, following the subscript convention for the CONTCOM indicating the window size used. In this modified contour space, every pitch is no longer distinct. A recurring pitch will have a different contour level, not some of the time, but almost always. Contour slices of all the same binary value (0s or 1s) mean the event in focus is an extreme within the window around it, but the next local minima or maxima is not likely to be at the same absolute pitch height. A conceptually challenging zero-value is produced by the left side of Figure 9.18.
An event that is equal to everything within its window is a local minimum. The contour level of 0 for the fifth note, though somewhat counter-intuitive, reflects a true statement: it is at the lowest pitch within a four-degree window around it. In contour level space, repeated pitches lose and gain height as they move away or towards pitch variance. The loss or gain of windowed pitch height by repeated pitches is not necessarily a fault of CONTCOM. To the extent this phenomenon is a detriment or benefit to the analysis of contour in music with repeated pitches (of which the Schoenberg piece is not) will be explored in further research, however, it can be avoided fairly easily by collapsing all consecutive repeated pitches to a single event, as on the right side of Figure 9.18. There is precedent for the elimination of repeated pitches in contour analysis by ethnomusicologists and music theorists (Polansky 1996:261). However, this produces an anomaly that is an issue for music without consecutive pitch repetition, such as this Schoenberg piece. In the right side example of Figure 9.18, levels 1 and 4 correspond pitches a step apart. The wide gap has nothing to do with reducing repeated pitches and everything to do with time-extreme events, which are under-contextualized in comparison with the interior contour slices of a CONTCOM. A pitch at the beginning or ending of a musical work is phenomenologically different than a pitch in the interior of the work. CONTCOM neither assumes a genesis of pitch or absolute finality. The initial
and final pitches may have more context, but that context is unknown to the CONTCOM. This does not pose any problem for the search algorithm, but it is a problem for the summation of slices into levels when some slices are incomplete. The maximum level of incomplete slices is less than the complete slices.

There are three ways to handle this problem when making a CLSEG. The first is to include cells for comparisons outside of the segment, if it is not at the beginning or ending of the piece. However, that will not describe all iterations of a recursive segment. The second is to allow the orientation of a window to shift to include more degrees in one direction or the other. This works for other pieces I have studied, such as the subject of Bach’s C-minor Fugue (BWV 847), but in this case, it does not (see Figure 9.19).

The third option is to include a range of possible values for incomplete contour slices. For the Schoenberg piece, the last option is ideal for modeling recursive segments. In Figure 9.20, the parent segment of Analysis 1 yields a CLSEG₄ of <0–2, 4, 3, 0, 1–2, 3–4> and
the child segment yields <0–2, 3–4, 2, 0, 3, 2, 0–1>, with the dash representing a range of possible contour level values.

![Recursive Segments from Schoenberg piece with contour levels mapped to staff](image)

While the contour level for slices with complete context for the adjacency radius can be calculated through a simple summing of the column, when there is incomplete context this is not the case. An interesting case of this is in the parent segment. It appears that the second slice could be level 3 or level 4 with more context, but if we consider the +1 comparison value (to the next event) it is above it and only level 4 is an option, which in turn means the maximum possible contour level for an event immediately prior to the segment is 3. There can be multiple local minima (or other levels), but there can be only one local maxima within a radius of each other (half the window size if it is symmetrical). This creates a bit of a paradox because a pitch prior to the segment could potentially be equal to or higher than the second event, which would create the inverse of the repeated pitch anomaly: adjacent pitches of different height with the same contour level. Once again, this is not seen as an invalidation of the model, but an interesting side
effect of a new abstraction of pitch space. What it appears to suggest is that, although the segment is extracted from the CONTCOM4 with outside context excluded, the outside context still has a phantom presence. The model for the parent segment excludes the possibility of level 3 for event 2. Although the possibility exists within the matrix formation for consecutive events with different pitches to have the same contour level, it does not happen in the actual analysis of the piece. It is only a possibility in the model of the recursive segment and may be constrained from surfacing in the music. To some extent, each segment model dictates surrounding events, or the environments in which it is realized. This extends to recursive behavior. The parent segment is more generative and less restricted in its realizations than the child segment. This can be attributed to a specific characteristic: the beginning CL range does not enclose the ending CL range (as it does in the child segment). If iterations of the child segment were to be arranged consecutively (which they are not in this piece), selection from of a level from the range of options for the ending of the first iteration would influence the level of the beginning of the second iteration or vice versa. For consecutive iterations of the parent segment, the realized initial or final contour level for one iteration does not restrict adjacent realizations.

Each realization of an underlying contour may have a unique relationship to the model CLSEG. The easier path to consider is a mapping from the actual pitches to the model, in which information is reduced. In Figure 9.21, a segment from the score is mapped to conventional contour space (CSEG class <354012>) and then on to contour level (CL) space.
As a further abstraction, the CLSEG can be calculated directly from the pitch series or the CSEG class. It matters not. The mapping from the symbolic data to the CSEG class is lossless in terms of the information needed to create a CLSEG model. The CSEG can be an intermediary step in the mapping, or bypassed. Defining the inverse mapping from the model to the actual pitches is more complicated because it requires the gaining of information: specification from a very general model to what we actually see in the music. The information gained is largely influenced by external musical constraints. For many instruments there are physical constraints, or a composer or improviser may be working with a specific pitch set or scale. There is infinite potential for realization of each CLSEG, as there are for CSEGs. This infinity is tangibly, but perhaps not truly, augmented by realizations that augment the cardinality. Is it possible to narrow this down to likely and unlikely realizations? A heuristic for this may be Prince and Smolensky’s *Optimality Theory* (2004), a phonological method inspired in part by Lerdahl and Jackendoff’s GTTM (1985), which in turn took cues from generative linguistics. In its nascent form, optimality theory evaluates candidate words by ranking constraints that explain why one word surfaces in the lexicon instead of other conceivable or improbable

---

64 Published 2004 but circulated since the 1990s.
words. Generally, constraints that exclude improbable words are easier to come by. When using OT, one does not know all of the inputs. They are hypothetical save the one that will be the successful candidate. More than the single output, the formulation and ranking of constraints is of interest here. Any CLSEG can generate as many inputs to be evaluated by a constraint ranking as one can imagine. For there to be recursion, multiple realizations of the CLSEG must succeed in the evaluation. Not to conflate music and language or words and melodic segments, but a ranking of constraints on outputs may be very useful for mapping a segment in contour space to pitch space. Any constraints would reflect important traits of style and idiomaticism, such as whether it is diatonic or non-diatonic music, and what voice range or instruments will perform the music. Exploring the inverse mapping from models to realizations is not a simple task, but may be very fruitful.

9.5 Contour Recursion in Tonal Music

In this section, the segmentation algorithm is applied Bach's C-minor fugue (BWV 847). Many analyses are available which identify the themes (subject and countersubjects) of the fugue and points of imitation. Bruhn (1993) is used as a ground truth. If it works, there is the potential to get past the preliminaries quickly so that one can dig in to detailed and thoughtful analysis. The output segments and indices from the algorithm make it possible to perform a variety of post-segmentation analyses without manual finding and entering locations and other data. All of the different contour interpretations of the subject in the previous chapter (COM-Matrix, C+ Matrix, fuzzy matrix, windowed
matrix/CONTCOM) as well as any number of non-contour interpretations can be done easily using the MIDI data, avoiding the onerous task of making such calculations manually.

In order to create a CONTCOM, we need a single pitch series (P) and to decide on a window size (w). The C-minor fugue has three voices, so each voice could form a pitch series in itself.

m. 1:

\begin{align*}
V1 & = \text{Treble Clef} \\
V2 & = \text{Treble Clef} \\
V3 & = \text{Treble Clef}
\end{align*}

Figure 9.22: Voices in BWV 847 numbered in order of entrance

The individual voices of a fugue are likely to be self-similar at larger time intervals, but the imitation most salient to listeners is found in quick succession between voices. Instead of a P for each voice, it is better to combine all voices into a single P for imitative polyphony. Figure 9.23 illustrates a concatenation of V1, V2 and V3 from Figure 3.
The window size \((w)\) is to some extent arbitrary. It can be up to the cardinality of \(P\) minus one \((n-1)\) but a size greater than the unique elements in \(P\) will not create a reduction of the pitch space, but an artificial expansion. It was already observed in Section 8.3 that a 10-degree window produced the same pairwise comparisons for the subject and all answers, but because there are only seven unique pitches in each, this adds artificial detail. Thinking forward, another consideration already addressed is to avoid searching for cardinalities beyond a saturation point at which every segment in the CONTCOM is unique. This brings a bit of a causality dilemma because we cannot know the cardinality saturation point until we decide on a window size. However, if you know the minimum cardinality you are interested in, then choose a window size that is smaller than that. If you do not, then it is better to err on the under-descriptive side and avoid rejecting contour recursion that may be of interest. For Schoenberg Op. 19, No. 4 a 4- to 6-degree window returned the same optimal analysis, so there is some latitude.

There are 44 unique pitches in \(P\) and the subject has cardinality 20, so the window size should be less than either of these. A window of 4 degrees produces 5 contour levels. This corresponds roughly to Pollack’s 1952 finding that participants conflated the middle pitches in an identification task involving collections of six or more pitches. Although this is tonal music and tonal relationships disrupt contour recognition to some extent

\[65\] In the list \(P\), simultaneities are treated ordered as individual pitches from low to high.
(Edworthy 1983), imitative polyphony conveys a musical aesthetic in which listeners attend to recurring contour patterns. Five contour levels are considerably more descriptive than a contour adjacency series, but far less descriptive than an Interval Adjacency Series with directions or a complete COM-matrix (with n-1 degrees of adjacency). This leaves room for recognizing inexact imitation that satisfies tonal constraints.

![Figure 9.24: CONTCOM for all three voices of Bach C-minor Fugue](image)

Figure 9.24 shows the CONTCOM with black squares at the boundaries of each voice (undercontextualized cells) and the vast interior of each voice substituted by ellipses. In P, V1 corresponds to the first 252 elements ($i=1–252$), V2 is the next 263 elements ($i=253–515$) and V3 the last 235 elements ($i=516–750$).

Once we have a CONTCOM, the auto-segmentation method from CRAS can be applied. The preprocessing crucial to the Schoenberg analysis, reducing redundant contour slices, is not applied here. A pool of recursive segment candidates (SEGPOOL)

---

66 C-2_p is the top row, C-1_p is the second row, C^1_p is the third row, C^2_p is the bottom row. Comparisons cannot be made to pitches before the start of V1, or the end of V3. Likewise comparisons are not made between the end of V1 and the start of V2, and the end of V2 and the start of V3. These are u (NaN) values.

67 Input parameters for auto-segmentation algorithm (for definitions refer to CRAS): numseg=3; cmin=10; cmax=25; degrad=2; delgap=0; elimRCS=0; delrep=1; overlap=0; minrec=4; covORpnts=0; simmin=1.
is produced by searching CONTCOM within a range of cardinalities. There must be a minimum to avoid over-segmentation by prototypical subsegments of cardinalities two and three which are irreducible using a method such as Morris (1993). The maximum arises from the cardinality saturation point at which every segment of that cardinality or greater is unique within the CONTCOM, that cardinality is 45 for this P. Cardinalities without recursion are automatically excluded from the SEGPOOL. A lot of computational expense can be saved by setting a maximum cardinality below the saturation point. In addition, a minimum amount of recursion can be specified. After SEGPOOL is created, the candidates are evaluated by ranked criteria. For the Schoenberg, Points was used to evaluate the SEGPOOL. The Points criterion favors candidate segments with immediately-adjacent iterations in P. In tonal music, a liquidation into conventional material (i.e. cadences) occurs towards the end of phrases, so an alternative criterion called Coverage is more appropriate. Coverage selects the candidate from SEGPOOL which, through its iterations, contains the greatest number of elements in P. Restated, the first segment returned by coverage represents the largest subset of P that can be explained by one recursive model contour. To be consistent with the aim of formalizing contour level space, Coverage can be stated as an equation:

After the primary segment (S1) is selected, all of the indices for other model segments that overlap with iterations of S1 are removed from the SEGPOOL, and all coverages must be recalculated. The evaluation of the remaining candidates and redaction of SEGPOOL continues until the number of desired segments is reached or there are no candidates left to evaluate. The Bach C-minor fugue has a subject and two counter-subjects, so I told the algorithm to stop after finding three segments. The expectation that
the three model segments returned would correspond to the subject and two counter-subjects proved to be false.

As Siglind Bruhn points out, the appearance of the first counter-subject (CS1) starting in Measure 26 starts in the highest voice (V2 here) but proceeds in the middle voice (V1 here) from the start of Measure 27 (1993).

Bach follows a path that is very idiomatic but the algorithm, which operates on a linear ordering of pitches in P, is insensitive to it. It would be possible and potentially fruitful to re-write the algorithm to also search for contours that are passed between voices, but this would be incredibly taxing in terms of processing. Instead, the sequential ordering of voices (1,2,3) in P is slightly compromised by moving the last seven sixteenth notes of V2 in measure 26 to V1 and the last eighth note of V1 in measure 26 to V2. Because P is ordered based on onsets and ignores offsets, the overlap of the F4 on beat 3 in V2 with the exchanged material is inconsequential.

In the revised P, an augmented V1 comprises the first 258 elements \((i=1–258)\), V2 is the next 257 elements \((i=259–515)\) and V3 the last 235 elements \((i=516–750)\). Making this alteration to the encoding of the Fugue into P produces a result closer to the expectation.
Table 9.6: The location and similarity for all iterations of the models in SEG

<table>
<thead>
<tr>
<th>Model Contour</th>
<th>Cardinality</th>
<th>Iterations</th>
<th>Coverage (cov)</th>
<th>% of P (n=750)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1 (Subject)</td>
<td>20</td>
<td>8</td>
<td>160</td>
<td>21.3</td>
</tr>
<tr>
<td>S2 (Counter-Subject 1)</td>
<td>20</td>
<td>6</td>
<td>120</td>
<td>16.0</td>
</tr>
<tr>
<td>S3 (episodic material)</td>
<td>16</td>
<td>4</td>
<td>64 (80)</td>
<td>8.5 (10.7)</td>
</tr>
<tr>
<td>Counter-Subject 2</td>
<td>12</td>
<td>5</td>
<td>60 (59)</td>
<td>7.9</td>
</tr>
</tbody>
</table>

The episodic material in S3 outperforms CS2 in the coverage evaluation. There is an even greater obstacle to CS2 being brought forth by the algorithm. Although Bruhn reports (and others would agree) that there are five iterations of the second counter-subject...
(CS2), one of these (mm.11–13, V1) is of a lower cardinality (hence the parenthetical 59 in Table 9.7) and otherwise modified. Two other iterations use octave displacement at one or more points (see Figure 9.26). CS2 is way down the list in the SEGPOOL with only two iterations that can be recognized by the CONTCOM. There are many other cardinality-twelve candidates with more repetition.

Measure 7, Voice 1:  
Measure 11, Voice 1:  
Measure 15, Voice 3:  
Measure 20, Voice 3:  
Measure 27, Voice 2:

Figure 9.26: All appearances of CS2

The variety of transformations applied to CS2 distorts the contour beyond a point recognized as equivalent at four degrees of adjacency, and even one (as in a CAS).

<table>
<thead>
<tr>
<th>Model</th>
<th>Cardinality</th>
<th>Measure</th>
<th>Meas. Position</th>
<th>Voice</th>
<th>Transformation</th>
<th>C+SIM (n=4)</th>
<th>C+SIM (n=-1)</th>
<th>Starting Pitch</th>
<th>Ending Pitch</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS2</td>
<td>12</td>
<td>7</td>
<td>0.625</td>
<td>1</td>
<td>Model</td>
<td>1</td>
<td>1</td>
<td>F5</td>
<td>G4</td>
</tr>
<tr>
<td>CS2</td>
<td>11</td>
<td>11</td>
<td>0.625</td>
<td>1</td>
<td>Syncopate</td>
<td>-</td>
<td>-</td>
<td>Ab4</td>
<td>Bb3</td>
</tr>
<tr>
<td>CS2</td>
<td>12</td>
<td>15</td>
<td>0.625</td>
<td>3</td>
<td>T4, OD</td>
<td>0.667</td>
<td>0.515</td>
<td>C3</td>
<td>G2</td>
</tr>
<tr>
<td>CS2</td>
<td>12</td>
<td>20</td>
<td>0.625</td>
<td>3</td>
<td>Octave Disp.</td>
<td>0.762</td>
<td>0.530</td>
<td>F2</td>
<td>C3</td>
</tr>
<tr>
<td>CS2</td>
<td>12</td>
<td>27</td>
<td>0.125</td>
<td>2</td>
<td>Epenthesis (App.)</td>
<td>0.977</td>
<td>1</td>
<td>F5</td>
<td>Ab4</td>
</tr>
</tbody>
</table>

Table 9.8: The location and similarity of all iterations of CS2. Sound change terms are applied from linguistics: syncope indicates the omission of a medial sound, epenthesis is the addition.

Table 9.8 includes C+SIM values for both a full (n-1) matrix and a four-degree window. The third and fourth appearance have low similarity under either measure, and the second appearance cannot be calculated because it is of a lower cardinality, and reduction of
redundant slices cannot fix this. A good way to characterize CS2 based on the four same-cardinality appearances is using Quinn’s C+ matrix (see Figure 9.27). Here each comparison represents the mean of four 0 or 1 value comparisons. There is a lot of consistency on the local level: the first three pitches maintain their relationship to each other in all iterations, as do the fourth through tenth. However, the relationship between these subsegments is not maintained in imitation.

![Figure 9.27: C+matrix with fuzzy values for the four cardinality-twelve appearances of CS2, consistent subsegments are bolded](image)

Recognizing the similarity of transpositions with frequent octave displacement is not a contour similarity task. Table 9.8 indicates there was some tolerance for octave displacement (using four degrees), two of the iterations of CS1 had displacement but they were still recognized as similar.

Although the segmentation algorithm failed to identify the second counter-subject, it confidently produced the subject, even without the modification of the P
needed to recognize the first counter-subject (the second segment in SEG). Model segments

### 9.6 Summary

While other music may pose unique challenges to the method, the intention for including variable parameters is that it can also be applied to a wide assortment of monophony and monophonic extractions from other textures. Schoenberg’s music is a good testing ground for finding contour recursion because the composer avoided exact repetition, a feature of popular music of which he disapproved (Jenkins 2016:3–5). Schoenberg’s music is certainly not alone in its recursive nature, Bach’s Fugue could not be reduced to a single segment family, but all iterations of the subject and first counter-subject were identified. The following are foreseen challenges to broader application accommodated in the initial formulation:

1. For Schoenberg, the analysis privileged candidate segments with close recursion, awarding “points” for iterations that are adjacent without overlapping. For tonal music, such as Bach, an alternative criteria of “coverage” is provided for segment evaluation (see Section 9.3).

2. These analyses were of pitch series of finite length. The ability to handle pitch streams of indefinite length is necessary for online processing or modeling real-time listening. Both of these applications make an adjacency radius around the focus (and comparisons to the past and future) less appropriate. A window of only pre-adjacent degrees can be used.
3. Analysis of monophonic recordings poses a unique challenge that cannot be addressed sufficiently here: segmentation of an audio signal into an ordinal pitch series. Proprietary software exists that can accomplish this task, notably Melodyne, but it is preferable to have more user control over the calculation of a pitch from a fundamental frequency analysis, as well as some signal-sensitive parameters. Different instruments, voices, musical styles, languages all present unique challenges in discretizing a continuous signal into pitch events.

Additionally, detailed exploration and discussion of the contour slice, contour level and the recursive segment model (CLSEG) is needed.

The processing load for the current algorithm is heavy. To approximate the ability of a trained music analyst to recognize what Morris terms “shape and referential affinity” is quite taxing for a computer. This load could be reduced some by knowing more about human perception of contour, which could in turn reduce the number of different analyses attempted with varying input parameters. What is lacking is a perceptual basis for how many degrees of adjacency are salient. One way to get more information is more behavioral studies. One could look for trends in similarity ratings for melodies that have the same contour at two but not three degrees of adjacency, three but not four degrees, four but not five degrees and so on. This is similar to methods used to study simple-adjacency by Dowling (1971) and Edworthy (1983), and Quinn for complex-adjacency (1999). Another method, similar to Edworthy (1985) is to ask listeners to make non-adjacent pairwise comparisons and look for a drop off in accuracy or response time. Experiments to reveal perceptual or cognitive limits of non-adjacent contour comparisons would provide greater explanatory adequacy for using a specific window size.
Experiments by Schmuckler (2010) compare contours of the same length of time instead of cardinality. No doubt, the duration of notes and metric placement affects salience. However, I prefer to consider them as parallel streams that are to some extent independent, like having a CSEG and a DSEG for the same segment (Marvin 1991). This reflects the role of contour in Schoenberg’s Op. 19, No. 4 as well as the tone-level systems of Niger-Congo languages that inspired this revision of contour theory. Further behavioral research would only lead to modest improvement in the method presented in this article because it is already clear from this and other analyses that a fairly small window (of two to six degrees) is most effective. My speculation is this is similar to human capacity for non-adjacency, which is without doubt variable.

The extent to which non-adjacent comparisons are perceptually salient is not known, but simple-adjacency has been thoroughly researched. A happy medium between simple-adjacency and exhaustive comparison reflects our knowledge of pitch perception and is productive analytically. Through automation, this analytical approach has potential for application to large corpora and to use machine learning to improve efficiency. Constraining the range of input parameter values to only the most productive would be much less costly in terms of processing than producing many analyses and ranking them. However, the flexibility of the window size and other parameters make this method extensible. The re-ranking of SEGPOOL evaluation criteria is only one way to adapt this method to other signals. Schoenberg’s Op. 19, No. 4 was selected as an object of analysis because of its brevity and to make a connection to prior work (specifically Morris 1993).

68 The folk songs are effectively time-warped through the transcription and may not be equitemporal in performance.
This does not reflect limitations on its use or efficacy. Unlike the COM-matrix, intended for music such as Schoenberg’s, CONTCOM and the segmentation algorithm are designed for application to diverse musical spaces, including Bach’s *Well-Tempered Clavier*, trumpet solos by Dizzy Gillespie (Chapter 10), Ìgbò choral music (Chapter 14) and Yorùbá praise poetry (Chapter 13). This modified toolbox for contour analysis maintains applicability to past analysis objects and holds promise for analyzing a wide range of monophonic sound and symbol.
10. REALIZATION IN A DYNAMIC REGISTER

Introduction

In Chapter 4, impressive intonational trends were observed in English and Yorùbá public speaking. Within a speaker's range, low tones may be higher than high tones, and vice versa. There are two ways to account for this phenomenon. First, if a disruption of an already initialized paradigm occurs (see Chapter 6), re-initialization of all tone levels may be required. Pitch reset often occurs when a breath is taken between utterances. Alternatively, a gradual but significant shift of tone levels may be accomplished through an accumulation of register transformations that are locally constrained. This is similar to the concept of voice-leading in music theory, but distinct in that it applies to a monophonic signal. In Chapters 6 through 9, I endeavored to demonstrate the relevance of the concept of phonological equivalence to melodic contour. In this chapter, I apply basic rules of voice-leading and selected transformational functions to model paralinguistic intonation. The next chapter draws more directly on David Lewin’s work; this chapter does so only indirectly, through Soderberg (1998), Santa (1999), and Hook (2007).

Lexical tone sequences and contour segments are amazingly generative. There is not just one possibility for the realization of a model, there are infinitely many. With broad constraints on the realization of lexical tones in speech and contours in music this infinity is reduced. Constraints may include fitting within an established paradigm like
initialized tone levels, a scale or harmonic progression, or more broadly the range of a
human voice or instrument. Setting frequency bands for each tone level is the simplest
way to realize speech tones. This is called a fixed register in this chapter. The rich
variation of paralinguistic intonation present in many (if not all) tone languages demands
an alternative model that accommodates more complexity. A dynamic register is
composed of frequency bands that shift over time through paradigm-preserving
transformations, similar to voice-leading prescriptions. The realization function presented
later in this chapter accommodates a mapping of a model contour to pitches within a
fixed or dynamic register. Formalizing ways in which a register may transform over time
is a necessary first step.

After initialization, a register may be dynamic within a finite pitch range,
represented by transformations between time-ordered polychords of fixed cardinality and
height-order. The polychord is transformed within a pitch space that may be discrete
(along a musical scale) or continuous (as in speech), but always has a finite range (e.g.
two octaves for a human voice). Height-order is maintained in each register
transformation. There is no octave equivalence because modular arithmetic has the
potential to change the order of pitches in terms of relative height. Height-order fidelity
means the highest element in an antecedent polychord is mapped to the highest element
in a consequent polychord, lowest to lowest, and each middle level(s) to corresponding
middle level(s). This excludes inversion as a function. Viable transformations include
logarithmic pitch transposition, wherein frequency ratios within the polychord are
preserved. However, the majority of the functions applied do not preserve intervallic
distance, including the warp function (Soderberg 1998) and modtrans function (Santa
A multiplying function that compresses and expands the intervallic content of the register is the primary means of changing interval content. I call this the *open/close* function in recognition of its bond to chord voicings. An open voicing has notes spread out and a close voicing has note close together. It has further semblance to open vowels where the mouth is open and the formants are far apart and close vowels that are the inverse. It is also a nod to Fela Kuti’s 1970 dance hit “Open and Close” in which listeners are instructed to open then close their arms. Clough’s $J$-function multiplies a fraction made up of coprime integers by positive integers from 0 to the lower of the coprimes minus one, then floors (rounds down) the values. The *open/close* function is similar, but with some key modifications. The $J$-function is specifically intended for generating diatonic scales (Clough and Myerson 1985, Clough and Douthett 1991), but it has been used as an alternative to neo-Riemannian transformations in Douthett (2008), Tymoczko (2013) and Yust (2013). These recent contributions are the most relevant here, particularly Tymoczko’s proposal, sustained by Yust, that nearest integer rounding be used instead of floor rounding. The main divergence here is that octave equivalence (modulo 12) is not applied after a transformation. The $J$-function is applied to pitch classes while the *open/close* function is applied to absolute pitches representing a physical register. In all there are four register transformations proposed: *modtrans*, transposition, *open/close* and *warp*.
10.1 Modeling Intonation

Figure 10.1 brings back an example from Chapter 4, three repetitions of a phrase by Yorùbá preacher Peter Owadayo. Each point is a mean pitch value for a syllable. Lines are added to the contour plot connecting all the high tones in blue, all the mid tones in green and all the low tones in yellow. There are two low tones so that trend is simple with a fixed slope. The trend lines for the high level and mid level are more complex. They are drawn lines and do not represent a regression. Clearly, no linear function can account for all of the high tones or mid tones within these phrases.

Intonation trends are fairly easy to illustrate but difficult to quantify. In a 1983 paper, “A Generative Model of Intonation,” Eva Gärding of the Lund School draws gridlines between corresponding points. Unlike the contour plots in Figure 10.1, Figure 10.2 presents continuous f0 plots of idealized, not actual speech. The dotted lines connect phonologically-equivalent points, not necessarily lexical tone.
Most of the plots in Figure 10.2 show gentle declination, but the top-right plot does not. Because the grids are linear, when the intonation trend changes direction, the gridlines must be broken. While the drawn lines in Figure 10.1 cannot be neatly quantized, linear gridlines between two time points can be. Figure 10.3 extends the broken gridlines from the top-right of Figure 10.2 and adds red vertical lines to indicate time points A, B and C.
The absolute pitch of each gridline is changing, the spacing between the two remains constant. This makes the relationship of A to B, B to C, and A to C similar to a chromatic transposition in music. The same value could be added to both gridline values at point A to yield the values at point B, and so on. This vertical sampling is very similar to the piecewise constant register presented by Ladd (1990:40).
In Figure 10.4, the register is of a constant span but it is abruptly shifted by a small interval up or down by new events outside of the register. Intonation may indeed be more continuous, as in the lines drawn on Owadayo’s sermon or Gärding’s gridlines. However, in order to have non-linear transformations of a register, the signal must be discretized. Ladd’s piecewise time discretization is an option, similar to the lexical-order window used in Chapters 6–9, but without hopping. The other option is pitch discretization: conforming frequency values to integers along a logarithmic pitch scale or other musical scale. Both time and pitch discretizations are used here in formalizing a dynamic register.

If a dynamic register is time discretized to measure intonation change, it is important that the piecewise portions are congruent so that they may be related through mappings from one to the other. The temporal scales from Chapter 6, particularly the lexical-order and phrase-order windows, are useful here. Ladd’s piecewise registers in Figure 10.4 are lexical-order, but one could also calculate phrase-order registers.

Figure 10.5: Phrase-order registers of Owadayo excerpt
Figure 10.5 shows phrase-order registers for the excerpt of the Owadayo sermon (shown in Figure 10.1). These could be labeled $r_{p1}$ (register for phrase 1), $r_{p2}$ and $r_{p3}$. In each repetition, the register is compressed (the span is reduced) and raised (height is added). In the next section, transformations are introduced to quantify these changes.

10.2 Intonation Prototypes and Register Transformations

Two- and three-level contours, as found in Niger-Congo languages, may be realized in pitch space through a dynamic register that links the underlying levels to different pitches at different times. In this section, a dynamic register that transforms over time is formalized.

![Figure 10.6: (Left to Right) Two-level system; Three-level system; Generic 7–pitch space; 2–octaves of a pentatonic scale.](image)

Each type of register transformation is presented in both continuous and discretized pitch space. In the discretized version, the pitch height of each level is constant within a lexical-order window. The intonation prototypes and transformations in the following subsections could be applied to registers at various temporal scales (lexical, phrase, section) and to tone-level systems or contours of different sizes. To talk about the shifting of lexical tone levels and contour levels, the lexical-order register is appropriate and is
used here. If a contour has only lexical-order specificity (e.g. two or three levels), the best
temporal scale for snapshots of the dynamic register is also the lexical order. The
prototypes are idealized and no single model can represent actual speech or music except
in rare cases. The transformations may be freely applied and combined as necessary from
one piecewise register to the next to describe actual speech and music. Constraints on the
application of register transformations are presented in Section 10.3.

The colored lines in the intonational prototypes for continuous pitch space (as in
speech) are based on a three-level system (as in Yorùbá). The three colored lines
represent the three tone levels: low-mid-high=yellow-green-blue. An appropriate
quantization of the three levels is −1 for low, 0 for mid and +1 for high. This reflects
neutrality of the medial level and clarifies that low and high are marked in terms of
production. However, I will use [0,1,2] following a three contour-level operationalization
of tone levels, also similar to the do-re-mi heuristic (see Chapter 2). Contour levels are
not the same as tone levels but they are close, and importantly we can reliably calculate
contour levels (by making a fixed number of contour comparisons for each tone-bearing
segment) while tone levels are a bit more elusive. At least while considering the
prototypes, tone levels may be considered interchangeable with contour levels. Three
contour levels form clear perceptual categories: 0 is a local minima, 2 is a local maxima,
and 1 is a medial (or non-extreme) segment. Three levels are also useful for musical
contours.

A register may be discretized in terms of pitch by using a polychord. If there are
three underlying levels or voices, then the polychord will be a trichord. The polychord is
implied and only appears if all levels are realized at once, which cannot happen in a
monophonic signal. To illustrate what I mean by a dynamic register in the musical sense, I refer to the accompaniment in the opening of Mozart’s C-Major Piano Sonata K. 545.

![Figure 10.7: Left-hand accompaniment at the opening of Mozart K. 545](image)

Through consistent figuration of an underlying harmony, the Alberti bass makes the presence and harmonic rhythm of the chord progression apparent even in a nominally monophonic texture. The contour LHMH, \(<0212>\) in levels or \(<+−+>\) as a contour adjacency series (CAS), is applied to each chord in Figure 10.8.

![Figure 10.8: Active region corresponding to Figure 10.7](image)

In this case, the dynamic register is analogous to voice-leading of tertian harmony. All three levels appear in each iteration of the “word” (LHMH). In some forms of analysis, the inversions might be considered trivial and the roots or prime form might be favored, but here the voice-leading is very important.

10.2.1 Register Initialization: Specification and MODTRANS

Initialization is the process by which the underlying levels are first specified as pitches. Chapter 6 outlined a syntagmatic process to establish three tone levels further considered
in the next chapter. For now, I want to bypass the process itself and simply consider the mapping from generic levels to specific pitches. If there are three underlying levels, then three pitches will be initialized. There is a strong precedent for a generic-specific distinction in intervals from scale theory (see Clough 1979). For the Mozart example above, three levels [012] are first mapped to [047] (normalized to C4=0). In this mapping the generic change of level +2 (for L to H or 0 to 2) is specified as +7. The specification of a the pitch of each level also specifies the intervals of the contour.

The number of specific trichords in continuous pitch space is infinite, but for a piano it is 109,736. The formula for unordered combinatoriality is the binomial coefficient \( C_{n,k} = \frac{n!}{k!(n-k)!} \), where \( n \) is the cardinality of the space and \( k \) is the cardinality of chord and \( C \) is the resultant combinatoriality. If \( n=88 \) and \( k=3 \) then \( C = 109,736 \). If a composer is writing a piece for piano where one hand (presumably the left) is arpeggiating trichords using LHMH or some other pattern, a reasonable constraint is to restrict the trichords to the span of a tenth (16 semitones). This ensures each trichord may be arpeggiated by a single hand without shifting positions excessively. Restricting the span reduces the combinatorialty considerably to 8120. This may be calculated so: \( C_{n,k,s} = C_{s,k} + (n+1-s)C_{s-1,k-1} \), where \( s \) is the span in semitones. This constraint could be called maximum span. A Yorùbá speaker may have two octave range, but they are likely to use less than an octave for three levels within a lexical-order register. Likewise, a piano has 88 keys, but it is not idiomatic or conventionally musical to have an arpeggiated chord than spans two or more octaves. If register is constrained to a single range of a tenth (\( n=16, k=3 \)), not any tenth within the entire instrument range, this
reduces the combinatoriality to 560. If the maximum span is a tenth in diatonic space instead of chromatic space (n=10, k=3), there are only 120 possibilities, and so on.

In Chapter 4, normalization of tone levels to different voice ranges was discussed. In the continuous pitch space of a human voice, the only firm constraint on initialization is range. For some speakers this is as much as two octaves (the Yorùbá priests in Chapter 4), for others this may be less. Some speakers may have a large range but only use a small part of it at one time. Figure 10.9 illustrates two possible registers within the same range. Maximum and minimum pitch indicated by red lines, the three levels, Low-Mid-High, by yellow, green and blue respectively.

![Figure 10.9: Close (left) and open (right) spacing of fixed registers](image)

The plots in Figure 10.9 represent a narrow (close-spacing) register and a wide (open-spacing) register. In each, the register is fixed over time, not dynamic. There is no transformation articulated between the two. These registers are incompatible with paradigm preservation. The total span of the first is smaller than the distance between
adjacent levels in the second. None of the pitches of the three levels are shared—there are no “common tones.” If these were to represent fixed registers for contiguous phrases in speech or music, pitch reset would have occurred between them and each would have its own initialization. In any order, the precedent in terms of span and inter-level distance would make the second hard to understand, requiring an even stronger re-initialization of the second phrase than in the initialization of the first. A mapping between these two fixed registers may be defined using the open/close function and transposition. However, because the transformation does not fall within constraints, that will be elaborated in section 10.3, the paradigmatic orientation of a listener to the levels might not be preserved through that mapping. Transposition in particular may cause a disruption of the paradigm if the interval of transposition is large (see Figure 10.10).

Figure 10.10: Fixed registers high (left) and low (right) within the range

However, I want to distinguish transformations that are abrupt, weakening or destroying the paradigm and requiring reinitialization, from those that are gentle enough to preserve the paradigm without any latency. To make this distinction, I prefer to relate initialization
and reinitialization to Santa’s *modtrans* function: a mapping of corresponding step classes to a different scale (or modular system).

Let us define MODTRANS \((x, y, z)\) as a transformation that maps each step class of a musical entity in modular system \(x\) onto a corresponding step in modular system \(y\), where \(z\) represents the “point of synchronization,” the pitch class in the starting modulus that is interpreted as step-class 0 (1999:202).

Figure 10.11 shows a mapping of three levels \([0,1,2]\) to three chromatic pitches \([7,8,9]\) and to three pitches that could be part of a C-Major scale \([4,5,6]\) (where \(C4 = 0\)) or a G-pentatonic scale \([0,1,2]\) (where \(G4 = 0\)).

![Figure 10.11: MODTRANS function mapping contour space to adjacent step-classes in scales](image)

The three underlying levels unify the pitch mappings. There is some continuity because G is the lowest in both pitch mappings. However, because of the ambiguity between the Bbb as the high level in the first and A as the mid level in the second, reinitialization may be necessary. Smoother register transformations that preserve an initialized paradigm are explored in the following sections.

### 10.2.2 Transposition: Ascend and Descend

The first dynamic intonation prototypes are ascent and descent through transposition. The pitches in the initial register are increased or decreased by the same amount in all the levels over time. This can be expressed as \(r_t = r_0 + k*t\), where \(r_0\) is the initial register, \(r_t\) is
the register at time $t$, and $k$ is a constant that is positive for ascent and negative for descent.

![Figure 10.12: Ascending (left) and descending (right) dynamic registers in continuous pitch space](image)

After initialization, a speech register can move up or down in continuous pitch space. But for a discretized space, such as a fretted or keyed instrument or music constrained to a scale, this is different. Figure 10.13 shows the descent prototype within a two-octave pitch space conformed to the G-pentatonic scale. The initial “point of synchronization” for the `modtrans` function initialization is D4, a mapping of levels [0,1,2] to pitches [D5, E5, G5]. This register is then transposed down one scale step at a time ($T_{-1}$) to the bottom extent of the range. Because the transposition is constrained to the pentatonic scale, pitches such as C5 are skipped.
The initial register in Figure 10.13 is a conjunct mapping of three levels to adjacent scale degrees. If transposition is applied without other transformations, the dynamic register will remain conjunct. The initialization could also be disjunct (as in Figure 10.14), in which case applying the same transposition by step downward would maintain a disjunct mapping. There are fewer disjunct trichords found within a finite and discrete range (e.g., two octaves of G-pentatonic). Thus, fewer transpositions can be made overall within the range.

Although the disjunct mapping runs out of space after fewer transpositions in the same direction, it is preferable because there is overlap without level ambiguity between +1 and –1 transposition. In Figure 10.13, the pitch of the mid-level in the antecedent register becomes the pitch of the high-level of the consequent, and the low-level becomes the mid-level. Depending on which level appears first in the new register, this may or may not lead to an ambiguity. If high raises or low down-steps first, then there is no problem.
Constraints on register transformation and contour realization are elaborated in Section 10.4.

### 10.2.3 Expansion and Compression: Open/Close Function

The *open/close* function sets an anchor point (a zero value within the register) and then multiplies all elements by the same factor. Figure 10.15 shows an expansion around a middle anchor between the left and right side plots. This can be expressed as \( r_2 = (r_1 - p_m) * k + p_m \). When \( k \) is greater than one, an opening (expansion) of the register is produced. In Figure 10.15, a \( k \) of 3 produces the right side from the left side.

![Figure 10.15](image)

Applying the *open/close* function without transposition never completely distorts the paradigm because the pitch of one level (the anchor) is preserved. As a continuous function of time this can be expressed as \( r_t = (r_i - p_a) * k t + p_a \). The anchor is indicated by \( p_a \), which could be the pitch of any of the elements in the register’s polychord.
Figure 10.16 shows continuous expansion and compression with low-, mid- and high-anchoring. Because there are three levels there are three possible anchors.

Figure 10.16: Six prototypes for open/close function in continuous pitch space
The possibilities of different anchors, and of values greater or less than 1 for the constant k, yields six prototypes associated with the open/close function in a three-level system. Similar to the transposition prototypes, open/close prototypes may be represented in discretized pitch space. Figure 10.17 and Figure 10.18 show compression with a top anchor and compression with a bottom anchor respectively, in both upper and lower octaves of a two-octave pentatonic scale.

![Figure 10.17: Compression-Ascent: Disjunct-to-conjunct mappings of 3-levels in lower octave and higher octave (no overlap between octave registers)](image1)

![Figure 10.18: Compression-Descent: Disjunct-to-conjunct mappings of 3-levels in higher octave and lower octave (no overlap between octave registers)](image2)

Each compression pattern of four in all, shows the effect of nearest integer rounding. In the compression-ascent transformations (Figure 10.17), the initial register in generic step intervals normalized to the top anchor is \([-5, -2, 0]\). If the initial register is multiplied by 4/5 (an irreducible fraction of coprimes), then it becomes \([-4, -1.6, 0]\). If multiplied by

69 Initially Clough and Douthett applied a floor function to round down all decimals. Tymoczko suggests an additive constant of .5 before using a floor function (nearest-integer rounding) (2013:129). In the same volume of the *Journal of Mathematics and Music*, Yust sustains Tymoczko’s alteration and notes the similarity to conventional rounding and the “Tie-breaker rule” (2013).
3/5 (also a fraction of coprime integers), it becomes \([-3, -1.2, 0]\). Finally, 2/5 produces \([-2, -0.8, 0]\). If floor rounding is applied to all of these consequent registers before normalization to the anchor pitch is removed, they turn out quite differently than when floor rounding is applied after the pitch height is restored. An additional concern with the floor function is that different intervals between levels will be produced depending on whether the anchor is at the top, middle or bottom. Nearest integer rounding avoids effects of different anchors or changes in the order of operations. It can be applied to normalized registers with negative values or registers measured in absolute height values (e.g. C4=60) yielding the same results.

Figure 10.19 shows expansion of an initial conjunct register in three different scales. All three different anchorings are included with the “zero” value at the low, mid or high level respectively.

Figure 10.19: Expansion using the open/close function in different scales
The expansions rely on Clough’s concept of generic step intervals (1979) but ignore octave equivalence. The difference in effect for each scale is quite stark.

10.2.4 Spacing Transformation: Warp Function

Tranposition changes the height of the register and the open/close function changes the span. The last function does not alter the composite range of the register, only the spacing of levels within it. The warp function introduced by Soderberg (1998:213) subtracts some value from one interval between the elements of a polychord and adds it to another. It operates on the Adjacency Interval Series (AIS) of set theory. For trichords there are only two intervals between vertically adjacent levels, for tetrachords there are three, and so on. The warp function cannot be applied in a two-level system because there is only one interval. Unlike the open/close function, which increases or decreases all intervals by the same factor with the spacing preserved proportionally, here the sum of the AIS (the span of the register) is preserved and the interior spacing is changed. One interval gets larger at the expense of the others. If applied to a trichord, one interval takes magnitude from the other interval, if there are more elements, the augmenting interval may take magnitude from one or multiple other intervals. The warp applied is indicated through AIS changes. The function \( \text{warp}(r,<-1,+1>) \) indicates the lower interval of a three-level register \( r \) is decreased by 1 semitone or scalestep and the upper interval increased by 1. There are only two prototypes for this function in a three-level system: the transformation where the lower interval gets larger, and the transformation where it gets smaller, both visualized in Figure 10.20.
10.3 Applying Register Transformations

To illustrate the utility of register transformations, I will start with an uncomplicated example. Returning to the Alberti bass example from Mozart K. 545, we can now apply machinery to explain the voice leading as register transformations. Using materials from the previous chapters, I can complete an analytical loop, including a reductive mapping from pitch space to contour space and a generative mapping in the inverse direction, from contour space to pitch space.

10.3.1 SEGMENTATION

From an unsegmented pitch series (P) for the excerpted material, we can produce a
CONTCOM$_2$, with pairwise comparisons one-degree forward and behind. This produces the potential for three contour levels $[0, 1, 2]$, however because of the frequent “pivot” notes, the contour level series produced is too reductive (see Figure 10.22).

The contour slice sums (contour levels) below the CONTCOM$_2$ in Figure 10.22 fail to present important features of the contour. The only recursive contour in the CONTCOM$_2$ is a dyad $<02>$ with multiplicity 15. This accounts for most of the excerpt, however, a dyad is too basic. In the last chapter, using a window of four degrees and setting a minimum cardinality of four was presented as a preferred method for contour level analysis.

In a CONTCOM$_4$ of the same pitch series $P$, there is only one recursive segment $<0222>$ which has a multiplicity of 6. The leftover pitches not included in the CLSEG model are in groups of four which can form segments in themselves. Another way to arrive at an optimal segmentation is to consider jagged CONTCOM blocks (as the algorithm did),
which are all the same even though the levels are not. Either contour levels segments (CLSEGs) with groupings of leftover pitches or jagged blocks yields the same segmentation. With a cardinality for all segments \( n=4 \) and index for the start points \( I=[1,5,9,13,17,21,25,29] \), \( P \) can be divided into a collection of segments \( S \).

\[
\text{makeSeg}(P, I, n) = \{[60,67,64,67], [60,67,64,67], [62,67,65,67], [60,67,64,67], ... [60,69,65,69], [60,67,64,67], [59,67,62,67], [60,67,64,67] \} = S
\]

10.3.2 REGISTER TRANSFORMATIONS

By reducing each melodic segment to a verticality, a collection of registers \( R \) is gleaned with a register for each segment in \( S \). \( R \) is a lexical-order discretization of a continuous dynamic register.

\[
\text{vert}(S) = \left\{ \begin{array}{c} 67 \quad 67 \quad 67 \quad 67 \quad 69 \quad 67 \quad 67 \quad 67 \\ 64 \quad 64 \quad 65 \quad 64 \quad 65 \quad 64 \quad 62 \quad 64 \\ 60 \quad 60 \quad 62 \quad 60 \quad 60 \quad 60 \quad 59 \quad 60 \end{array} \right\} = R
\]

![Figure 10.24: discretized dynamic register (R) for Alberti bass excerpt](image)

Every element of \( R \) has the same cardinality (3), so a one-to-one mapping can be defined between every pair in \( R \). It makes sense to define these chronologically: \( r_1 \rightarrow r_2 \), \( r_2 \rightarrow r_3 \), ..., \( r_{n-1} \rightarrow r_n \). In the equations below, square brackets \([\]\) indicate nearest integer rounding, \( h \) and \( l \) (small \( L \)) are the anchors.
\[ r_1 = r_2 \]
\[
[(r_2 - h_2) \times 5/7] + h_2 = r_3
\]
\[
[(r_3 - h_3) \times 7/5] + h_3 = r_4
\]
\[
[(r_4 - b_4) \times 9/7] + l_4 = r_5
\]
\[
[(r_5 - b_5) \times 7/9] + l_5 = r_6
\]
\[
\text{warp}(((r_6 - h_6) \times 8/7] + h_6), <2,+2>) = r_7
\]
\[
\text{warp}(((r_7 - b_7) \times 7/8] + b_7), <1,+1>) = r_8
\]

The first two registers are the same, but all other adjacent pairs are different and require the open/close function at the very least. The last two pairs also require the warp function. A common tone is present between each verticality so transposition is not necessary. The common tones may be used as an anchor in each application of the open/close function. The register transformations can be viewed more clearly in Figure 10.25. In this figure, the registers are normalized to C4=0 and (=) indicates the anchored pitch.

<table>
<thead>
<tr>
<th>AIS</th>
<th>&lt;4,3&gt;</th>
<th>&lt;3,2&gt;</th>
<th>&lt;4,3&gt;</th>
<th>&lt;5,4&gt;</th>
<th>&lt;4,3&gt;</th>
<th>&lt;3,5&gt;</th>
<th>&lt;4,3&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>9</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Mid</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Low</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Multiplier</th>
<th>5/7</th>
<th>7/5</th>
<th>9/7</th>
<th>7/9</th>
<th>8/7</th>
<th>7/8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Warp</td>
<td>&lt;2,+2&gt;</td>
<td>&lt;1,+1&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 10.25: Mozart Alberti bass voice-leading modeled through \( f \)-function and warp function

10.3.3 MODEL SEGMENT

Chapter 9 demonstrated that a CLSEG with inner contextualization (allowing the window of comparison to wrap inward) made it possible to describe all iterations of Bach’s C-
minor fugue with one model. In Chapter 10, wrapping inward to contextualize a model
CLSEG of the Schoenberg primary segment would not work, so ranges of CL values
were used for undercontextualized sums. In both cases, some version of the CLSEG
succeeded to model equivalence where the CSEG failed. In this case, after using
windowing to arrive at the segmentation, a conventional CSEG may be used to model the
recursive contour segment. The same COM-matrix is produced by all melodic segments
in S.

\[
com(S) = \text{com}([60,67,64,67], [60,67,64,67], [62,67,65,67], [60,67,64,67], ... \\
[60,69,65,69][60,67,64,67], [59,67,62,67], [60,67,64,67]) = \begin{bmatrix}
0 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 \\
0 & 0 & 0 & 0
\end{bmatrix}
\]

Columns of the COM-matrix can be summed to produce the CSEG <0212>.

\[
\text{sum} \left( \begin{bmatrix}
0 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 \\
0 & 0 & 0 & 0
\end{bmatrix} \right) = \langle 0212 \rangle = c
\]

The recursive CSEG may be visualized as so:

![Contour figure for Alberti bass pattern](image)

Figure 10.26: Contour figure for Alberti bass pattern
10.3.4 **Realization Function**

The realization function combines a register with a contour. The cardinality of theregister’s polychord and the number of unique elements in the contour must be the same.

If the realization function is only supplied a single verticality or a single contour, then themapping is from contour pitches to real pitches. This is realization within a fixed register.

\[
\text{real}(R(1), s) = \text{real} \left( \begin{array}{c} 7 \\ 4 \\ 0 \end{array}, \langle 0212 \rangle \right) = [0, 7, 4, 7]
\]

*Figure 10.27: Realization on a C-major triad*

\[
\text{real}(R(5), s) = \text{real} \left( \begin{array}{c} 9 \\ 5 \\ 0 \end{array}, \langle 0212 \rangle \right) = [0, 9, 5, 9]
\]

*Figure 10.28: realization on an F-major triad in second inversion*

A realization of contour pitches in a fixed register is a one-to-one mapping. A realization of contour levels in a dynamic register is not a one-to-one mapping, it requires theCLSEG be divided into CSUBSEGS. This will be demonstrated in the last section. In theMozart example, the same recursive contour <0212>, the word LHMH, is sent to everydiscretization of a dynamic register R with cardinality 8. By realizing the contour c withall elements in R we produce P, but P is now enhanced by L, a series that highlightscontour equivalence between notes (color-coded in Figure 10.29).

\[
\text{Real: } R \times c \rightarrow P, L \\
\text{or} \\
\text{real}(R, c) = P, L =
\]
This L is different than the contour levels produced by summing slices in CONTCOM\textsubscript{2} or CONTCOM\textsubscript{4}. It is produced by stringing together CSEGs (of contour pitches), but I would still call it contour-level space. Once CSEGs are combined, contour pitches become contour levels because different pitches may have the same contour value assignment. Contour pitches are unspecified but have distinct absolute pitch height, much like scale degrees.

Realization in a fixed register with the same number of pitches as there are unique heights in the underlying contour is trivial. Welmers (1973) avers that one can map a tone sequence in a three-level language to a C-Major triad. This is the initial register in the Mozart, but it does not remain static, it morphs. Despite the dynamism, the level equivalence throughout is doubly secure, both through smooth voice-leading and the insistent repetition of the Alberti bass contour pattern. Either one of these features alone would be sufficient to sustain the paradigm, together these features make the levels explicit from the moment of full initialization. The Alberti bass pattern could indeed be a tone-to-tone mapping of a word with lexical tone LHMH and it would preserve the meaning well. On top of this, the pattern LHMH, with all three levels presented immediately, is a clear initialization (see Chapters 6 and 11).
10.4 Constraints: Ese Drum Example

A paradigm of phonological equivalence may be sustained through transformations that use anchoring (have common tones). If there are no anchors, quasi-voice-leading constraints on register transformations are all the more necessary. Additionally, there are realization constraints: high tones should be the first to raise and low level the first to lower. Constraints on register transformations and contour realization may be demonstrated through possible mappings of the two tone-levels of Ìgbò to sets of drums, each with fixed-tension and a different pitch height, unlike the variable-tension dundún drums of the Yorùbá. In Meki Nzewi’s article “Beyond Song Texts--the Lingual Fundamentals of African Drum Music” (2001) several sets of talking drums are detailed, including ese, ukom and mgbá, all found among the Ngwa group of the Ìgbò. Each of these instruments consist of a drum row with distinct tunings. The ese has four tuned drums and a variable-pitch “spirit drum,” the ukom and mgbá each have nine tuned drums in a row (2001:95). Nzewi calls the “instrumental processing of language” metasong (2001:91).

Nzewi’s study of the tuned drum rows of Ìgbò, in which a two-level tone system can be realized dynamically in a larger inventory of pitches, is consistent with the dynamic register I propose. Each component drum in these sets does not correspond to a tone level (H or L) all the time, but does at points in time. Thus, the speech surrogacy of tuned drum rows is an example of a discretized dynamic register.

---

70 Co-authored by Israel Anyahuru and Tom Ohiaaraumunna.
71 I also saw a demonstration at the 2014 African Literature Association conference in Johannesburg, South Africa.
In Figure 10.30, all dyads in a generic seven-pitch space are shown. This does not correspond specifically to any of the tuned drum rows Nzewi describes, but instead is a generic model based on Nzewi’s documentation of various tuned drum rows. This finite and discretized pitch space limits the number of possible registers and makes the transformations between them apparent.

![Figure 10.30: Mappings of two tone-levels to a seven-pitch collection](image)

While the drums Nzewi describes do have tuning systems, the lines and spaces of the staff in Figure 10.30 do not correspond to any particular tuning for this composite representation. The pitches are ordered in height but otherwise unspecified similar to contour pitches, in other words generic steps and intervals. If [0] is the lowest pitch, the first registral dyad in Figure 10.30 is [0,1] and the last is [0,6]. The number of dyads in a seven-pitch space is the sixth triangular number (C=21). Each dyad represents a potential realization of two tone-levels, however they cannot occur in any order if the paradigm is to be preserved.

A sequence emblematic of “terracing” in Ìgbò speech is all dyads with [0] as the lower element, arranged from [0,6] to [0,1] in descending interval size. No transposition is needed with a constant low anchor and the warp function is irrelevant to dyadic registers. Only the open/close function is used. To take the first register to the second, multiply it by 5/6, the second to the third, times 4/5, third to fourth, times 3/4, fourth to fifth, times 2/3, and fifth to sixth, times 1/2.
A static low tone and falling high tone create a wedge-shape. Often in speech, high tone is *downstepped* in the course of a phrase, and then reset to near the initial high pitch with each new phrase. Welmers (1973) and Clements (1979) both provide descriptions of “tone-terracing” similar to this. The interval of downstep is not constrained within speech, but it is here because of the discrete pitches of the drum set. Also the number of downsteps which can occur without high-raising or total reset is constrained: there can only be five downsteps, moving the high-tone from the highest pitch (6) down five times to the pitch just above the lowest (1).

Realizing a string of HL or LH contour dyads in the wedge-shaped dynamic register creates no ambiguity in either case.

For other registral sequences, this is not the case. Ascending conjunct dyads are sensitive to the contours realized in them. An HL contour applied to all registers in the collection is
fine because high leads the raising. However, realizing an LH contour leads to ambiguity because the low tone raises first.

Rahn calls different generic intervals with the same specific interval size (e.g. an augmented fourth and a diminished fifth) an ambiguity (1991:36). The assignment of adjacent tones of different level to the same pitch is also ambiguous. For a descending sequence of conjunct dyadic registers, the low tone should lead the downstepping to avoid ambiguity. When no anchor is present, constraints on which tone may appear first in a transformed register are necessary.

### 10.4.1 SUMMARY OF CONSTRAINTS, CONTRAFACTUM VOICE-LEADING

Constraints on transformations of a dynamic register overlap with common practice voice-leading rules to some extent, but not fully. The dynamic register is similar to implied harmony not proper homophony, wherein multiple voices are always present, sounding at once and moving at the same time. Constraints on the dynamic register and contour realization preserve initialized levels and avoid a need for reinitialization. Huron
(2001) presents perceptual phenomena that correlate to common practice voice-leading rules. Based on empirical evidence, perceptual principles that underly voice-leading practice and include pitch proximity, tonal fusion, and co-modulation. The realized signals under consideration here are stratified by phonologically-equivalent levels but ultimately are monophonic, so tonal fusion and co-modulation are not pertinent. An analogy between levels and voices is particularly salient in the Mozart accompaniment because a contour pattern embellishes a harmonic progression. However, in the case of the Ìgbò tuned row drums, a harmonic progression in the Western sense is not implied. Pairs of drums are surrogates for speech tone levels of a single speaking voice. Voice-leading rules strongly tied to vertical sonority, which Huron demonstrates relate to minimum masking, tonal fusion, and co-modulation, are not relevant to the ìse drum example or, more generally, register transformations. For example, rules about parallel or similar motion are not germane. Conversely, perceptual effects of pitch proximity, and non-proximity, and associated rules are highly relevant to paradigm-preservation within a dynamic register. In fact, the whole essence of levels might be said to stem from the pitch proximity principle (see Chapter 5 & 6). George Miller and George Heise reported a “trill threshold” (1950, 1951).

Two tones of different frequencies were alternated successively five times per second. When the difference in frequency was small, the alternation sounded like a continuous up-and-down movement of the pitch. When the difference in frequency was large, the alternation sounded like two unrelated, interrupted tones. Many studies have reported consistent findings since (Huron 2001:19). Huron summarizes the “pitch proximity principle” as “the coherence of an auditory stream…maintained by close pitch proximity in successive tones within the stream”
The pitch proximity principle also informs constraints and methods to promote contiguity and avoid reinitialization in contour realization.

Constraints and methods concerning register transformations include:

1. No register can exceed or extend beyond the full range;
2. Anchoring a level when the register compresses or expands provides continuity;
3. Do not transpose by an interval greater than the intervals between levels in the input register, such that L would be higher than M.

Constraints and methods concerning contour realizations within a dynamic register:

1. If ambiguity is possible between the pitch of a level in the antecedent register and the pitch of a different level in the consequent register, downstepping low or raising high should expand or shift the register, in other words the level at the extreme in the direction of the shift should lead the shift;
2. contour recursion, like an “ostinato” repetition of a single contour (e.g. “L’oruko Jesu” repetition, Alberti Bass), creates expectation and may aid in transferring a paradigm to a new register that may otherwise be ambiguous, this is somewhere between a smooth register shift and full reinitialization.

The number of realizations for any contour is infinite, but can be reduced to a more plausible array of possibilities through these constraints and methods. The constraints and methods here are similar to the markedness and faithfulness constraints in optimality theory.
10.5 Musical Gestures in Jazz Improvisation

Julian Hook states “any two contour-equivalent lines are related by transposition between suitably defined GISes” (2007:19). By contour-equivalent, Hook means contours that produce the same COM-matrix and CSEG class. Hook makes apt use of parodies of Bach and Beethoven by Nicolas Slonimsky (1894–1995) to demonstrate this transpositional equivalence, similar to the modtrans function discussed earlier (Santa 1999). In the formalization introduced here, a Bach melody and a Slonimsky parody are the same contour realized in different fixed registers. The number of unique values in the contour is the same as the number of unique pitches in the respective registers the contour is realized in. The methods in this chapter accommodate this uncomplicated realization or iterating it in a series of piecewise registers as in the Mozart example. A further possibility is to proliferate (or even reduce) the vertical realization of a single underlying contour by chopping it up into contour subsegments and spreading them across consecutive discretizations of the dynamic register. This makes it possible to map three levels to a number of different spaces, which can be demonstrated through application to a trumpet solo by Dizzy Gillespie.

Figure 10.34: (a) mapping of 3-level contour space to two-octaves of a pentatonic scale (treble clef omitted) to a specified register, (b) mapping of three levels to 2–octave pentatonic scale without the constraint of an active region, (c) mapping of three levels to a smaller space.
The dynamic register explanation applied to Alberti bass in Mozart and ìgbò tuned drum rows can also be applied to improvisational gestures in jazz. The segmentation algorithm presented in the previous chapter was applied to a transcription by Jacques Gilbert available on the EJMA Woodwinds Website (2015). The transcription is one among many available on the website as MIDI files, a great resource for computational analysis. The file analyzed here is a transcription of the trumpet solo from Dizzy Gillespie’s recording of the standard “Blue Moon” by Rodgers and Hart.

A reduction of redundant contour slices was applied before the pattern-finding algorithm, making it possible to find similar contours of different cardinality. All of the melodic segments in Figure 10.35 correspond to one of the reduced continuous matrix segments in Figure 10.36. Because redundant contour slices were reduced, all of the melodic segments in Figure 10.35 produce cardinality-six contour segments. The retrograde is all
cell values changed and rows in reverse order. The retrograde-inversion is all rows in 
reverse order from the primary segment, or all cell values inverted from the retrograde.

Figure 10.36: Contour segments as they appear in the reduced continuous matrix

The +1 and –1 rows of the primary segment matrix are the immediately adjacent 
comparisons.

Figure 10.37: Three-level contour reduction

By excluding non-adjacent comparisons, we get a three-level reduction, with a sum of 0 
representing a local minimum, 1 a medial pitch, and 2 a local maximum. In terms of low 
(L), mid (M) and high (H) levels this contour is <LMHMLM>, in integers it is 
<012101>. The reduced form is consistent with all of the realizations in Figure 10.35: an 
uninterrupted ascent, uninterrupted descent and a hook (single ascent) at the end.

The recursive contour produced from automated analysis of the solo corresponds 
to contour recursion in the melody of the song itself (Figure 10.38). They produce the 
same primary recursive segment.
The melody includes a sequence based on the same contour, but with a repeated pitch. The repeated pitch accommodates the text of the song, but in instrumental variations on the theme, it is not necessary. The last iteration of the sequence in Figure 10.38 is like a tonal answer, with do asserted strongly as the first note and final note.

With repetition removed and chains of ascents or descents reduced, much of the song’s melody and Gillespie’s improvisation on the tune can be modeled with a single recursive contour with three levels.

The first two presentations of this contour in the melody are simple realizations. A three-level contour is sent to a three-pitch fixed register.
Other realizations are more complex. Because it is fairly long, 6 elements in reduction and as many as 11 in realization, this recursive contour is not lexical-order, but phrase-order.

The segment can be subdivided into lexical order segments, LMH, HML, LM. Here M is merely a placeholder for strings of medial pitches between local extremes (or pivots). The middle element in both of the first two subsegments can be expanded, making a longer ascent or descent (e.g. LMMMH or HMMML). The subsegments are overlapping in the full segment with an anchor between each, represented here by parentheses: LM(H)M(L)M. The subsegmentation of a phrase-order segment allows it to be realized across multiples states of a dynamic register. Similar to the Mozart example, the open/close and warp functions can be applied as transformations between states of a
dynamic register. The difference here is that realizations of the contour segment may be spread across multiple states of the register instead of sending the contour to single states.

The final iteration of the sequence in the melody (Figure 10.38) has a span that widens after the first three notes. The first and second registers in the dynamic register R are related through a top anchor and a factor of 7/4.

\[
\text{real } (R, C) = \text{real} \left( \begin{bmatrix} 69 & 69 & 69 \\ 67 & 65 & 65 \\ 65 & 62 & 62 \end{bmatrix}, \{012, 210, 01\} \right) = [65, 67, 69, 69, 65, 62, 62, 65]
\]

\[
\]

The subdivision of the contour creates duplicate notes at the anchor points that must be deleted to produce the realization in Figure 10.42.

All of the realizations in Gillespie’s improvisation are complex, requiring subdivision of a contour and a dynamic register. In Figure 10.43, melodic segments from the solo are reduced to cardinality-six and normalized to Middle C (C4) as the starting pitch. The transposition is noted at the top of the transformation diagram for each melodic segment. Like the diagram for Mozart’s Alberti bass figure, discretized registers are presented vertically with anchors indicated by the equals sign (=). The multiplier for the open/close function is indicated between each register and a warp (indicated with angle brackets < >) is noted if necessary. Each measure represents a single realization (in a reduced and normalized state) and they are not contiguous. They are in order with the

324
measure number, the contour transform (e.g. P=prime, RI=retrograde inversion) and the transposition indicated.

This application of anchoring, with a pitch acting as a pivot in direction and an overlap between adjacent registers, carries a high level or low level into the next register. In the subsegmentation, H and L are not only pitch extreme but time extreme within each register, on the cusp of a register transformation.

The segments in Figure 10.43 are at an intermediary stage of realization. Adding
transposition back in is trivial. To reach the surface level in Figure 10.35, the enclosed medial levels (elements 2 and 4 of the reduced contour) must be expanded in many of the iterations. An example of horizontal expansion is in Figure 10.44, in which the first subsegment <012> is expanded. True to BeBop-style improvisation, the expansion of the medial-level note is sensitive to chord changes, but otherwise quite free.

\[
R = \left[ \left( r_1 \times \frac{1}{3} \right) + 8 \right] = r_2
\]

\[
(r_2 \times \frac{1}{2}) = r_3
\]

\[\text{real} (R, C) = \text{real} \left( \begin{bmatrix} 12 & 12 & 10 \\ 5 & 10 & 9 \\ 0 & 8 & 8 \end{bmatrix}, \langle 012, \langle 210, \langle 011 \rangle \rangle \right) = \begin{bmatrix} 0.5, 12, 12, 10, 8, 8, 9 \end{bmatrix} = P\]

\[\text{delrep}(P) = [0.5, 12, 10, 8, 9]\]

\[\text{expand}_{\text{Fmaj/chromatic}} ( -3) = \]

Figure 10.44: Realization in a dynamic register with expansion of the medial level

In Chapters 8 and 9, contour levels were used to explain variation in related contours that conventional contour analysis cannot, such as tonal answers. Here, contour level models are shown to be highly generative. The inverse mapping of CLSEGs is considerably more complex than the realization of CSEGs because contour levels do not have a one-to-one correspondence to pitch heights. The relationship between contour-level space and
pitch space is complex. A dynamic register serves as a conduit between the two: shifting, compressing, expanding and warping over time.

Figure 10.45: Visualization of Figure 10.44
11. Interlevel Cycles and Robust Signals

Introduction

This chapter addresses the relationships of non-equivalent contours to each other. Here, non-equivalent indicates contours do not even form the same contour adjacency series (CAS). Uniform cycles are used to generate intervals between levels, called *interlevels*. Circles and cycles based on repetition of the same transformation (as in the circle of fifths) or a series of transformations (as in Richard Cohn’s hexatonic cycle, 1996) play an important role in music theory. Instead of transposition or Riemannian transformations, binary state changes are used based on Lewin (1995). The intention is to relate non-equivalent contours found in a single musical piece, a performer’s improvisational vocabulary or the lexicon of a tonal language.

Instead of intervals with direction and magnitude, interlevels are comprised of direction and a specific paradigmatic inventory. LM and MH interlevels are distinct. Thus, interlevels are unlike generic step intervals (Clough 1979) or refined contours (Huron 1999) between levels, wherein LM and MH would both be movements of +1 levels. Although LM and MH are similar in their syntagm (a contour of +), they are distinct in the levels they span (Low to Mid and Mid to High). This reflects the difference between a movement from a marked level (LM) and a movement to a marked level (MH). The term *interlevel* is also motivated to distinguish these intervals from the state changes which constitute the Generalized Interval System used in this chapter. Because
state changes are the intervals of the GIS used to relate interlevels, they are intervals between intervals. Interlevels arise from an aggregate cycle formed by combining a paradigmatic cycle and syntagmatic cycle of state changes.

The problem addressed by *interlevels* was raised by the experimental findings in Chapter 5. The syntagmatic interval between Low and Mid tone levels (interlevel LM) and Mid and High tone levels (interlevel MH) is ostensibly the same, but the paradigmatic inventory in each sequence is not. The simple contour (directional comparison) produced by the tone sequences LM and MH is ascent (+). The refined contour (directional comparison and generic magnitude of change) produced by these tone sequences is “+1”. In both, there is a rise to the next higher tone level. The aspect of the signal that differentiates LM from MH is not syntagmatic but paradigmatic. Similar statements can be made about “–1” contours (HM and ML). The study findings in Chapter 5 indicate syntagmatic direction plays a major role in word identification, but we did not test perception of ML-HM or MH-LM homograph pairs because (in theory) two syllables would need to be manipulated to modulate the word. There are very few homophone pairs formed between like syntagms, a tacit acknowledgement in the lexicon of the perceptual similarity of “+1” interlevels to each other and “–1” interlevels to each other. Pairs with contrasting syntagms, like ML and MH (− v. +) and HH and HL (0 v. −) are very common, but homophone groups include LM and MH. When LM and MH do appear in a homophone group, LH is also present. In the population of all disyllables in a Yorùbá dictionary (UI 1990), I found four complete ascent groups (see Table 11.1).
<table>
<thead>
<tr>
<th>Homophone</th>
<th>MH</th>
<th>LM</th>
<th>LH</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apa</td>
<td>Arm</td>
<td>à+pa+ (compound)</td>
<td>scar</td>
<td>MM: a+pa</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LL: Prodigality</td>
</tr>
<tr>
<td>egbè</td>
<td>Club, society</td>
<td>Smoked or dried food</td>
<td>Side of body</td>
<td></td>
</tr>
<tr>
<td>ebi</td>
<td>Blood-relation</td>
<td>guilt</td>
<td>Acted as midwife</td>
<td></td>
</tr>
<tr>
<td>Ewe</td>
<td>Leaf</td>
<td>Young folk</td>
<td>v. wé</td>
<td></td>
</tr>
</tbody>
</table>

Table 11.1: All-ascent homophone groups in Yorùbá

An *interlevel* is formed by the smallest unit with both paradigmatic and syntagmatic values: two tones. Single segments with a known paradigmatic value (L, M, H) have relative height that influences the syntagmatic relationship formed with other tones. L can only be equal to or below other tones within a lexical-order window, H can only be equal to or above other segments within a lexical-order window. In themselves, single segments have no syntagm. As argued throughout this dissertation, direction of pitch change is the most reliable form of pitch perception but it does not describe all contrasts, paradigmatic cues are needed for the perceptual discretion of syntagmatically-equivalent tone sequences such as LM and MH. At the same time, because of the ambiguity of these tone sequences, their appearance in the lexicon is constrained: they rarely appear together. In this chapter, tone contrasts are generated through aggregate cycles of paradigmatic and syntagmatic state changes. Groupings and paths are defined in the context of the circle that informs tonological organization including homograph sets and tone sequences. It behooves us to arrange 2–tone sequences in a way that proximal contours are contrastive, and to separate potentially ambiguous sequences. The paradigmatic states distinguish *interlevels* that are syntagmatically equivalent, making them distant within the cyclic model.
Figure 11.1: Descent (blue) and ascent (yellow) triangles in the three-level 12-cycle

LM, MH and LH have the same syntagmatic direction (+) but different paradigmatic inventories (@LM_, @MH and @L_H). By incorporating a paradigmatic inventory, I am able to explain the difference between different “+” syntagms without using refined contour. This is important because the distance between LM and MH may not be the same. Each is unique in the position of markedness: LM is marked on the first tone, MH is marked on the second, and LH is marked on both tones. Necessitate paradigmatic features that may arise from pitch trajectory, registral differences in timbre, or level-tracking after initialization as in Chapter 6 and Chapter 10. Contour levels are a quasi-paradigmatic option that is explored, but this interpretation of the signal is sometimes misrepresentative (see Section 11.5).
11.1 Interlevel Cycles

11.1.1 Binary States

Binary states were introduced to transformational theory in a 1995 article by David Lewin, “Generalized Interval Systems for Babbitt’s Lists, and for Schoenberg’s String Trio.” In it, Lewin establishes a new way to model musical objects, binary states, and an interval specific to binary states, the state change. The binary states and state changes form a Generalized Interval System (1995:82–83), as evidenced by the Cayley table in Table 11.2.

<table>
<thead>
<tr>
<th></th>
<th>&lt;00&gt;</th>
<th>&lt;01&gt;</th>
<th>&lt;10&gt;</th>
<th>&lt;11&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;00&gt;</td>
<td>&lt;00&gt;</td>
<td>&lt;01&gt;</td>
<td>&lt;10&gt;</td>
<td>&lt;11&gt;</td>
</tr>
<tr>
<td>&lt;01&gt;</td>
<td>&lt;01&gt;</td>
<td>&lt;00&gt;</td>
<td>&lt;11&gt;</td>
<td>&lt;10&gt;</td>
</tr>
<tr>
<td>&lt;10&gt;</td>
<td>&lt;10&gt;</td>
<td>&lt;11&gt;</td>
<td>&lt;00&gt;</td>
<td>&lt;01&gt;</td>
</tr>
<tr>
<td>&lt;11&gt;</td>
<td>&lt;11&gt;</td>
<td>&lt;10&gt;</td>
<td>&lt;01&gt;</td>
<td>&lt;00&gt;</td>
</tr>
</tbody>
</table>

Table 11.2: Cayley table for a two-item binary state GIS

An important feature of binary-state GISes is Boolean addition. The state change <01> applied to the state @01 yields the state @00, not @10. Following Lewin’s notation, “@” indicates a list of states, 0 indicates an item present, and 1 indicates the item is not present. Binary state changes (indicated with angle brackets < >) are involutions. Any state change applied twice returns an object to its initial state. Binary states have been applied to contour recently by Mailman (2015). Mailman identifies a prominent arch contour in Schoenberg’s Op. 11, No. 2 and divides it into three subsegments: a front (ascent <012>), a cap (upper-neighbor figure <010>), and a back (descent <210>) (2015:12). In Mailman’s article @111 indicates the entire arch contour present, @010
indicates only the cap is present, and so on. In Mailman’s use of the states, each item in
the binary code is a chronologically-ordered CSUBSEG (contour subsegment) that is a
part of a larger arched CSEG. The application of binary states to paradigmatic and
syntagmatic components of interlevels is quite different.

11.1.2 INTERLEVEL CYCLES

For a two-bit binary code, the state changes include: <00> change nothing (the identity),
<11> change all states, <01> change the first-position state and <10> change the second-
position state. Interlevel cycles do not necessarily utilize the entire group of state
changes. They are restricted to single state changes (e.g. <01> for two-item changes and
<001> for three-item changes) with a state change sum of 1 ($\Delta 1$). Because each cycle is
formed by a sequence of single state changes, they may be called uniform.

Two cardinal cycles are used, a 4-cycle and 6-cycle. The 4-cycle is for two-item
states (with a binary place for each item) and the 6-cycle is for three-item states (with a
binary place for each item). Every item in the cycle has a maximum of two “1” values
(e.g. @011 for two-item states) because an interlevel is comprised of two tones.
Figure 11.2 shows the two-state 4-cycle on the left and the three-state 6-cycle on the right. The 4-cycle fully exploits the group with all states and state changes present. The 6-cycle omits the null-value state (@000) and the all-present state (@111).

11.1.3 Interlevel Generation

A paradigmatic inventory (e.g. @11=@LH or @111=@LMH) indicates which levels are present in a contour segment or tone sequence. In connection with Chapter 10, it also indicates which elements of a registral polychord will be utilized in a realization. A syntagmatic inventory indicates the directional relationship(s) formed by the paradigmatic items. An interlevel for a two-level system has a two-state paradigm (for L and H) and a two-state syntagm (for + and –). The same-level syntagm (0) is not included. For interlevel generation, it is instead invoked by the empty syntagm (@00 or @____) or full syntagm (@11 or @+-) in which the + and – cancel each other out. @+_ indicates no descents are present, @__– indicates no ascents are present, @+– indicates
there is at least one pivot and @__ indicates monotone. Combining a paradigmatic and syntagmatic inventories all interlevels are generated.

<table>
<thead>
<tr>
<th>@</th>
<th>@L</th>
<th>@H</th>
<th>@LH</th>
</tr>
</thead>
<tbody>
<tr>
<td>@__</td>
<td>LL</td>
<td>HH</td>
<td></td>
</tr>
<tr>
<td>@+</td>
<td></td>
<td>LH</td>
<td></td>
</tr>
<tr>
<td>@–</td>
<td></td>
<td>HL</td>
<td></td>
</tr>
<tr>
<td>@+–</td>
<td>LL</td>
<td>HH</td>
<td></td>
</tr>
</tbody>
</table>

Table 11.3: Interlevels in a two-level system

Because the empty and full syntagmatic inventory are both conceived as same level, they produce redundancies (as shown in Table 11.3).

11.1.4 SYNTAGMATIC INVENTORY

The syntagmatic inventory is used to represent a single syntagm (or contour) above, but it could also describe an entire contour adjacency series. It is similar to Friedmann’s Contour Adjacency Series Vector (CASV) in that it is not chronologically ordered. A CASV of <1,2> would mean there is a single ascent (+ is the first place) and two descents (– is the second place) in the CAS. Because the syntagmatic inventory uses binary states, it does not indicate multiplicity. Each of the two contour comparison types (+ and –) are merely present (1) or absent (0). Instead of using 0/1 for each place, I will use the item (+ or –) to indicate presence and underscore (_) to indicate absence. This will clarify the difference between the paradigmatic and syntagmatic inventories. Both will behave as binary states and <0/1> state changes may be applied. Thus, @+__ indicates that only
ascent is present and @– indicates only descent is present. The state change between all ascent (@+) and all descent (@–) is <11>.\textsuperscript{72}

\[ \text{int}( @+, @– )= <11> \]

Likewise the state change (interval) between neutral (monotone, @–) and a multidirectional contour (@+ –) is <11>. <11> changes everything. A CAS for a melodic segment or tone sequence of more than two segments is likely to have both contour directions present, so a syntagmatic inventory of @+ – is not very descriptive in this case. However, the other states of syntagmatic inventory are highly descriptive @– (a flat contour), @+ (an ascent) and @– (a descent). For a short contour of only two segments, @+ – = @–, the + and – cancel each other out. Because we are dealing with contour segments of cardinality two, only one contour comparison is possible. Hence the progression of syntagmatic directions for the 4-cycle is 0, –, 0, + in ternary contour categories.

11.1.5 PARADIGMATIC INVENTORY

The paradigmatic inventory is similar to the syntagmatic inventory in that it is not chronologically ordered. Also similarly, each item is either present or not present in the contour segment or tone sequence, it occurs at least once to be present and multiplicity is not indicated. While phonological tone levels and formalized contour levels (introduced in Chapter 8) cannot be conflated, they are often consistent with each other. Secion 11.5\textsuperscript{72} Lewin (1995) uses state change and interval interchangeably, here the term state change is preferred because of the other meanings of interval. The function \text{int}() measures the interval between states, which is the state change.
addresses distortion of tone levels in contour level reduction. The number of positions in the binary code corresponds to the number of levels in the tone system or contour reduction. Here, I will focus on two and three-level systems corresponding to Ìgbò and Yorùbá respectively. A two-level system can be represented by @LH states and three-level system by @LMH states.

Together, the paradigmatic inventory and syntagmatic inventory completely describe a tone-level or contour-level interval (an interlevel). In a syntagmatic approach, the interlevel is tonemic. In addition to the perceptual evidence for syntagmatic direction in Chapter 5, evidence is found in the lexicon: LM and MH rarely form minimal pairs because they are difficult to distinguish. Yet, both are lexical tone sequences available to a three-tone system. LM and MH commonly form minimal pairs with other interlevels. They are partially indistinct to each other through a common syntagm, but the paradigmatic inventories are distinct: @LM_ and @_MH.

11.2 Interlevel cycle for a Two-level system: LH

A two-level system of Low (L) and High (H) has four distinct interlevels ($2^2=4$): HH, HL, LL, and LH. Each of these intervals has a paradigmatic inventory and a syntagm. Unlike a three-level system there are not many interlevels that share a syntagm, only two: LL and HH. LL has the paradigmatic inventory @L_ and the syntagmatic inventory @_ _ (or 0 for flat). HH has the paradigmatic inventory @_H and the same syntagmatic inventory. Of the four interlevels (HH, HL, LL, LH), there are two with the same syntagm (HH and LL) and two with the same paradigm (HL and LH = @LH). If we
consider the syntagmatic inventories @+ – and @ _ _ as equivalent (both 0 or same level), a uniform cycle that alternates <01> and <10> state changes is formed with two static syntagms, one ascent and one descent [0 + 0 –].

<table>
<thead>
<tr>
<th>State</th>
<th>@+ –</th>
<th>@ – _</th>
<th>@ _ _</th>
<th>@ _ +</th>
<th>@+ _</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change</td>
<td>&lt;10&gt;</td>
<td>&lt;01&gt;</td>
<td>&lt;10&gt;</td>
<td>&lt;01&gt;</td>
<td>&lt;10&gt;</td>
</tr>
</tbody>
</table>

Table 11.4: Syntagmatic states and state changes for 4-cycle (bold indicates completion)

A uniform cycle can be formed for the paradigmatic inventories using the same state changes, however the null state @ _ _ (no tones present) does not form an interlevel and is therefore excluded from the cycle. By modifying the state-change cycle as below (swapping the first and last changes), the null is replaced with another “all present” (@LH).

<table>
<thead>
<tr>
<th>State</th>
<th>@LH</th>
<th>@ H</th>
<th>@LH</th>
<th>@L</th>
<th>@LH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change</td>
<td>&lt;10&gt;</td>
<td>&lt;10&gt;</td>
<td>&lt;01&gt;</td>
<td>&lt;01&gt;</td>
<td>&lt;01&gt;</td>
</tr>
</tbody>
</table>

Table 11.5: Paradigmatic states and state changes for 4-cycle (bold indicates completion)
Figure 11.4: Paradigmatic 4-cycle

Not all objects in the space of 2–bit binary codes are encountered in the modified cycle. Instead, there is a redundancy of @LH (@11) along the east-west axis.

By combining these cycles into an aggregate, four unique interlevels are produced, the syntagmatic cycle determines direction and the paradigmatic cycle determines range.

Figure 11.5: Aggregate cycle for two-level system
This arrangement in Figure 11.5 reflects features of each interlevel including height, direction and markedness. Height is reflected through the placement of all high (HH) and all low (LL) along the north-south axis, and mixed (HL and LH) along the east-west axis. Syntagmatic direction follows a clockwise path along the cycle, with HL at 3 o’clock and LH at 9 o’clock. If we apply the model to the Ìgbò language, we see a progression from least marked at the top to most marked at the bottom. The HH toneme is the least marked (high is the neutral tone). The east-west tonemes (LH and HL) both have an unmarked (H) and marked (L) tone and the southmost toneme is marked in both elements.

A homograph group like /a.kʷa/ may is generated by the complete cycle (see Table 11.6). The consecutive intervals in the cycle each form a minimal pair that is differentiated by a single (minimal) state change of paradigm and of syntagm. These minimal pairs are strongly contrastive. The east-west poles and north-south poles do not form minimal pairs because each requires a double-state change, east-west is a double-state change of syntagm (int(@+_ ,@_–)=<11>) and north-south is a double state change of paradigm, int(@_H, @L_) = <11>.

<table>
<thead>
<tr>
<th>Word</th>
<th>ákwáá</th>
<th>ákwáá</th>
<th>ákwáá</th>
<th>ákwáá</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gloss</td>
<td>cry</td>
<td>cloth</td>
<td>bed</td>
<td>egg</td>
</tr>
<tr>
<td>Tone</td>
<td>HH</td>
<td>HL</td>
<td>LL</td>
<td>LH</td>
</tr>
</tbody>
</table>

Table 11.6: /a.kʷa/ homophone group

Another way to visualize the aggregate 4-cycle is with a sine wave. This is more reflective of the syntagmatic cycle with each zero-crossing corresponding to a syntagm of “0”, the peak to + and the trough to –.
11.2.1 PAIRS AND PATHS

The relationships between intervals in the cycle are two-fold: contrastive and sequential. The contrastive relationships (pairs) are indicated by the solid line of the circle. Each consecutive pair of interlevels forms minimal pairs differentiated by a single state change in both the paradigmatic inventory and syntagmatic inventory. Additionally, a pathway is formed by progressing clockwise through the circle (indicated by green arrows in Figure 11.7). This pathway indicates the order in which interlevels may appear in a tone-sequence signal. Each interval has an antecedent tone in the first position and a consequent tone in the second position. In a contour, the consequent tone of one interlevel becomes the antecedent tone of the next interlevel. The pathway in Figure 11.7 is quite apparent: HH may be followed by itself or HL but not LH; HL may be followed by LL or LH but not itself.

Figure 11.6: Single sine-wave cycle representation of aggregate cycle
HL and LH do not form a minimal pair but they share the property that they may precede or follow each other. They are freely sequential with each other and can even create a loop. This property is not shared by HH and LL, the north-south poles. HH and LL may loop themselves.

11.2.2 DOWNSTEP AND HIGH-RAISING

A problem in applying the aggregate 4-cycle to Ìgbò tonology is accounting for the step-tone. This is a tone that only appears after the high tone but is not at the same level as a previous high tone. Instead, it initiates a new pitch for the high level. In the orthography used by Williamson (1972), step is treated much like a middle tone in Yorùbá. It is given its own diacritic (e.g. álú v. álū). However, in Clark (1990) step tones are not on a separate level but are down-stepped high tones (e.g. álú v. ál'ú) where “!” indicates when downstep occurs between syllable. There are homophone pairs of HH-HS and HS-HL. Thus, the downstepped tone behaves like a distinct level in some cases. In other
homophones, HH and HS (or H!H) yield equivalent meaning, which supports Clark’s orthography. Clark’s downstepped high tone works better for this formal model. Though somewhat counterintuitive, a combination of a paradigmatic state of @_H and a syntagmatic state of @–_ produces the H!H (HS) interlevel. At first, I feared the step tone challenged the interlevel formalization, but in fact it is a great application.

If @_H is north and @–_ is east, this places H!H north east. In Figure 11.8, the northeast interlevel is placed outside of the cycle as an alternative path from HH to HL to avoid disruption of the normative path of the two-level 4-cycle. In the alternative path, the syntagm is changed first, int(@+–, @–_)=<10>, with no change to the paradigm, yielding a new interlevel, H!H. Then, the paradigmatic state is changed, int(@_H, @LH) =<10>, completing the transformation from HH to HL.

A similar phenomenon to downstep is high-raising documented for Yorùbá by Laniran and Clements (2003). It is also possible in a two-level system like Ìgbò, but does not form a distinct toneme. In Figure 11.9, a “high-raising” interlevel is placed to the
northwest of the circle because it shares a paradigmatic state with the interlevel HH (to the north) and a syntagmatic state with the interval LH (to the west). An interval with a raised high as the consequent level does not form a homophonic minimal pair with either HH or LH so the “links” between are weaker than the others (indicated by dotted black lines in Figure 11.9). The presence of raising and lowering makes arrows a better symbol than “!””. H↑H is not a distinct toneme from HH, it does not produce unique words, but it can substitute for HH or follow or precede HH in a contour. As in Figure 11.7, the green arrows in Figure 11.9 indicate the pathways between intervals.

![Figure 11.9: Two-level 4-cycle with satellites for raising and lowering High level](image)

In Figure 11.7, there were 6 distinct paths from one interlevel to the next (counting bidirectional arrows as two paths). With the addition of H↑H and H↓H in Figure 11.9, the pathways proliferate to 15.
11.3 Interlevel cycle for three-level systems: LMH

In a three-level system there are nine interlevels \(3^2=9\). Three ascend (LM, MH, LH), three are static (LL, MM, HH) and three descend (HM, ML, HL). To model the tonology for a three-level system like Yorùbá. I will not use a 9-cycle, which is an odd number. Instead, I will divide interlevels into two categories: dynamic (ascents and descents) and static. By alternating dynamic and static syntagms in our cycle we are able to use the same syntagmatic 4-cycle: 0, –, 0, +. This requires redundancy in the static interlevels because there are only three compared to six dynamic.

Instead of a paradigmatic 4-cycle, the 6-cycle is necessary for three levels. There are eight possible paradigmatic inventories for a contour segment using three levels \(2^3\). Because the aggregate cycle is representative of single interlevels and not longer contours, the “all present” state is not used (@111, in this case LMH), nor is the null state (@000). Six paradigmatic states are left: @001, @011, @010, @110, @100, and @101, which are @____H, @_MH, @_M, @L_, @L_, and @L_H respectively. Like the alternation of static and dynamic syntagms, single level and two-level paradigms alternate. The aggregate cycle for interlevels in a three-level system has length, twice through the paradigmatic 6-cycle, and thrice through the syntagmatic 4-cycle (see Figure 11.1). Because there are only 9 unique intervals, in the three-level 12-cycle there are redundancies, in the contributing syntagmatic and paradigmatic cycles and in the surface aggregate. Each two-tone paradigm produces two distinct interlevels depending on the syntagm it is combined with (@+_ or @_–). However, single-tone paradigms only produce the same interval whether they are aggregated with the null (@_ _) or neutralized (@ +–) syntagm. The single tone paradigms start and end on the same tone,
and thus are static. Each single-tone paradigm is combined with the null (@_ _) and the neutralized (@+–) syntagms, but the surface interlevel is the same.

Figure 11.10 shows the aggregate 12-cycle for the three-level system. MM starts the cycle because it is the unmarked interlevel, just as HH is the unmarked interlevel in the two-level system of Ìgbò. MM forms a north-south pole because of the redundancies, in the two revolutions of the paradigmatic 6-cycle, each of the single-tone paradigms is repeated. LH and HL occupy the same east-west poles in the three-level cycle. The three-level 12-cycle has a number of symmetries reflecting features of the interlevels. The first of these is pitch height. The height of all four directional poles sums to zero if [L,M,H]=[–1, 0, +1]. Halves of the cycle divided at the north-south pole or east-west pole also sum to zero. The quadrants northeast and southwest sum to +3 and the quadrants northwest and southeast sum to –3, similar to the positive-negative valence of the quadrants of the Cartesian plane.
In addition to the zero sums at the four poles in Figure 11.10, other geometric shapes are formed by this interlevel cycle. The ascending intervals form a triangle, as do the descending intervals. Also, there are two complete sets of static intervals, each of which forms a triangle.

Like consecutive interlevels in the two-level cycle, consecutive interlevels in the three-level cycle form strongly contrastive pairs. Each interlevel aggregated from a single-tone paradigm and static syntagm (LL, MM, HH) has four interlevels with which it can form strongly contrastive pairs. This is not true of the two-tone paradigms.
The colored chords in Figure 11.12 indicate pairs with a state change sum of greater than 1 (e.g. \( \Delta(<101>) = 2 \)) between respective paradigms. The green chords between ML-MH and HM-LM indicate that these pairs are strongly contrastive because their syntagms are inversions of each other. They differ from the consecutive intervals in the cycle because they require \( \Delta 2 \) state change of paradigm and syntagm (e.g. \( \text{int}(\text{LM}_-,\text{MH})=<101>, \Delta(<101>)=2 \)). Therefore, they cannot form interlevel sequences, only minimal pairs. The yellow chords connect interlevels that share the same syntagm but have different paradigms (with a state change sum of 2). These minimal pairs are weakly contrastive. They exist in the Yorübá lexicon (and other three-tone languages), but as evidenced by the study in Chapter 5, the judgment of magnitude is not so reliable as the judgment of ternary direction. Cues that may or may not be embedded in the pitch dimension are needed to strengthen the contrast between such homophones.
The Yorùbá lexicon has no disyllable homophones with a unique word for all nine interlevels. On the upper end is the homophone /i.ɡba/ which produces five distinct words (see Table 11.7).

<table>
<thead>
<tr>
<th>Word</th>
<th>igbá</th>
<th>igba</th>
<th>igbà</th>
<th>igbá</th>
<th>igbá</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gloss</td>
<td>calabash</td>
<td>two-hundred</td>
<td>climbing rope</td>
<td>time</td>
<td>garden egg</td>
</tr>
<tr>
<td>Tone</td>
<td>MH</td>
<td>MM</td>
<td>ML</td>
<td>LL</td>
<td>LH</td>
</tr>
</tbody>
</table>

Table 11.7: /i.ɡba/ homophone group

If /i.ɡba/ homophones are placed on the three-level 12-cycle they occupy consecutive positions from LH to MH. This incorporates two of the three ascent interlevels (LH and MH, but not LM). LH spans the full range of levels, but MH and LM do not. As discussed in the introduction to this chapter, and indicated in Figure 11.12 (above), LH-MH and LM-LH form weakly contrastive minimal pairs, but MH-LM does not usually form minimal pairs. I only found four all-ascent homophone groups in a population of over 2000 disyllables. If a homophone includes LH-MH it is unlikely to include LM-LH because that would entail a third pair MH-LM.
Most minimal pair types and many larger homophone collections occupy consecutive positions in the cycle. Cycles of single-state changes create a meaningful geometry of tone sequencing and lexical contrast.

11.3.1 INTERLEVEL PATHS

As with the two-level 4-cycle, the three-level 12-cycle. It also indicates the order in which interlevels may appear in a tone sequence. Unlike the two-level 4-cycle, it is not a single clockwise pathway around the circle but two paths: one clockwise down the east side and one counter-clockwise down the west side. Any of the static interlevels (LL, MM, HH) may be skipped or be a source of indefinite prolongation (e.g. MMMMMMMMM...). I static interlevels are parenthetical in the pathways. Following the path along the westside, skipping the static interlevels, produces the sequence MLHM, composed of interlevels ML, LH and HM.
Following the eastside, the sequence is MHLM, an inversion of the westside along the MM-axis (north-south pole). H is sent to L and L is sent to H, M is unaffected. The westside and eastside contours are also retrogrades of each other. If the paths are combined, westside then eastside (MLHMHLM) or eastside then westside (MHLMLHM), two palindromes are produced which are inversions of each other.

Not all contours follow the path along the rim of the cycle. Paths are likely to include diametric chords from one interlevel, across the geometric center of the circle, to an interlevel of the same paradigm. The diametric chords in Figure 11.15 can form loops (e.g. MLMLMLML..., MHMHMHMH... and HLHLHLHL...).
Later in this chapter, a path that incorporates consecutive crossings along the same diametric chord (back and forth) is shown to reduce the robustness of the signal underspecifying level identity.

### 11.4 Interlevel cycles for larger level systems

So far, my attempts to form interlevel cycles for systems with more than three levels have failed, and the utility is uncertain. I can however speculate on the size of these cycles. A two-level system (e.g. @LH) has four interlevels: two static and two dynamic. The number of static and dynamic interlevels is the same, so redundancies of the static intervals are not needed to articulate all intervals in the cycle. A three-level system (e.g. @LMH) has three static and six dynamic interlevels, thus the static interlevels occur twice in the cycle. The number of interlevels in larger systems can easily be determined:
the number of levels times itself. The number of static intervals always equals the number of levels. The number of dynamic intervals is found by subtracting the number of static intervals (column 3 in Table 11.8) from the total number of intervals (column 2).

<table>
<thead>
<tr>
<th>Levels</th>
<th>Intervals</th>
<th>Static</th>
<th>Dynamic</th>
<th>Sets of Singles</th>
<th>Cycle Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>3</td>
<td>6</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
<td>4</td>
<td>12</td>
<td>3</td>
<td>24</td>
</tr>
<tr>
<td>5</td>
<td>25</td>
<td>5</td>
<td>20</td>
<td>4</td>
<td>40</td>
</tr>
<tr>
<td>6</td>
<td>36</td>
<td>6</td>
<td>30</td>
<td>5</td>
<td>60</td>
</tr>
<tr>
<td>7</td>
<td>49</td>
<td>7</td>
<td>42</td>
<td>6</td>
<td>84</td>
</tr>
</tbody>
</table>

Table 11.8: Inventory of n-level systems

A 4-level system has 16 interlevels: 4 static and 12 dynamic. If a cycle could be created with single state changes of paradigm, as in the two-level 4-cycle or three-level 12-cycle, then static and dynamic would need to be alternated requiring each of the single-tone paradigms to occur three times in the cycle. This produces 12 static interlevels to complement the 12 dynamic interlevels. Unlike the three-level cycle, a four-level cycle cannot simply loop the same paradigmatic cycle, each paradigmatic cycle would need to be unique. If we number the four levels \([0,1,2,3]\) instead of using letters, a four-level cycle could follow this order of single-tone paradigms, abbreviated \(@2\) for \(@_ _ 2 _\).

\(@0 @1 @2 @3 @0 @2 @3 @1 @0 @3 @1 @2\) (loop back to @0)

This 12-cycle of double state changes is turned into a cycle of single-state changes by inserting two-tone paradigms between each one-tone paradigm (i.e. @01 between @0 and @1). Each paradigm appears twice to capture the different syntagms that may combine with the two-tone paradigms. The one-tone paradigms are redundant because

\(^{73}\)Alternatively, T for top or B for bottom could be added to L, M and H.
their syntagm can only be static. The two-tone paradigms are dynamic, either ascending (+) or descending (–). If the paradigmatic cycle is aggregated with a syntagmatic cycle, it cannot be a loop of the 4-cycle. Each two-tone paradigm should be aggregated once with (+) and once with (–), but this does not happen if the 4-cycle is looped over the paradigmatic cycle I have described.

An explanation for the failure to devise a single interlevel cycle for four levels in the mold of the three-level 12-cycle is in the number of unique pairs of levels compared to the number of levels. In a two-level system there is one two-tone paradigm, in a three-level system there are three two-tone paradigms, but in a four-level system there are six two-tone paradigms (as shown with the tetrahedron above). With four or more levels there are more pairings than levels. If a shape is formed with each level as a point and lines connecting each level to every other level, there are more than two line segments that converge in each point representing a level. Any order of two or three levels can be placed end-to-end in a circle and each level will be next to every other level, but this is
impossible for four levels. The only possibility is to enter three-dimensional space with the aggregate of paradigms and syntagms as in Figure 11.17.

![Figure 11.17: 3D interlevel network, arrows indicate equivalent points in the wrap-around](image)

### 11.5 Robust Signals

Contour-level reduction is not a lossless process. Portions of a melodic segment may be distorted in the mapping to contour levels. Contour level reduction detects local maxima, local minima and medial pitch heights within a window of comparison. Under the right circumstances, a 2–degree window will return contour levels that correspond to three tone levels as in Yorùbá or three scale degrees of a melody (if that is the extent of the input). Sometimes contour level reduction is faithful, other times it is not. As of this date, I know of no more reliable process to glean tone levels from a signal than contour levels. Contour levels are formalized where tone levels evade formalization. Aberrant realizations of underlying tone are a problem for contour levels, much as they are for
frequency banding. For instance, LHHH realized as a chain of ascents (observed by Laniran and Clements 2003) would be interpreted as a <0112> or <0122> by a CL2SEG, depending on the window orientation, not <0222> as the underlying tone suggests. For smaller window sizes (such as two degrees), repeated or “trilled” tones may be distorted, regardless of the clarity of the realization. As demonstrated in Chapter 9, the level of repeated notes may change with each successive repetition (as in Figure 11.18).

Figure 11.18: Anomalies in contour levels (sums of contour slices): (left) consecutive repeated pitches, and (right) insufficient context for time-extrema.

This issue is remedied by collapsing all repetition to a single segment. However, if tones are the same level but have different pitch, how can that be done? This is the disconnect between underlying and realized tones that is only partially emended by using contour levels, instead of frequency bands, for speech analysis.

Because contour-level analysis is designed to reflect perception, tone sequences unclear in reduction may also be unclear perceptually. The sender is aware of the paradigm and the syntagm but the receiver may only be aware of adjacent and non-adjacent pairwise comparisons up to some limit (see Chapter 7). A basic example of what might be called paradigm failure or loss (depending on whether the levels were initialized) is sequences that toggle (or “trill”) between two tone-levels out of three: HMMHMHMH (2121212), MLMLMLM (1010101) and HLHLHLH (2020202). After
contour level reduction and consecutization all of these reduce to 1010101. In fact, consecutization could be applied to the input integer encodings (e.g. HLHLHL = 2020202 = 1010101), so the contour-level reduction is extraneous. If such sequences are the extent of the signal, then the three levels are never initialized. If these sequences follow initialization of all three levels, then the paradigm may be lost. When analyzing tone sequences that toggle between two levels, there is nothing that non-adjacent comparisons tell us that simple-adjacent (+1/–1) comparisons do not. For a signal to be robust for perception and contour level reduction (to the extent these overlap) then interlevels of the same two-tone paradigm should not be in clusters greater than two. This implies multiple diametrical crossings along the same axis should be avoided.

![Figure 11.19: Preferred (green) pathways for a robust signal](image)

Figure 11.19: Preferred (green) pathways for a robust signal
Contours that follow the green pathways shown in Figure 11.19 are more robust for encodings with register transformation and decoding realized tone through contour level reduction.

Even with a moderately robust contour like HLMLHLM (that follows suggested pathways), some distortion may occur with specific window sizes. The larger the window, the safer the signal. However, too large a window creates artificial detail in a reduction and may read similar contours as incongruent, like the fugue subject and tonal answer example in Chapters 1 and 8. If trying to decode a three-level tone sequence, one should not use a four-level reduction because it does not actually reduce anything. Orientation of the window around the focused segment makes a considerable difference. If we are modeling listening to speech in a tone language or listening to novel music for the first time, it makes sense to make backward, and not forward, comparisons. Below the same tone sequence HLMLHLM encoded into contour pitches as 2010201 is decoded three different ways, all with the same window size (depth of three, two comparisons) but with different orientations around the event in focus. Both the “–2” (backward comparison) window and the “+2” (forward comparison) window maintain the integrity of the signal, but the window with an even radius around the focus introduces a false maximum.
If the CAS (<–+–+–>) is maintained but the H and M tones are swapped around, the *false maximum* problem may persist. The pallendromic tone sequence HLMLMLH is also obfuscated when a uniform radius (“+1/–1”) window is applied in contour level reduction. The introduction of two *false maxima* degrades the three-level contour to a two-level contour.

For the “+1/–1” window, the consecutize function may be applied to the result:

\[ \text{cons}(2020202) = 1010101. \]
Another palindromic tone sequence with the same CAS (<–+–+–+>) but a different assortment of H and M tones is MLHLHLM. It suffers from a different problem: an H tone that becomes a *false medial*.

In this case, a “+1/–1” window does not corrupt the tone levels. Instead, it is the +2 and –2 windows that fail. Depending on the orientation of the window around the focus, one of the H tones in the middle of the sequence is read as medial (1) instead of maximal (2).

Yet another signal loss is possible when an M tone is placed between two high tones and the “+1/–1” window is used: a *false minimum* (middle reduction below).

360
MLHMHLM reveals a fourth type of distortion in which two different levels have the same relative height (left and right reductions). Among other problems, this obfuscates pairwise adjacent relationships, with M and H returning the same sum: 1. This problem is also found in LMHLHML when a “+2” or “−2” window is used. This signal is preserved with a “+1/−1” window.

![Diagram showing window reductions for MLHMHLM](image)

Following a path along a single side will always produce a robust signal in itself. However, following the westside then eastside will produce local maxima at LML when using a −1/+1 window.

![Diagram showing window reductions for MHLMLHM](image)
An eastside to westside path is more robust than westside to eastside, unless starting midway through a side. Sequences of three tones, one at each level, are always robust. Nonadjacent recurrence of tone levels within a local window may result in ambiguity.

A reliable way to produce a signal that will not be distorted by contour level reduction is to simply avoid repetition of H within the window.

![Figure 11.26: “–2” (left), “–1/+1” (middle) and “+2” (right) window reductions for HMLHLMH](image)

When using C+ ascent to model contour, repetition of M and L tones within the window is unproblematic, but redundant H tones are distorted by the reduction. If using C–descent (1 for lower, 0 for higher or same pitch), low tones will distort. If I knew more precisely how tone is perceived in a three-level system like Yorùbá, it would be easier to define a robust signal. Is binary C+ ascent really plausible? Exactly how many degrees of adjacency are pertinent? Robustness here only refers to the fidelity of these signals within the C+ ascent comparison and 2–degree window model. Although this model has been arrived at through empirical evidence and reason, it still may be a ways off from reality.
11.6 Contour Transformation

Interlevel cycles relied on binary states as inventories. A state of @LMH (@111) would mean all three tone levels are present, and @_ _ _, nothing present. Another use of binary states and state changes is a contour adjacency series (CAS). A CAS is a time-ordered series, unlike inventories, which ignore chronology. Conventionally, a CAS includes pairwise adjacent comparisons using ternary contour categories. If all repetitions (0) are removed, there are only two directions left (–, +). A CAS can then be converted into a binary state (indicated by @) and state changes (indicated by <> ) may be applied: <+--+-+> = @100101, such that <100010> applied to <+--+-+> is <---+++> (@000111)

CAS notation and state change notation both use <> as brackets, but the presence of +/- as opposed to 0/1 may be used to differentiate them.

State changes may also be applied to contour slices in a binary state form: @1|0. The vertical bar indicates the position of the focused pitch segment. Places to the left are pre-adjacent comparisons and right are post-adjacent comparisons. @11| indicates the focus is above the two previous segments, @|11 means the focus is above the two segments after it. The state change <01> applied to the slice @1|0 produces the slice @1|1 (a local maximum).

Binary states for series and slices are not isolated entities. A change in one contour slice will have ripple effects in the rest of the contour. Changing an ascent to a descent in a CAS at one point will require reconciliation with other states, at least locally. This is particularly true when a contour is constrained by a number of levels. If the descent from a high tone to a mid tone is changed to an ascent, what happens to their identity? There is no better testing ground for a theory of contour realization and
transformation than Yorùbá oral poetry. Yorùbá oral poetry will be explored in more
detail in Chapters 12–14, including ethnographic information. For now, I consider a
poetic device that can be modeled with state changes.

11.6.1 TONAL COUNTERPOINT IN YORÙBÁ ORAL POETRY

*Tonal counterpoint* is a common device in Yorùbá oral poetry documented by
Ọlátunji (1984) among others. Like many other terms in this dissertation, there may be
some semblance of meaning for musicians, but the nuance may not be immediately
apparent. Both “tonal” and “counterpoint” are terms relevant to music, but the meaning
here is the linguistic tonal (as in lexical tone) and a rhetorical counterpoint, not a
polyphonic one.

Ọlátunji describes couplets in which each phrase is parallel if not identical in
terms of phonic content and the first sets up a tonal expectancy for the second (1984:35).
The contrast might also be between words within a single phrase. The tonal expectancy
for a final low is set up by first ending a similar word or phrase on a non-low tone, H or
M. H is countered with L in the next phrase. M is also countered with L. L has finality.

After I first read about this device several years ago, I started to notice it. I have
been astonished by its prevalence in a variety of Yorùbá vocal styles, including neo-
traditional popular music. Examples from recent field recordings will be examined in
Chapters 13 and 14. Ọlátunji offers examples of his own and from out-of-print texts. For
parallelism between two phrases in a couplet he offers the text in Table 11.9.
Many H and M tones are changed to L tones in this couplet. This is done by replacing words in the first phrase with syntactically-congruent and semantically-analogous words in the second phrase (Olatunji 1984:35). Butterflies become big insects, ear becomes head. Tonal counterpoint enlivens the pitch dimension in poetry.

The richest use of tonal counterpoint may be the use of tonally-contrastive homophones.

In Table 11.10, fuss (or matter, kétékété) is changed to donkey (kétékété) by inverting the tone. In this case, the counterpoint is interior to a phrase and L tones precede the H-tone version.

Tonal counterpoint may also create a meaningless contrast. Meaning is not changed by the transformation if the word is not part of a homophone group.
The meaning of láéláé is not changed in Table 11.11 because “there is no word like láèlè” (Ọlátunji 1984:35). Contrastive tone that does not alter meaning is simply a form of wordplay. Because the number of syllables is the same in both phrases of the couplet in Table 11.11, it is easy to compare the CAS of each. In Table 11.2, the tone sequence for the first phrase is in the top row and the sequence for the second phrase is in the bottom row. A CAS for each is in-between.

<table>
<thead>
<tr>
<th>Tones 1</th>
<th>M</th>
<th>M</th>
<th>L</th>
<th>H</th>
<th>L</th>
<th>M</th>
<th>M</th>
<th>H</th>
<th>H</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAS 1</td>
<td>0</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CAS 2</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td>-</td>
<td>+</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Tones 2</td>
<td>M</td>
<td>M</td>
<td>H</td>
<td>H</td>
<td>L</td>
<td>M</td>
<td>M</td>
<td>L</td>
<td>L</td>
<td>L</td>
</tr>
</tbody>
</table>

Table 11.12: Contour Adjacency Series for couplet in Table 11.11

The change of the third tone from L to H changes two contours around it. However, the change of the last four tones from H to L only changes one contour. It is most efficient to view the latter as a single change of a syntagm instead of a change in four paradigms.

The transformation in the last example from Ọlátunji is very similar to an application of the device by Mayowa Adeyemo, a female poet and musician who was studying at Lagos State University when I met her in 2013. In a 2013 performance at the University of Lagos, she rendered Oríkì Ejiré (the praise of twins) among other praise poems. The tonal counterpoint is quite striking when it is heard (as opposed to read). Even though I was just beginning to study Yorùbá in depth at the time, the tonal counterpoint was quite apparent to me.
This couplet contrasts tone between a near homophone: “jó” and “yò.” This is unlike Ọlátunji’s examples which are either homophones, the same word with distorted tone or simply words of similar length and analogous meaning and type. The tonal expectancy for L tone is built up until the very end. The incomplete paradigmatic inventory of the first phrase (@MH) is completed by the second (@LMH), a state change of <100>. The reduced CAS (with same level syntagms removed) morphs from <+–+> to <+–→>, a state change of <001>.

The first phrase is underspecified in terms of tone levels. Three levels are never initialized so the paradigm cannot be derived from the syntagm. The state change from the first makes the second a complete paradigm and also makes the second phrase robust.
Full initialization does not occur until the very end of the couplet, quite an effect! Despite all the contrast, each component phrase is balanced in terms of markedness, with two marked tones in each.

I have taken a liberal stance on contour equivalence, allowing for contours of different cardinality and pitch uniqueness to be formally the same. However, I am more skeptical of the perceived similarity of transformed contours, such as those found in these tonal counterpoint couplets. Beyond that, I also doubt transformations such as inversion and retrograde are perceptually similar for most listeners. Contour transformation through state changes is the last bit of formal machinery I will introduce in this dissertation. They are fairly modest in themselves. More change than flipping a single syntagm here or there and the relationship between two contours is lost. Proximity, homophony and semantic equivalence relate couplets in tonal counterpoint. It is contour transformations that distinguish them and make them contrastive. Although changing the final Hs to Ls in Figure 11.28 can be represented with a modest state change, these are different contours. After all the shades of similarity and equivalence explored, we have arrived at dissimilarity and contrast.
III. Tone and Tune
12. VOCAL ARTS IN SUB-SAHARAN AFRICA

12.1 Vocal Arts and African Cultures

In Sub-Saharan Africa, singing has been a casual pastime, formal practice, and specialized profession for centuries. This is embodied in the call-and-response choruses found throughout the continent, involving a fluid exchange between leader (often a professional praise singer) and an inclusive group of all persons present, building community in perhaps the best way: through singing together. Unlike North America and Europe, in Africa one rarely hears the phrase “I cannot sing.” The average singing diet is complex, including Arabic chant or English hymnody (depending on whether one is Muslim or Christian), popular music in English or French (depending on the colonial past), and choruses in one or more of Africa’s 2000 languages.

Throughout Sub-Saharan Africa there are historic practices of praise poetry, chanted and sung, that thrive to this day, including azmari bet in Ethiopia, oriki in Nigeria, izibongi in South Africa and the griot of the Sahel region (just south of the Sahara). There is also great diversity. Ethiopia has a rich inventory of wind and string instruments and unique scales that have coalesced with African-American music in Ethio-jazz. Singing in Nigeria is often accompanied by talking drums that mimic the pitch contours of the voice. The complex vocal polyphony of South Africa, notably Zulu mbube and isicathamiya, is the sound often identified with Africa in the West through Paul Simon's album Graceland and Disney's movie musical The Lion King.
12.1.1 ORALITY AND LITERACY

The legendary Saint Yared of Ethiopia is reported to have devised a musical notation in the sixth century. Evidence of notated chant, or melekket, exists in the form of manuscripts dating from the sixteenth century, but this notation is not used outside of the Ethiopian Orthodox Church (Shelemay and Kimberlin, 2013). The vocal arts in Sub-Saharan Africa remained a predominantly oral tradition into the twentieth century. Missionaries first introduced Western notation into Sub-Saharan Africa in the nineteenth century, and Tonic Sol-Fa notation is now widely used, as is shown in Figure 12.1.

![Music notation](image)

*Let us now rejoice and hail this great morn.*

Figure 12.1

Today, vocal forms that were transmitted orally, like praise poetry and choruses, are now a mix of orality and literacy (Kaschula 1999:56). Digital recording and the Internet have reinvigorated and transformed oral traditions, and facilitated music and film industries independent from the West that are distinctly African.

12.1.2 LINGUISTIC DIVERSITY

Several language families are unique to Africa: Niger-Congo A, Niger-Congo B (Bantu), Nilo-Saharan, and Khoe-Saan (Khoisan). There are also large numbers of native
and non-native speakers of Afro-Asiatic languages (notably Amharic, Arabic, and Berber) and Indo-European languages (notably Afrikaans, English, French and Portuguese). All in all, there are over 2,000 languages spoken in Africa, over a quarter of the 7,000 languages estimated to be spoken in the world today. In Nigeria alone, there are approximately 500 languages spoken, while Ethiopia has a formidable 87 languages. South Africa’s language count is lower, but its linguistic climate remains complex with a number of native languages, including Sotho, Xhosa and Zulu, that are more robust than the colonial offshoot Afrikaans (Lewis et al. 2013).

12.1.3 VOCAL TECHNIQUE

A wide spectrum of sounds are heard in African vocal arts, from speech to chanting to singing to ululation (a uvular trill) to imitations of percussion and animal sounds. All of these sounds might be heard in a single performance, spread amongst the members of an ensemble, or coming from a single voice, like an azmari (Amharic praise poet) in an Ethiopian tej bēt (honey-wine tavern) (Kebede 1975:52). However, many forms and the artists who specialize in them are defined by specific vocal techniques.

In describing African vocal arts, it is important to identify the language, the movements involved, accompanying instruments, the content or purpose, and vocal style. The language of a vocal performance suggests further details about the phonetic inventory used and the ethnic and geographic origin. Xhosa izibongi might include clicks while poetry from West Africa will not. Description of vocal style may include the type of phonation, registration, and the presence of timbral effects like nasalization. Is it
speech-like or sung? Is it full voice or falsetto? Is it tense or relaxed? Whether it is Zulu choral music or Yorùbá poetry, vocal quality is often more important to classification than subject matter (Babalola 1966; Erlmann 1999; Vidal 2012). The same words repeated with a different vocal style become a different type of poetry. For example, while essentially all Yorùbá oral poetry involves oríkì (praise), it is varied by the vocal style (Okpewho 1992:129).

12.1.4 SOLO PERFORMANCE

Vocal performance has been a profession in many African cultures for centuries, including the iconic griots of the West African Sahel (Hale 1998), the azmari of Ethiopia (Kebede 1975) and the imbongi of South Africa (Kaschula 1999). These vocalists use a wide variety of voice production and are variously referred to as oral poets, praise singers and verbal artists in academic literature depending on the discipline of the researcher.

Okpewho explains that pre-colonial communities were governed by a king or chief that had “absolute authority over all matters” and employed court poets whose “major role was to swell the image of the ruler.” (1992:25) Oral poets also carried news from one place to another, recanted folktales, and praised gods. But fundamental changes to African life have drastically changed the role of the poet, who is now largely relegated to “cultural performances.” (Okpewho 1992:25) Instead of praising kings, the poet now praises the rich; instead of indigenous gods, he praises Allah or Jehovah. One constant is social critique, which remains an important aspect of African oral poetry. Poetry of dissent can incite listeners to protest, as it did during the apartheid era in South Africa,
leading to the imprisonment of poets (Kaschula 1999:57–60). During the occupation of Ethiopia (1936–1941), the Italians gathered *azmari* and executed them for fear of their sway over public opinion (Shelemay and Kimberlin 2013). One hears echoes of the alternation of praise and social critique in African-American hip-hop (Smitherman 1997:4), and now African hip-hop, which ranges from self-praise and boasting to mocking of competitors and social and political indictments.

Vocal style often distinguishes one type of oral poetry from another. Across the continent, Africans are acutely aware of the many sounds the voice can produce. Yorùbá poetry is “classified according to the manner of voice production employed...” (Babalola, 1966:vi). Vidal aptly describes these:

*Ìjálá* is characterized by nasal colouring and acoustically intense open voice tone quality... *Ewi* is characterized by a high falsetto and wailing voice quality... Tense vocal quality with slight nasal colouring can be observed in *rara*... *Iyere Ifa* has the purest tone quality of all the four chants (2012:71–72).

To help demonstrate some of the differences, four video links of examples of Yorùbá praise poetry are provided below:

- Muri's *Ìjálá* Example 2:  <http://youtu.be/0Ry5CCmKQNE>
- Máyòwá Adéyẹmọ praises Ogun:  <http://youtu.be/2nEa9v5Np6k>
- Máyòwá Adéyẹmọ praises Eledumare:  <http://youtu.be/rpnPkZDA0I8>

The recent emergence of women in the largely male field of oral poetry (Kaschula 1999:56) has complicated the identification of genre because male and female voices have registrational differences. A woman performing *Ìjálá* in head voice might be misinterpreted as an *ewi* poet, thus a high belt might be more appropriate for *Ìjálá*.
12.1.5 Ensemble Singing

Similar to solo performance, ensemble singing is often characterized by vocal style. Types of Zulu choral music are “frequently conceptualized by... vocal registers and timbres... *isikhwelajo* is characterized by high-pitched, almost yelling sounds, *cothoza mfana* and *isicathamiya*... generally feature more soft-touched, low-intensity vocals.” (Erlmann 1999:189) Recording ensembles drawing on the oral tradition continue to distinguish between the vocal styles that defined the various oral genres. Amateur and semi-professional choirs exist in churches, schools and universities and many fine composers who pay close attention to the details of language (such as tone contrasts) have emerged. However, the distinct vocal qualities that define the oral singing tradition have been lost in the written choral tradition.

The less formal *chorus* singing is now disconnected from the praise singers (or oral poets) that traditionally led call-and-response singing. Many Nigerian Christians and Muslims would refuse to listen to an *Ìjáìlá* poet praise the indigenous god of iron, *Ògún*, let alone sing a refrain in affirmation. Call-and-response choruses are now sung in large churches filled with thousands of worshipers, often alongside American Contemporary Christian music, which is part of the international growth of evangelical and Pentacostal Christianity.

12.1.6 Popular Music

Urban centers are now the attraction for the vocally gifted (Okpewho 1992:41). Youssou N'Dour descended from *griots*, but embarked on an international recording
career instead of learning the trade of his forebears (Duran 1989:277–8). In major cities like Addis Ababa, Dakar, Johannesburg, Kinshasa, Lagos, and Nairobi, there are recording studios everywhere, producing music locally that is consumed locally, sold by compact disc vendors in markets and on street corners. From the 1960s to the 1990s, Africans were largely attracted to imported music like James Brown, Bob Marley and Michael Jackson. Now, in the small shops that dot the neighborhoods, African-produced music overwhelms music from the outside. Because of the high premium for Internet access, Bluetooth file exchange between cellphones has become a primary means of sharing music. If youth are going to listen to hip-hop, they would rather it be hip-hop that relates to their lives and the complex linguistic environments they live in, which may include multiple indigenous languages in addition to formal and informal codes of English or French. Contrastingly, music on television has more of a Western sensibility with the proliferation of amateur singing competitions, including Project Fame and The X-Factor.

12.1.7 SUMMARY OF VOCAL TECHNIQUE

The oral tradition in Africa is really an *aural* and an *oral* tradition. It is a tradition of singers hearing and imitating sound, not just words. The alphabet and musical notation filter out many aspects of sound and are inept at representing sounds they were not designed for, such as languages with pitch contrasts or music in non-western scales. We are fortunate that technology has advanced to make recordings that nearly capture the full range of human hearing and vision, and has provided the means to share information with
others around the world. The academic study of African vocal arts is more complete with transcriptions and recordings because vocal style is cultivated as much as content or structure. Thus, the term oral poetry is a somewhat misleading way to refer to the historic voice profession in Sub-Saharan Africa.

### 12.2 Tone and Tune Elaborated

Many of the languages spoken in Sub-Saharan Africa are tone languages of the Niger-Congo A and B families. Ìgbọ and Yorùbá are the focus in this dissertation, but there are hundreds of others. Since Herzog (1934), ethnomusicologists and linguists have compared the tone contours of speech with the melodic contours of singing and found similarity. However, adherence to speech-melody by musicians varies greatly across and even within cultures (Agawu 1988). Schellenberg suggests there is a genre-dependent trend, with traditional music as the most faithful and contemporary popular music as the least faithful to linguistic features (2012:271). The pitch contrasts of speech may also influence vocal harmony because, for the sake of linguistic accuracy, similar contours are usually present in all renderings of a lyric. As nineteenth century missionaries found, the contrary motion favored by European hymnody, in which the same text is set to varying pitch contours, is at odds with the phonological systems of African tone languages (Èkwúèmé 1974c:337, Agu 1992:14). Similarly, strophic forms are not common in the music of tone languages because it requires different lyrics to follow the same melodic contours. Ìgbọ musician Laz Èkwùèmé identifies parallelism and antiphony (call-and-
response) as devices sympathetic to tone and thus common in African ensemble singing (Èkwúémé 1974:344–5).

12.2.1 What Is The Difference Between Tone And Tune?

Ignorance among missionaries about the importance of tone is manifested in metric translations of hymn texts into African languages. The singing of the translations to standard tunes has produced hymns of utter nonsense across West Africa (Parrinder 1956:37) as well Asia and the Americas. Violations wherein a melody ascends in pitch when the intended speech tone would descend are problematic. For a well-documented language like Yorùbá, with a published vocabulary available since 1843, one must speculate that those who undertook these translations had the knowledge and resources available to be more sensitive to tone, but chose to ignore it. The response among Yorùbá Anglicans was to add ‘native airs’ to the hymnbook, explained by Rev. J. J. Ransome-Kuti (grandfather of Fèlá Kuti) in the preface:

No [hymn] tune… can possibly express the meaning of words in a ‘tonic’ language such as Yorùbá, so well as one written specially for the words. (Ìwé Òrìn Mímó (Book of Holy Songs) 1923)

Recall von Hornbostel suggested that, in lieu of European hymns or hymns composed in a European fashion, converts be encouraged to ‘sing and play after their own natural manner’ (1928: 62). This musical practice was pioneered in the Africanized Aláduúrá churches early on, but it was not until the post-colonial era that the ‘natural manner’ spread to Catholic and Protestant churches where European hymns in both foreign and indigenous languages are sung alongside praise choruses today. Agu indicates that the
true fault of the hymns is not the linguistic defects—people have come to recognize the intended meaning—the problem is one cannot dance to them (1992:14). However, I have seen this shortcoming of hymns overcome on many occasions.

12.2.2 Musical Analogy

If one wonders whether the analogy between music and language in Yorùbá culture originated with Crowther’s comment in his preface to the 1852 orthography—"the Yorùbá language is very musical"—the answer is most likely no. Speech surrogate instruments that serve both as signal and entertainment are found in many tone language cultures in Africa, from the icon of Yorùbá culture, the double-membrane hourglass-shaped talking drums (dùndún), to sets of pitched ideophones among the Ìgbò (including the ògénè bells and ùdù pot drums). Long into the past, Yorùbá praise poets and dùndún players (who often work in tandem) must have understood a connection between the pitch effects of decreasing and increasing pressure of the leather bands of the drum and manipulating the human voice because they would mimic each other. Nigeria’s first professor of music Ígwë Laz Èkwùèmé was the fifth dissertation advisee of Allen Forte (Carson Berry 2009:214). Èkwùèmé rejects missionary-turned-musicologist A. M. Jones’ contention that Africans are “utterly unconscious of any organized theory” behind their music (1949:18). Èkwùèmé points out that the Chopi refer to a hombe, a tonal center around which notes revolve (Tracey 1948 in Èkwùèmé 1974b:35–36). Later in the same article, Èkwùèmé comments on scales:

[In] many cases the music of Sub-Saharan Africans is diatonic- that is, uses whole steps and half steps-but may be said to be modal in that the ordering of these
whole steps and half steps may not be in keeping with the ordering of the Western European major or minor scale. Scales may be tetradonic, pentatonic, hexatonic, or heptatonic. (Èkwúèmé 1974b:52)

Èkwúèmé's attempt to reconcile the diatonic scale with pre-colonial African music is anachronistic. However, the Tonic Sol-Fa method integrated quickly into many African cultures, so it seems likely that it was to some extent compatible with existing musical practices. The responsiveness of each culture to Tonic Sol-Fa likely varied, and may have done so in accordance with the nature of the instruments already present within the culture. Schubert and Wolfe (2014) argue that stable pitch in singing is influenced by fixed-pitch musical instruments. If this is the case, at least in some African cultures, stable-pitch in singing was likely already present in cultures with fixed-pitch instruments, such as the lamellophone (thumb piano). In others, Tonic Sol-Fa may well have been the introduction to singing on a relatively stable pitch. Èkwúèmé points out that spoken pitch is less definite than sung pitch and that slides and glissandi are present in singing, but states these “should be ignored in an attempt to determine a scale” (Èkwúèmé 1974b:52).

Speech surrogate instruments in Ìgbòland are predominantly fixed-pitch (as in the ogene bells and gongs, ese tuned drum rows and ùdù pot drums). Informants suggest speech tones in Ìgbò tend to be discrete (delineated and stable pitch) like these instruments. Thus, the assumption of Tonic Sol-Fa—people sing on level pitch—is compatible. However, in Yorùbá sloped pitches (falling or rising contour) are found within a vowel segment (Bamgbose 1990) and there are variable tension speech surrogates like the dùndún. Yet, Tonic Sol-Fa, including the notation, was heartily adopted and adapted into Yorùbá Christian culture and is now a common method for choral singing. While the
portamenti and glissandi of Yorùbá tones and talking drums are still extant, the solmization syllables *do-re-mi* have formed a heuristic for speech tones (see Chapter 2).

12.2.3 What is the depth of the mapping?

The notion that speech tone may determine musical melody, as Èkwúèmé (1974c) suggests, continues to fascinate many language and music researchers, including myself. The more interesting question about tone and tune is not whether tone does or should determine tune, but when it does occur, what is the depth of the mapping? Is it restricted to directional relationships, like Friedmann’s contour adjacency series? Is it also reflected in the size of melodic intervals or non-adjacent relationships? Are intonational trends as well as lexical tone realized in music?

To look for deeper tone-tune relationships, I first applied refined contour, which includes not only direction but magnitude in pairwise comparison of adjacent pitch events. Quintary contour categories (leap up, step up, same, step down, leap down) are one search parameter in Huron’s Themefinder algorithm that may be used to search for similar or contour equivalent melodies in the Essen folk song collection. Refined (or quintary) contour distinguishes between interlevels of LH and MH but not MH and ML.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Refined Contour</th>
<th>Tone Sequence</th>
<th>Interval Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>↑</td>
<td>+2</td>
<td>LH</td>
<td>Leap up</td>
</tr>
<tr>
<td>↑</td>
<td>+1</td>
<td>MH, LM</td>
<td>Step up</td>
</tr>
<tr>
<td>→</td>
<td>0</td>
<td>HH, MM, LL</td>
<td>Repeated</td>
</tr>
<tr>
<td>↓</td>
<td>−1</td>
<td>HM, ML</td>
<td>Step down</td>
</tr>
<tr>
<td>↓</td>
<td>−2</td>
<td>HL</td>
<td>Leap down</td>
</tr>
</tbody>
</table>

Table 12.1
Distinguishing between smaller and larger magnitude interlevels adds depth to analysis of tone-and-tune. However, while LH and MH and LM interlevels are all likely to yield ascending musical intervals, the magnitude of the musical intervals is not consistent from one composer or performer to the next. Interlevel realization may not be consistent within a single piece. As Chapter 5 demonstrated, different listeners may hear the same magnitude differently, likewise creative artists hear tone differently, and therefore realize it differently. Notions from scale theory of ambiguity and contradiction are useful here (Rahn 1991:36). If the interlevel LM (+1) was realized as a M3 (4 semitones) and the interlevel LH (+2) was realized as a M3, this causes ambiguity. If LH was also realized as a m3, this would be a contradiction of LM as a M3. A few ambiguities arise even in Laz Èkwúèmé’s careful realization of Ìgbò tones (see Chapter 15). Less methodical tone-tune mappings are rife with ambiguities and contradictions. In my experience, creative realizations of tone, whether chanted or sung, distinguish themselves from speech realizations, particularly in the clarity and stability of pitch. However, they are similar in that magnitude between adjacent tones is not a highly reliable indicator of what interlevel was intended or perceived (see Chapter 5). For these reasons, contour level analysis is preferred to methods that incorporate syntagmatic magnitude, such as refined contour. Through non-adjacent comparisons, it is possible to recognize a depth of mapping beyond simple-adjacency without using interval magnitude. In the coming chapters contour level analysis as well as linear correlation (Pearson's r) and refined contour are applied.
12.2.4 MAKING APPROPRIATE COMPARISONS

Conventional tone-tune analysis is to compare syntagmatic direction between a transcription of the underlying tones and a transcription of a melody. Using this method, Agawu (1988) finds evidence against tone-and-tune in Ewe. Schellenberg (2012) summarized a number of studies by linguists and music researchers who have used techniques similar to this and found a general trend that commercial and church music have low correspondence and traditional music has high correspondence. Ironically, some linguists who prefer a paradigmatic approach for phonological analysis use a syntagmatic approach for tone-tune analysis (e.g. Leben 1983). In response to Richards (1972), Leben shrewdly points out that comparing realized melody to underlying tone sequences is problematic because lexical tone is often altered when words interact in speech. Thus, according to Leben, one should consider phonotactics that affect tone realization in speech when analyzing tone realization in melody (1983:151). This suggests that comparing speech and melody, which are both realized tone, may be more appropriate than comparing realized tone in melody to underlying tone, which are in different realms. Comparing speech and melody is difficult because song lyrics are typically a rarefied, restricted and repetitive use of vocabulary that are not normal speech. It makes sense to make comparison between different versions of the same realized tone, but people do not usually speak in song lyrics. Gather (2013) points out that European hymns use syntactic dislocation, which makes hymn texts unlike speech. Poetry and lyrics are generally different from speech, but can you speak a poem or lyric? Yes. And, according to Barber (1990) among others, poetry in small excerpts, such as praise names (oríkì ọsọkì), often enters into everyday speech among the Yorùbá.
Recently, I recorded spoken and sung versions of 22 chorus texts, and also spoken versions of 18 hymn texts to compare with the melody of hymn tunes used by Holy Trinity Anglican Church Ikate-Lagos, recorded in October 2013. Unlike Locke’s study of Ewe texts (1992), an independent speaker (who is not a musician) was recorded speaking transcriptions of the song texts. The speaker, Adélékè Adéékó, is an expert on Yorùbá language and literature and is, quite importantly, able to read Yorùbá diacritics for tone (a rarefied skill actually). The Yorùbá choruses (popular praise songs) were performed by ‘Kúnlé Kuti, music director of Christ International Community Church in Columbus, Ohio. The melodies of the hymn tunes were encoded as MIDI data based on the recordings by Holy Trinity Anglican Church Ikate-Lagos. The hymn recordings could not be analyzed directly because there were four parts (SATB) with organ accompaniment.

To prepare the speech and choruc recordings for analysis, much like the other recordings presented thus far, the audio signal was segmented into tone-bearing units (syllable segments) manually in Melodyne. Then, the onset and offset values were exported as a MIDI file to be used in MATLAB. In MATLAB, the YIN algorithm was used to calculate mean fundamental frequency for each syllable segment of the continuous audio signal. The values were then converted into semitones (C4=60).

Three measures were used to compare the two versions of each text: the contour adjacency series (CAS), C+SIM with a range of degrees of adjacency, and Pearson’s $r$ (see Table 12.2 and Table 12.3). Similar to settings of the Ìgbò text “Oha Kele” (see Chapters 1 and 16), generally original tune-text combinations had higher correspondence (across measures) than translations of hymn texts sung to pre-existing European hymn tunes. A few of the hymn-tunes have low contour similarity and negative correlation,
suggesting the tonemes are unintelligible. There are a few hymns that may be somewhat intelligible. Some of the choruses had very high correspondence to the speech suggesting the melody was carefully created to follow the tone contour implied by the lyrics.

The lowest speech-song linear correlation of those examined is “Ase iṣẹ ọrun,” with a negative coefficient ($r = -0.436$). The text is 439 in *Iwe Orin Mimo* and the tune used by Holy Trinity is “Narenza.” In Figure 12.2 and subsequent plots, log pitch is normalized to the minimum and maximum value within each signal and mean frequency values for each syllable are plotted in time-order with durational information removed. The blue squares are for speech and the green circles are for song.

![Figure 12.2: Hymn 12: Spoken version of “Ase iṣẹ ọrun” text (blue) and Narenza hymn tune (green)](image)

In Northern European hymn-singing practice, a hymn text may be sung to any tune of the same meter (in the poetic sense meaning number of syllables not musical meter). It would make sense for choir directors to prefer hymn tunes that are a better fit for the speech tones, but that is not always the case. For one thing, correspondence in one verse may not equate to correspondence in all verses. In this analysis, only the first verses of hymns are studied.
As it happens, the hymn tune *Swabia* (see Figure 12.3) provided in *Iwe Orin Mimo* for “Ase ifẹ orun” has higher correlation for the first verse ($r = -0.087$). Whether no correlation (around 0) is more or less bothersome to tone language speakers than negative correlation is unknown. Once the tone is wrong, it may not matter how wrong it is.

Figure 12.4 has the strongest positive correlation and highest CSIM at 4, 6 and n-1 degrees. This is 136 in *Iwe Orin Mimo*, “Krist’ ki’jọba Ṛẹ de,” sung to the hymn tune prescribed in the hymnal, *St. Cecilia*. The close correlation between segments 7–11 is
astonishing, though most likely by chance. The speaker was not aware of what hymn
tunes were to be compared and is not a chorister. The rest of the hymn tunes studied, all
selected based on recordings by Holy Trinity Church in Ikate are in Table 12.2.

<table>
<thead>
<tr>
<th>Hymn</th>
<th>n</th>
<th>CAS</th>
<th>C+SIM by Degrees of Adjacency</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>-1/+1</td>
</tr>
<tr>
<td>1</td>
<td>27</td>
<td>0.478</td>
<td>0.609</td>
<td>0.593</td>
</tr>
<tr>
<td>2</td>
<td>28</td>
<td>0.417</td>
<td>0.583</td>
<td>0.607</td>
</tr>
<tr>
<td>3</td>
<td>49</td>
<td>0.512</td>
<td>0.659</td>
<td>0.602</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
<td>0.529</td>
<td>0.529</td>
<td>0.525</td>
</tr>
<tr>
<td>5</td>
<td>38</td>
<td>0.387</td>
<td>0.548</td>
<td>0.500</td>
</tr>
<tr>
<td>6</td>
<td>24</td>
<td>0.550</td>
<td>0.650</td>
<td>0.563</td>
</tr>
<tr>
<td>7</td>
<td>30</td>
<td>0.385</td>
<td>0.500</td>
<td>0.533</td>
</tr>
<tr>
<td>8</td>
<td>28</td>
<td>0.250</td>
<td>0.417</td>
<td>0.357</td>
</tr>
<tr>
<td>9</td>
<td>40</td>
<td>0.303</td>
<td>0.545</td>
<td>0.525</td>
</tr>
<tr>
<td>10</td>
<td>24</td>
<td>0.550</td>
<td>0.600</td>
<td>0.667</td>
</tr>
<tr>
<td>11</td>
<td>40</td>
<td>0.353</td>
<td>0.441</td>
<td>0.450</td>
</tr>
<tr>
<td>12</td>
<td>26</td>
<td>0.364</td>
<td>0.500</td>
<td>0.519</td>
</tr>
<tr>
<td>13</td>
<td>32</td>
<td>0.214</td>
<td>0.393</td>
<td>0.406</td>
</tr>
<tr>
<td>14</td>
<td>42</td>
<td>0.417</td>
<td>0.500</td>
<td>0.512</td>
</tr>
<tr>
<td>15</td>
<td>28</td>
<td>0.500</td>
<td>0.833</td>
<td>0.696</td>
</tr>
<tr>
<td>16</td>
<td>30</td>
<td>0.346</td>
<td>0.500</td>
<td>0.433</td>
</tr>
<tr>
<td>17</td>
<td>56</td>
<td>0.265</td>
<td>0.429</td>
<td>0.473</td>
</tr>
<tr>
<td>18</td>
<td>56</td>
<td>0.521</td>
<td>0.563</td>
<td>0.598</td>
</tr>
<tr>
<td>Mean</td>
<td>34.333</td>
<td>0.408</td>
<td>0.544</td>
<td>0.531</td>
</tr>
</tbody>
</table>

Table 12.2: Hymn to Speech comparison

In summary, the mean Pearson’s $r$ for all hymns is 0.184, which indicates that overall,
correlation between spoken versions of the hymn texts and the hymn tunes selected by
the church choir director is weak. The Yorùbá choruses, selected by a music director at a
non-denominational church which does not typically sing hymns, have an appreciably
mean correlation coefficient (0.407).
Of the choruses, “Aba mi nibi iyinrere” had the highest contour similarity at most window sizes.
Comparing two segments of the same signal, whether speech or music, is fruitful. So is comparing different realizations of the same text, such as the speech-song comparisons above. In many of the examples of Nigerian vocal arts examined in Chapters 13–15 the realization of tones is more systematic than in speech. In the case of choral composers, each composer may have their own system of realization. The analysis approach in the coming chapters is not to compare artistic realizations to non-artistic realizations of the same text. I do not have speech versions of the lyrics and I do not attempt to infer how the texts would be spoken. Instead, the primary form of analysis is to apply contour level reduction to the tone sequence of the transcribed text and the log pitch series of the artistic realization in order to compare them.
13. ORÍKÌ ÒGÚN

Introduction

In the Yorùbá language of southwestern Nigeria, *oríkì* can refer to a single praise name (Abraham 1962:481) or an entire string of “attributive epithets” (Barber and Waterman 1995:241). Because of the variety of vocal modes, styles and instrumental accompaniments associated with it, oríki defies classification. It has been studied as poetry, and studied as music. Historically, oríki was delivered by a specialist in a particular vocal style (Babalola 1966:vi). For example, *ìjálá* is acoustically open and intense while *ewi* is in a high falsetto, wailing voice quality (Vidal 2012:71). According to Waterman, “the words that placate gods and drive kings to suicide [are] made more potent by the patterning of timbre, texture, pitch, and rhythm” (1990:217). Yorùbás have oriki for “almost everything... even food” (Vidal 1969:56), but the praise of gods and mortals is the most lucrative for singers and engaging for audiences, whether it is received as persuasive or controversial. Votes are gained when a politician hires a well-known singer to elaborate their heritage. Christians and Muslims pray for salvation if they hear the praise of indigenous Yorùbá gods. This chapter presents findings consistent with trends observed in the past, but also unexpected turns in the ongoing and praise-singing culture of Nigeria. Similar to Waterman’s observations in the 1980s (1990), developments in music technology continue to be incorporated into neo-traditional music. Meanwhile digital video has made it possible to share interpretations of historic
performance practice on YouTube. Pan-Yorùbá identity, political elections, mass
celebrity and monotheism have changed the addressees, the focus on kings and
indigenous gods is diminishing, but the role of praise in religion and social hierarchy
largely remain intact. Oríkì continues to be a discourse on individual and collective
identities (Barber and Waterman 1995:241, 249).

13.1 Oríkì in the past

In Yorùbá culture, names are chosen for their meaning and are presumed to be fateful,

hence the proverb “orúkọ ni ì ro ọmọ” (a child's name affects him) (Ọlátunji 1984:68). A

possible etymology of oriki is a combination of ori, meaning the head or destiny, and ki

or ki, a salute or greeting (Abraham 1962:367, 371, 481). Thus, to call an oriki may also

be understood as hailing one's destiny. Common given names (orúkọ) include names of
good fortune:

Abiólá: one who is born into wealth,
Adékúnlé: one who possesses a home with many crowns;

names honoring Yorùbá gods:

Ôgúndana: Ògún has created a path,
Ifáṣeun: Ifá has done well / thanks be to Ifá;

and names honoring the Abrahamic god:

Olúwaṣeun: Lord has done well / thanks be to god
Olúṣégún: the Lord is victorious,
Tolúlpé: all thanks be to god.

Oríkì sókí are predetermined names based on the circumstances of birth, such as the

names for twins: Táiyé and Kéhìndé (Ọlátunji 1984:68).
From the humble beginning of a single orúkọ or oriki ọsọki, one’s personal collection of praise names grows throughout life. A person may accumulate as many as twenty praise names composed by different people at different times (Barber and Waterman 1995:251). In addition, a Yoruba person enters the world with oriki orilẹ, associated with place of origin, and oriki ǹdilẹ, detailing family heritage (Vidal 2012:56). Most oriki orilẹ pre-date the written record and are difficult to place historically, but some are believed to have originated in the thirteenth century (Babalola 1966:12). In contemporary society, Yoruba still form group identity around a common origin in an ancient town (Barber and Waterman 1995:260). Often, orilẹ, idilẹ and personal oriki are combined to make a greater whole, a complete portrait of one's homeland, lineage and life. This becomes an inventory of “praise units” that can be included, omitted, reordered and revised based on the occasion (Ọlátunji 1984:101–102, Barber 1991:508–9).

Proverbs (ǹwe) and riddles (àlọ) are freely incorporated as favorite utterances of the subject or to simultaneously illuminate and obscure his or her character (Barber 1991:510) enhancing the tendency of oriki to be esoteric or even cryptic.

John Comaroff described two codes among Tswana poets in South Africa. The formal code states “shared values and ideals,” the evaluative code discusses people and events in the present (1975:150). Kaschula extrapolated this model to apply to both imbongi of South Africa and the griots of the West African Sahel (1999:66). The formal code is static and the evaluative code is dynamic. We see both in oriki. Orilẹ and idilẹ are largely fixed while personal oriki is ideally evaluative, with epithets added based on new knowledge of the person or even happenings of the moment. Oriki is most potent when
poet and addressee belong to a community. The evaluative code relies on inter-personal relationships.

Urbanization and migration challenge the evaluative code in oríkì. While praise-singing remains part of collective life, the generation of new praise units is increasingly rare. Since mid-twentieth century, the personal oríkì of fathers have been recycled to praise sons (Barber 1991:246–7). Oríkì has been a staple of Yorùbá-language radio and television since the mid-twentieth century. These new mediums necessitate the use of and contribute to the formation of a more homogenous pan-Yorùbá identity. For commercial recordings or appearances on radio or television, it’s not desirable to be too specific. In nightclubs, where every single person is a possible addressee, an intimate knowledge is not possible. The singer is doing well to riff on the name, profession and place of origin of an addressee. Thus, oríkì artists adapt content for heterogeneous audiences, developing “a generic corpus of standardized filler texts” (Waterman 1990:18).

13.2 Oríkì Ògún: past in the present

Lagos, the former capital and current financial center, attracts Nigerians of all ethnicities and means, including an array of artists. While a large sector of the booming entertainment industry targets national and international audiences with English-language movies and music, Yorùbá-language media are still viable because the Yorùbá language is robust, even in cosmopolitan Lagos.

In contemporary society, the poet is often relegated to “cultural performances” (Okpewho 1992:25). Nollywood Yorùbá-language movies often call for traditional
dance, music and poetry, deemed necessary to depict a timeless sort of village life.

University theatre students study cultural performance alongside the works of modern playwrights such as Olá Rótimí, Femi Ôṣófisan and Wólé Ţóyínká. The works of Olá Rótimí and others incorporate cultural performance such as oríkì into dramatic literature.

Máyọwá Adéyẹmọ, a student in the Dept. of Theatre Arts and Music of Lagos State University, has developed skill as an ijálá poet. Ìjálá is known as the “hunter's call” and is largely associated with Ôgún, the God of Iron. The University does not have an oral-poet-in-residence to mentor students, so when we interviewed her, we were convinced the practice must have been handed down from generation to generation in her family. “No,” she said, “I went to the library and got on the Internet!” Adéyẹmọ's experience shows us Yorùbá poetry is no longer a strictly oral art-form. There are many resources for studying the art, from published collections such as Òrìkì Àwọn Timi Èdè to recordings on the Internet. Comparison of three recorded performances of Òrìkì Ôgún by Adéyẹmọ shows each rendition is nearly identical. This suggests her Òrìkì Ôgún is in Comaroff's formal mode, a recitation of a memorized text (the full text is available in Table 13.2). Adéyẹmọ's version is not exhaustive, it substantially shorter than other versions I have studied (such as that of Alabi Ogundepo), but it is fixed. This oríkì is in her performance repertoire and could be useful for a home movie or cultural event because Ôgún is a significant Yorùbá god.

For many Nigerians, praising Ôgún and other indigenous gods is distant from contemporary life, expelled by Christian and Muslim proselytization. One of my favorite lines about Ôgún (from Adéyẹmọ’s rendition) is “He has water at home, but bathes with blood”—what rich imagery! I have heard Yorùbá sermons warn against keeping the
charms of “lesser gods” hidden in your pocket: “Iránṣẹ kan ọ lè sin baba méjì.”

When Máyòwá Adéyémọ performed at the inaugural celebration of World Voice Day Nigeria on April 16, 2013, she began with Oriki Elédùmàrè (praise of the creator) and followed with Oriki Êjiré (praise of twins) before praising Ògún, the seminal oríkì for ìjálá artists. Praising the creator or twins is acceptable to Yorùbá Christians and Muslims while praising Ògún is not. Twins hold a special place in Yorùbá culture (see Mobolade 1971), an instance where Yorùbá cultural beliefs and Abrahamic religion co-exist.

Figure 13.1: Adéyémọ performing at the University of Lagos, April 13, 2013

74 From a sermon at All Saints Anglican Church Yaba on June 2, 2013, translation: “A servant cannot serve two masters.”
Tension was building already, but when Adéyemọ came to the refrain of Oríkì Ògún, the indignation of the audience escalated. Breaking out of the chanting vocal mode, Adéyemọ sang it five times. With each repetition the crowd’s roar grew:

Mo f'owó r'ajá a mọ ri b'Ògún lo de! I used money to buy a dog to worship Ògún outside!

Members of the audience were shouting “fire-fire” prayers to prevent their ears from perceiving “devil worship.” An oríkì battle ensued between Ògún and the Abrahamic god. She ended with “Ě so ti yín!” (Now you say your own).

Adéyemọ recites Oríkì Ògún as an historic cultural artform. As a Christian, she is more able to “freestyle” (a term for improvisation borrowed from American culture), for Elédùmàrè (Creator) or Olúwa (Lord). By adding new names that evince the Abrahamic god’s power in the present, she renders oríkì in the formal and evaluative codes. In Adéyemọ's repertoire, Oríkì Ògún is static, while Oríkì Olúwa is dynamic. A strong relationship between poet and subject brings the poetry to the present, deepens evaluation and enables the creation of new praise epithets. In short, oríkì that is generative.

13.2.1 INTONATION IN ÌJÁLÁ

In my time in Nigeria, I also had the privilege of meeting and making recordings of performances by Alabi Ogundepo. A distinguished researcher and performer within the Yorùbá poetry tradition, Ogundepo had a major role in documenting the oral tradition.
Figure 13.3 includes an analysis of a recording of Ogundepo reciting one of his many, many transcriptions. The windowed mean pitch never deviates more than three semitones from the global mean pitch, indicated by the dotted line. Figure 13.3 also includes an analysis of Máyòwá Adèyèmọ’s University of Lagos performance referred to in the previous section, and a sermon by Owadayo (first analyzed in Chapter 4). Adèyèmọ’s voice is considerably higher than the male voices, showing the performance practice of Ìjálá, also known as Hunter’s poetry, is expanding with new voices.
For the most part, Adéyẹmọ keeps within a tight radius around the global mean pitch, but also deviates widely at points. This is actually three consecutive performances, each fairly internally consistent in terms of range and tone dispersion, but distinct from the others, as if each is in a different key. The first is praise of the creator, Eledumare. The second is the praise of twins, Ejire, a little higher than the first. The final is the canonic Oriki Ogun, praise of Ogun, and boasts an even higher tessitura. If we compare both Ìjálá poets and the first sermon by Owadayo, it is almost as though the sermon enters an Ìjálá like vocal mode. The Ìjálá consistently occupies the high arousal portion of this model. And based on this and other evidence about the effects of paralinguistic intonation on tone levels, let’s make some revision to the Autosegmental account of tone levels. So for speech analysis, a syntagmatic model has the problem of initialization, and a paradigmatic model only works well for chant, and even then is really messy. So, let’s see what perception has to say about it.
13.2.2 COMPARATIVE ANALYSIS OF “ORÍKÌ ÒGÚN”

Finally, let’s return to Ìjálá poetry to compare multiple performances of the same text.

![Figure 13.4](image)

In Máyọwá Adéyemọ’s performance that we looked at earlier, three praise poems were rendered. The last was the praise of Ogun, the god of iron, who has water at home, but bathes with blood. Ìjálá by Máyowá Adéyemo

<table>
<thead>
<tr>
<th>Date</th>
<th>Notes on Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>4/16/13</td>
<td>Large Audience, Interactive, Handheld Mic</td>
</tr>
<tr>
<td>7/27/13</td>
<td>Video and Audio, No Audience, Ambient mic, Standing</td>
</tr>
<tr>
<td>7/27/13</td>
<td>Audio Only, No Audience, Gooseneck Mic, Sitting</td>
</tr>
</tbody>
</table>

Table 13.1

So I have three recordings of the praise of Ògún rendered by Adéyemọ, and a transcription of the poem into text by her. The first performance was in front of a large live audience, which is important because Ìjálá is quite interactive, and also of note she...
was using a handheld microphone and the recording is quite noisy. The second is a canned video recording a few months later with no audience and a distant directional microphone on the camera (video on YouTube: <https://youtu.be/2nEa9v5Np6k>). The third was recorded about an hour after the canned video with a high quality Countryman gooseneck condenser microphone. For the audio only, she was sitting down.

Each recording was segmented into 210 events that follow the first 24 lines of the text in Table 13.2.
This text is based on the July 27th, 2013 audio-only recording. From phrase (line) 25, the April 16th performance diverges. That performance is unique in that a large audience was present. The audio only recording is unique in that Adéyemọ was sitting.

Table 13.3: Similarity and Correlation calculated in comparison to 7/27 1st Perf. (Video)

<table>
<thead>
<tr>
<th>Date</th>
<th>Notes on Performance</th>
<th>C+SIM (209 deg.)</th>
<th>C+SIM (1 deg.)</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>4/16/13</td>
<td>Large Audience, Standing</td>
<td>0.913</td>
<td>0.963</td>
<td>0.950</td>
</tr>
<tr>
<td>7/27/13 (2nd)</td>
<td>No Audience, Sitting</td>
<td>0.952</td>
<td>0.973</td>
<td>0.979</td>
</tr>
</tbody>
</table>

Table 13.3 includes some of the measures applied to the spoken and sung versions of the Christian choruses: contour similarity at 1 and n-1 degrees and Pearson’s $r$ linear correlation coefficient. The canned video (recorded on 7/27) was used as the model because it had aspects in common with the other two. The performances from the same day are more similar in terms of contour than the live (audience present on 4/16) and canned (no audience on 7/27) standing performance. All three were recorded with different microphones and with different amounts of environmental noise present. However, I do not think that had an appreciable impact on the fundamental frequency analysis.

Table 13.4

<table>
<thead>
<tr>
<th>Date</th>
<th>Performance Notes</th>
<th>Min</th>
<th>Max</th>
<th>$\mu$ pitch</th>
<th>$\sigma$</th>
<th>$\mu$ int</th>
<th>$\mu$ dur</th>
</tr>
</thead>
<tbody>
<tr>
<td>4/16/13</td>
<td>Video, Audience, Interactive, Handheld Mic</td>
<td>C4</td>
<td>C5</td>
<td>67.5 (G4+)</td>
<td>2.640</td>
<td>2.215</td>
<td>0.244</td>
</tr>
<tr>
<td>7/27/13</td>
<td>Video, No Audience, No Mic, Standing</td>
<td>B3</td>
<td>C5</td>
<td>67.3 (G4+)</td>
<td>2.701</td>
<td>2.421</td>
<td>0.253</td>
</tr>
<tr>
<td>7/27/13</td>
<td>Audio, No Audience, Gooseneck Mic, Sitting</td>
<td>Bh3</td>
<td>Bb4</td>
<td>65.3 (F4+)</td>
<td>2.766</td>
<td>2.464</td>
<td>0.278</td>
</tr>
</tbody>
</table>
The same day performances were more similar in terms of contour and some other measures. The recordings in which Adéyemọ was standing were much closer in terms of mean pitch and range, though months apart.

The holy grail of autosegmental theory and a paradigmatic approach to modeling tone languages is showing that the same speaker speaks the same text at the same pitches on different occasions. If this were consistently observed, frequency bands would be a highly descriptive, even explanatory, model. Pitch similitude happens, but not reliably. Diana Deutsch (2004, 2006) has linked speaking a tone language to a greater incidence of absolute pitch perception. A paradigmatic approach is less problematic for Ìjálá than for casual speech or the Yorùbá sermons because the register is much more stable within a single poem and declination is to some extent mitigated. Although declination is less of a factor, physiology does still seem to play a role in production. When Adéyemọ sat down, her overall register dropped by about two semitones. This suggests that the physiology of
production may override pitch memory and perceptual feedback in terms of establishing an intonational register for ìjálá.

Based on Adégé’s performances, body position, specifically whether one is sitting or standing, has greater impact on intonation than whether an audience is present. I propose ìjálá belongs to the high arousal end of the intonation model, and to really belt it out with a high tone-level range, one should stand.

<table>
<thead>
<tr>
<th>Correlation</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tones to Pitch</td>
<td>0.658</td>
<td>0.679</td>
<td>0.690</td>
</tr>
<tr>
<td>Contour Levels (–1/+1)</td>
<td>0.754</td>
<td>0.776</td>
<td>0.794</td>
</tr>
</tbody>
</table>

Table 13.5: Correlation between tones (encoded LMH=012) to MIDI pitch values, and correlation between contour level reductions of each.

Having a transcription by the poet and three performances of that text offers a great opportunity to explore the depth of the mapping from tone-to-tune and determine the
extent to which the contour of the pitch realization is preserved from performance to performance.

Table 13.6 includes contour similarity values for various degree windows (indicated in the leftmost column), comparing each performance to the underlying and to each other. The numbering (1,2,3) refers to the chronological order of the performances (listed in Table 13.4).

<table>
<thead>
<tr>
<th>Degrees</th>
<th>T/1</th>
<th>T/2</th>
<th>T/3</th>
<th>1/2</th>
<th>2/3</th>
<th>3/1</th>
</tr>
</thead>
<tbody>
<tr>
<td>–1</td>
<td>0.870</td>
<td>0.888</td>
<td>0.888</td>
<td>0.953</td>
<td>0.972</td>
<td>0.944</td>
</tr>
<tr>
<td>–1/+1</td>
<td>0.738</td>
<td>0.771</td>
<td>0.766</td>
<td>0.888</td>
<td>0.939</td>
<td>0.864</td>
</tr>
<tr>
<td>–2</td>
<td>0.715</td>
<td>0.757</td>
<td>0.771</td>
<td>0.869</td>
<td>0.935</td>
<td>0.836</td>
</tr>
<tr>
<td>–2/+1</td>
<td>0.615</td>
<td>0.662</td>
<td>0.667</td>
<td>0.808</td>
<td>0.906</td>
<td>0.756</td>
</tr>
<tr>
<td>–2/+2</td>
<td>0.528</td>
<td>0.566</td>
<td>0.566</td>
<td>0.717</td>
<td>0.858</td>
<td>0.689</td>
</tr>
<tr>
<td>–4</td>
<td>0.542</td>
<td>0.580</td>
<td>0.571</td>
<td>0.769</td>
<td>0.877</td>
<td>0.750</td>
</tr>
<tr>
<td>–3/+2</td>
<td>0.464</td>
<td>0.512</td>
<td>0.507</td>
<td>0.673</td>
<td>0.829</td>
<td>0.654</td>
</tr>
<tr>
<td>–3/+3</td>
<td>0.362</td>
<td>0.452</td>
<td>0.419</td>
<td>0.624</td>
<td>0.795</td>
<td>0.605</td>
</tr>
<tr>
<td>–6</td>
<td>0.429</td>
<td>0.481</td>
<td>0.462</td>
<td>0.705</td>
<td>0.829</td>
<td>0.686</td>
</tr>
<tr>
<td>–4/+3</td>
<td>0.306</td>
<td>0.402</td>
<td>0.349</td>
<td>0.593</td>
<td>0.770</td>
<td>0.569</td>
</tr>
<tr>
<td>–4/+4</td>
<td>0.298</td>
<td>0.370</td>
<td>0.317</td>
<td>0.548</td>
<td>0.731</td>
<td>0.510</td>
</tr>
<tr>
<td>–8</td>
<td>0.327</td>
<td>0.394</td>
<td>0.356</td>
<td>0.606</td>
<td>0.736</td>
<td>0.601</td>
</tr>
<tr>
<td>–5/+4</td>
<td>0.285</td>
<td>0.353</td>
<td>0.295</td>
<td>0.546</td>
<td>0.710</td>
<td>0.498</td>
</tr>
<tr>
<td>–5/+5</td>
<td>0.248</td>
<td>0.296</td>
<td>0.252</td>
<td>0.524</td>
<td>0.684</td>
<td>0.490</td>
</tr>
<tr>
<td>–10</td>
<td>0.243</td>
<td>0.262</td>
<td>0.277</td>
<td>0.539</td>
<td>0.689</td>
<td>0.500</td>
</tr>
<tr>
<td>–6/+5</td>
<td>0.215</td>
<td>0.268</td>
<td>0.249</td>
<td>0.507</td>
<td>0.673</td>
<td>0.473</td>
</tr>
<tr>
<td>–6/+6</td>
<td>0.230</td>
<td>0.235</td>
<td>0.245</td>
<td>0.490</td>
<td>0.637</td>
<td>0.461</td>
</tr>
<tr>
<td>–12</td>
<td>0.206</td>
<td>0.211</td>
<td>0.250</td>
<td>0.529</td>
<td>0.623</td>
<td>0.461</td>
</tr>
<tr>
<td>–7/+6</td>
<td>0.202</td>
<td>0.217</td>
<td>0.236</td>
<td>0.468</td>
<td>0.601</td>
<td>0.443</td>
</tr>
<tr>
<td>–7/+7</td>
<td>0.193</td>
<td>0.213</td>
<td>0.218</td>
<td>0.450</td>
<td>0.574</td>
<td>0.436</td>
</tr>
</tbody>
</table>

Table 13.6: CONTCOM C+SIM

Values for the –1 window (top row) are all fairly high. T/2 and T/3 comparisons yield the same similarity quotient, and 2/3 has the highest similarity. Performances 2 and 3 are from the same day. The second performance (2) has the highest contour similarity to the underlying tones (T) for most windows, with some notable exception of –2, –10 and –12 in which the third performance (3) is more similar.
Declination does not effect Yorùbá poetry as much as speech, however there still may be pitch reset between phrases. Certainly, there is less attention to either paradigm preservation or interphrasal syntagms than to tone paradigms and syntagms within the phrase. In the next subsection (12.2.3), I will compare the third performance, for which the audio quality is the highest, to the underlying tone phrase-by-phrase, and even segment-by-segment. First, consider phrase comparisons for the two 7/27 performances (in Table 13.7). These had the highest contour similarity for entire series comparisons.

Seven out of 24 have equivalence at n–1 degrees (conventional C+SIM). For a single degree radius (–1/+1) there are 15 equivalent phrases, for a two-degree radius, there are 10, and for a three-degree radius, 4. NaN values indicate that the window is larger than the cardinality. That was the case for three of the –3/+3 windowed comparisons. The same NaN phrases had equivalence at n–1 degrees, accounting for the lower number for a three-degree window (7–3=4). A 6–degree window (–3/+3) is a bit too large for phrases of this cardinality, but each of the other window sizes has merits. The 2–degree window (–1/+1) only includes adjacent comparisons, but equivalence indicates that “pivots” (or turning points) are found in analogous locations. A 4–degree window reduces the number of equivalent phrases from 15 to 10, but adding non-adjacent comparisons raises the similarity rating for phrases 6 and 13. Non-adjacency may be necessary to distinguish local minima that are M from L and local maxima that are M from H (see Chapter 11).
In the next section, I consider the transcription by the poet, which I checked with several dictionaries. However, a dictionary is somewhat immaterial to Yorùbá poetry. There is poetic license to distort tone, and there are many words rarely used outside of poetry, particularly in contemporary urban culture. Performances such as this set the standard for ìjìnlè Yorùbá because there is no more authoritative source. That may change as more renditions of Oríkì Ògún and other canonic poems are recorded and/or transcribed.

### 13.2.3 Closer Analysis of Adéyemọ’s “Oríkì Ògún”

This section analyzes the high-quality recording of Adéyemọ’s “Oríkì Ògún” from July 27, 2013. Table 13.8 reveals that the mean pitch for tones of a certain level are ordered as

<table>
<thead>
<tr>
<th>Phrase</th>
<th>n</th>
<th>–1/+1</th>
<th>–2/+2</th>
<th>–3/+3</th>
<th>n-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>1.00</td>
<td>NaN</td>
<td>NaN</td>
<td>1.00</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>1.00</td>
<td>1.00</td>
<td>0.983</td>
<td>0.944</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>0.900</td>
<td>0.950</td>
<td>0.950</td>
<td>0.956</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>1.00</td>
<td>1.00</td>
<td>0.917</td>
<td>0.964</td>
</tr>
<tr>
<td>5</td>
<td>12</td>
<td>0.958</td>
<td>0.896</td>
<td>0.917</td>
<td>0.909</td>
</tr>
<tr>
<td>6</td>
<td>8</td>
<td>0.875</td>
<td>0.906</td>
<td>0.938</td>
<td>0.893</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>1.000</td>
<td>1.000</td>
<td>NaN</td>
<td>1.000</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td>1.000</td>
<td>1.000</td>
<td>NaN</td>
<td>1.000</td>
</tr>
<tr>
<td>9</td>
<td>15</td>
<td>0.967</td>
<td>0.917</td>
<td>0.922</td>
<td>0.890</td>
</tr>
<tr>
<td>10</td>
<td>8</td>
<td>0.938</td>
<td>0.938</td>
<td>0.958</td>
<td>0.929</td>
</tr>
<tr>
<td>11</td>
<td>8</td>
<td>0.938</td>
<td>0.938</td>
<td>0.958</td>
<td>0.946</td>
</tr>
<tr>
<td>12</td>
<td>12</td>
<td>1.000</td>
<td>0.979</td>
<td>0.958</td>
<td>0.955</td>
</tr>
<tr>
<td>13</td>
<td>13</td>
<td>0.923</td>
<td>0.962</td>
<td>0.962</td>
<td>0.955</td>
</tr>
<tr>
<td>14</td>
<td>7</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>15</td>
<td>8</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>16</td>
<td>9</td>
<td>1.000</td>
<td>0.944</td>
<td>0.963</td>
<td>0.958</td>
</tr>
<tr>
<td>17</td>
<td>9</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>18</td>
<td>11</td>
<td>1.000</td>
<td>0.977</td>
<td>0.985</td>
<td>0.991</td>
</tr>
<tr>
<td>19</td>
<td>10</td>
<td>1.000</td>
<td>1.000</td>
<td>0.967</td>
<td>0.978</td>
</tr>
<tr>
<td>20</td>
<td>7</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>21</td>
<td>13</td>
<td>0.923</td>
<td>0.904</td>
<td>0.910</td>
<td>0.891</td>
</tr>
<tr>
<td>22</td>
<td>10</td>
<td>0.950</td>
<td>0.950</td>
<td>0.967</td>
<td>0.967</td>
</tr>
<tr>
<td>23</td>
<td>7</td>
<td>1.000</td>
<td>1.000</td>
<td>0.952</td>
<td>0.976</td>
</tr>
<tr>
<td>24</td>
<td>8</td>
<td>1.000</td>
<td>0.969</td>
<td>0.979</td>
<td>0.964</td>
</tr>
</tbody>
</table>

Table 13.7: C+SIM by phrase between 7/27 performances
one might expect from low to high. The means are about a standard deviation apart, which indicates that there is much overlap in the absolute pitch of tones within a level.

<table>
<thead>
<tr>
<th>Tone Level</th>
<th>Mean</th>
<th>St. Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>67.7</td>
<td>1.8</td>
</tr>
<tr>
<td>Mid</td>
<td>65.1</td>
<td>1.9</td>
</tr>
<tr>
<td>Low</td>
<td>63.1</td>
<td>2.4</td>
</tr>
</tbody>
</table>

Table 13.8

Instances of tonal counterpoint offer an opportunity to compare analogous syllables with like or unlike tone.

<table>
<thead>
<tr>
<th>Yorùbá</th>
<th>Tones</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ò fìkán sánko</td>
<td>H LH HM</td>
<td>One for cutting grass</td>
</tr>
<tr>
<td>Ò fìkán yènà</td>
<td>H LH LL</td>
<td>One for making marks</td>
</tr>
</tbody>
</table>

Table 13.9

This example of tonal counterpoint follows Ọlátunji’s description, the HM at the end of the first phrase map to LL at the end of the second phrase. H and M are both sent to L. The phonic content of the last word is also changed, so it is not nonsensical word play but a more sophisticated instance of the device, wherein each sentence has distinct meaning. The first three syllables of each parallel phrase of Table 13.9 are identical in phonic and tonal content. The contour equivalence of these three elements is apparent in Figure 13.7. However, neither the f0 trace or mean pitch has the same absolute pitch height in each. The mean pitch is about two semitones higher for the first three syllables in the consequent phrase. Also note that segments 2–4 in the antecedent phrase are realized with
<012> CSEG though the underlying tone is LHH. This is quite common in speech and is addressed further in the conclusion (Chapter 17).

By producing contour level reductions of both the tone level sequence and pitch series for each phrase I have meta-paradigmatic values for each element of a phrase segment, based on both underlying tone and the pitch realization. The contour levels for each can be compared one-to-one to see if they are a match. Table 13.10 compares tone to tune at the smallest possible level: the tone-bearing segment. 1 indicates underlying tone and pitch realization produce the same contour level within the 0–2 range for a 2–degree symmetrical window, 0 indicates they did not. NaN (not a number) entries are present when the phrase has fewer elements than the segment number indicated by the column. Initial tones are the first five of the phrase, and final tones are the last five. For phrase 2,
which has 10 elements, all are present. For phrase 5 which has 12, match values for elements 1–5 and 8–12 are shown. The rightmost column ($\mu$) is the mean of 0/1 matches for the phrase, this may also be taken as CLSIM. A 1.000 value, returned by 7 out of 24 phrase comparisons, indicates complete agreement between the CL$_2$ series for the tone and the tune. These are not the same seven that had n–1 equivalence for the comparison of performance 2 to performance 3.

<table>
<thead>
<tr>
<th>Phr</th>
<th>n</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>$\mu$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>NaN</td>
<td>NaN</td>
<td>NaN</td>
<td>NaN</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.000</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.600</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.750</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>12</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.917</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0.875</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.800</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.600</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.800</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>8</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.750</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>12</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>13</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.846</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>9</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0.889</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>9</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.778</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>11</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>10</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.800</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>13</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.923</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>10</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.900</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.875</td>
<td></td>
</tr>
<tr>
<td>$\mu$</td>
<td>9</td>
<td>1.000</td>
<td>0.958</td>
<td>0.875</td>
<td>0.826</td>
<td>0.870</td>
<td>0.957</td>
<td>0.870</td>
<td>0.750</td>
<td>0.625</td>
<td>0.958</td>
<td>0.855</td>
</tr>
</tbody>
</table>

Table 13.10: Comparison of CL2 (+1/–1) values for corresponding tones and realized pitch (1 = match, 0 = different)

The bottom row indicates that the 1$^{st}$, 2$^{nd}$, 5$^{th}$ from the end and last elements had contour level equivalence to their underlying tone in over 95% of the phrases (at least 23 out of 24 phrases).
Calculating CL$_2$ similarity for each phrase and averaging the values returns a higher quotient (.855) than calculating similarity between unsegmented tone sequence and pitch series (.794 from Table 13.5). The lowest CL$_2$ similarity is .600 in phrases 3 and 8. Phrase 8 is the consequent phrase of the tonal counterpoint example above. The tone is not contradicted in the pitch realization, but declination has a strong presence in this particular phrase, as it might in speech.

The longest phrase with CL$_2$ equivalence (1.000) is phrase 19 with cardinality 11.

<table>
<thead>
<tr>
<th>Yorùbá</th>
<th>Tones</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ògún élémonà n'í gbeèsun'su</td>
<td>LHMHMLHMLMM</td>
<td>The Ëlemóna offer Ogun roasted yam</td>
</tr>
</tbody>
</table>

Table 13.11: Phrase 18

In Yorùbá poetry, declination does not distort tone level to the same extent it does in speech. In some of the short phrases, equivalent tones could be fit to frequency bands within the phrase, and there is no sign of declination. That is not the case for phrase 19, yet the tones are clear in terms of local adjacent and non-adjacent contour. The last three M level tones (indicated by the green dotted line in Figure 13.8 are at roughly the same pitch as L level tones earlier in the phrase. This does not produce ambiguity in the signal because the proximal L level tone is also lower. This is a paradigm-preserving use of a dynamic register.
Adéyěmọ represents a new generation of ijálá artist on multiple levels. First, ijálá is known as the hunter’s call and usually associated with male praise singers. On the other hand, it is not uncommon for a female praise singer to praise a male deity. The Abrahamic god is generally anthropomorphized as male among Nigerian Christians and Muslims. She honed her skills partially through YouTube videos and also recognizes a connection between Yorùbá oral poetry and hip-hop, referring to improvisation as “freestyling.” Taking all this into account, it is difficult to assess how neo- and how traditional, her neo-traditional performance is. To the extent that her performances reflect performance practice of the past, they indicate that tone realization is carefully practiced in Yorùbá poetry. Whilst elder generations complain that youth cannot speak ijinlè (deep) Yorùbá, we can only hope that young performers like Adéyěmọ continue to have interest in this historic practice and that those offended by the praise of Ògún can at least appreciate the vitality ijálá brings to the Yorùbá language.
14. CONTEMPORARY PRAISE-SINGING

14.1 Praise of “Big Men”

Nigeria is now a fairly saturated media market, with a film and music industry, radio and television stations, many newspapers and widespread internet usage (especially on smartphones). Though Nigerian culture has changed a lot, oriki has integrated with new media as it has emerged. Notably, 'Lánre Adébojú popularized the form in radio in the 1970s and the musical genres Fújí and Jújú are primarily associated with praise-singing. Thus, it is hard to call it orature anymore because it is often transmitted through recordings. Contrary to other transitions out of strict orality in the past, wherein orature was mapped to the written word, Yorùbá poetry has maintained its aurality. Oriki has remained vital because of its ability to build reputations--although in the scholarly sense, oral poetry may not be an authoritative record, for much of the public, it is.

Fújí and Jújú have been popular since the 1960s. Though aurally distinct, the genres share two common features with one another and oriki: they are largely panegyric and feature exchange between lead vocalist and lead drummer, performing on a talking drum in speech surrogacy.76 At large events, the Fújí or Jújú praise-singer has overtaken the less-equipped poet. With amplification, a voice can be heard praising a big man from far off.

76 There are a variety of “talking” drums used in Yorùbá music. Fújí uses sákárá and drums from the bátá ensemble. Jújú uses drums from the dundún and gangán ensembles.
At an aríyà (life-cycle celebration), band captains are expected to know biographical information for each important guest. If they do not, the band manager gathers information on key participants. Along with themes of prayer (àdúrà), money (owó), honor (olá), individual destiny (orí), jealousy (ílara), and competition (ídíje), the band captain sings praises (iyìn) for prominent figures and abuses (èébú) for their enemies (òtá) (Waterman 1990:186). A successful praise “swells the head” of the subject. Often, he or she will rise up to dance (perhaps even on the stage) and “spray” money on the singer's head and neck.

The primary source of income for Fúji and Jùjú musicians is parties for businessmen and politicians, not sales of recordings. Money earned from praising big men has made many Jùjú and Fúji singers gbajúmòn themselves, creating a dichotomy between dual identities as big men and beggars (Waterman 1990:22).

While Yorùbá music is aurally diverse (Thieme 1969:38–9), social criteria are the most important to musicians and audiences in characterizing music, particularly who plays the music, who patronizes the musician, and where and when it is played (Waterman 1990:16). Fúji and Jùjú are distinct musical genres, but both are referred to as “big manism” in conversations about contemporary culture. In the 1970s and '80s, King Sunny Adé, Chief Ebenezer Obey and Alhaji Barrister often dedicated albums to a big man, sometimes plastering the big man’s face on the cover along with the musician’s own. Alaja-Browne (1989:238) offers “Dr. Şehindemi” as an example of “overt publicity” from early in King Sunny Adé’s career.

77 Money from sales of recordings is in large part siphoned off by piracy.
Table 14.1

<table>
<thead>
<tr>
<th>Yorùbá</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>B’áláboyún bá lo sódòre, á bì wèrè, 78</td>
<td>Pregnant women treated by him are delivered easily,</td>
</tr>
<tr>
<td>Ṣèhìndèmì, doctor tò mòye</td>
<td>Shehindemi is a brilliant doctor.</td>
</tr>
<tr>
<td>B’áláisàn bá lo sódòre, á sàn kíá,</td>
<td>Sick people treated by him get well quick,</td>
</tr>
<tr>
<td>Ṣèhìndèmì, doctor tò mòye</td>
<td>Shehindemi is a brilliant doctor.</td>
</tr>
<tr>
<td>Èyìn onitúlú,</td>
<td>Whoever is sick, 79</td>
</tr>
<tr>
<td>E jé lo fún ’wosàn kíá,</td>
<td>Should hurry to him for treatment.</td>
</tr>
<tr>
<td>Kò sóhun méjì tó wón j’áláfíà lo.</td>
<td>For good health, he is second to none!</td>
</tr>
<tr>
<td>Ṣèhìndèmì Doctor, (2x)</td>
<td>Shehindemi Doctor,</td>
</tr>
<tr>
<td>Ṣèhìndèmì Doctor, mò gbà fún o</td>
<td>Shehindemi Doctor, I congratulate you,</td>
</tr>
<tr>
<td>Ṣèhìndèmì doctor o.</td>
<td>Shehindemi Doctor oh.</td>
</tr>
</tbody>
</table>

Praise-singing grew from building local reputations in the 19th century to active commercialism in the late 20th century. While oríkì as business marketing is new, the draw of praise-singers to royal and aristocratic people is not (Finnegan 1964:111). Yet, “big-manism” has become a derisive term for what is perceived as the problem with praise-singing in contemporary culture. Although panegyrics in some form are common to all human cultures, Yorùbá culture does seem to have an exceptional proclivity. The conveyance of praise from oral tradition to popular music genres may be somewhat unique to Yorùbá culture and has influenced the popular music of proximal ethnolinguistic groups. Through the 1950s and 1960s, the ubiquitous West African genre Highlife had praised beautiful women and musicians praised themselves, but had avoided public endorsement. The success of Jùjú musicians in the 1970s may have influenced Ìgbò guitar highlife musician Oliver de Coque to record “People's Club of Nigeria” with His Expo '76 Ogene Sound Super of Africa.

---

78 Starts at 2:30 in the recording.
79 Tùlu refers to neuralgia.
14.2 Praise-singing and Politics

In modern democratic election cycles, praise-singers are often employed by campaigning politicians. If there are two candidates, each may hire their own praise-singer, usually a Fúji singer who will make recordings and perform at events. Popular Fúji artists often have long-standing rivalries and these are mirrored by the political rivalries, which for the election period, are engendered in music. If the musical rivalry is not perfectly matched in terms of the popularity of the singers (especially within a voting area), that may translate into victory for the candidate who hired the more popular singer.

The incumbent governor of Ògùn state, Ìbíkúnlé Amósùn, hired the Fúji singer Ọṣefiu Ṣefiu Àlàó for the current election cycle. Unlike many Fúji singers based in Lagos who convey a certain cosmopolitanism, Àlàó rejects modernity and the draw of the megalopolis, choosing to stay in Abéokúta, the capital of Ògùn state. This has endeared Àlàó to the residents of Ògùn state. Thus, Amósùn's association with Àlàó gives him an added advantage with voters. Àlàó calls himself ọmo ọko, “the real bushman.” While ‘bushman’ is often pejorative among more cosmopolitan Nigerians, Àlàó uses the term as an oríki for himself, evincing his ọjọ̀nlè (deep) sense of Yorùbá culture. It is a rather ingenious appellation for one who specializes in a neo-traditional art. When we visited him in the studio in 2014, he was recording a song in support of Amósùn (the incumbent governor of Ògùn state).
The song he was recording is an oríkì orílè for Amósùn, in praise of the governor's family lineage and place of origin. It is also a carefully calculated response to a campaign concern: in the last election, people outside of Amósùn's father’s local-government area (LGA) and the nearby region did not support him as strongly as those within the area. The recording is a revisionist oríkì orílè, establishing a new intertextual connection to the oríkì orílè of Amósùn’s mother, which will include mentions of her place of origin. The goal is to impress the voters with the incumbent’s connection to his mother’s birthplace. Although Amósùn won the 2011 election, this campaign move increases the likelihood of success in 2015 by swaying voter’s in and around his mother’s LGA. In America, politicians make similar appeals to the public, letting them know that they were born here or there, and the endorsement of a local celebrity certainly (who may be working for the campaign) certainly helps. So, what is the significance of this politically-motivated message being formed into an oríkì orílè? Well, it would not be as effective in a state with an ethnically diverse population. Politicians hire praise-singers for campaigns...
because presentation matters. Oríki orílè lends veracity to a politician's claim of heritage and connection to a village that simply stating it does not. Oríki has adapted to political marketing in democratic elections, and in Yorùbáland, politicians make similar appeals to the appeals made elsewhere. The meeting of political campaigning and oríki performance is not a novel phenomenon, it is a natural progression from the past. Nowhere in our interview with Ṣefiu Àláó did he indicate that there was anything unconventional about modifying Amósùn’s oríki orílè to increase his chances at election. Àláó was eager to discuss this particular song, though he was recording others that day. Similarly, his fans would not likely find it untoward or manipulative. Professional praise-singing (Fújì and Jújú music included) has been dismissed by cosmopolitan Nigerians because of the quid-pro-quo nature of paying someone to sing praises. However, the marketing of reputation by elite singers for powerful people is not patently deceptive. The relationship between Àláó and Amósùn is quite transparent and Amósùn’s relationship to his mother’s village is not a fabrication. Furthermore, the message that Amósùn cares about his mother’s side as much as his father’s side may be very sincere and germane to his political agenda beyond winning the election. The most manipulative aspect of a politician hiring a praise-singer may be the harnessing of song to convey a message. Although the goal of gaining more votes remains the same, using oríki orílè instead of simply saying “my mother's from your town so you should vote for me,” substantially changes the message and its effectiveness. Thus, we see exchange between art patronage and political patronage: the politician patronizes the musician financially, the musician patronizes the politician by attracting his audience as voters. Ìbíkúnlé Amósùn was reelected as governor of Ògùn state in April 2015.
14.3 Analysis of Fújì Singing

Schellenberg (2012) concluded the degree of correspondence between speech and melody is to a large extent dependent on genre. Cross-cultural comparison showed folk or traditional music has higher correspondence than commercial music. Regarding Yorùbá music, informants report Contemporary Christian, Naija Hip-Hop, translated European hymnody have very low correspondence, while Jújú, and particularly Fújì have high correspondence, similar to Yorùbá poetry and choral music. Fújì defies generalizations that are otherwise appropriate. Among the recordings analyzed here, recent Fújì music by Saheed Òṣùpá has more tone-tune correspondence than even earlier examples from the same genre. Granted, it is a very small sample. However, the findings are consistent with discussions with musicians and enthusiasts. While I cannot generalize from these brief excerpts, my thought is that a practice of deep tone-tune mapping is not turned on or off, it is embedded in a genre, and even more fully in individual style.

In a 2002 article, noted Yorùbá music expert Christopher Waterman explores the cultural identities associated with Afrobeat, Fújì and Jújú music, lumping Fújì/Jújú together under the “Big Man” identity. While it may be true that the practitioners of these two genres spend much of their careers portraying themselves as, or praising the names of, big men, they are aurally distinct genres. As a forthcoming study by David Àíná and myself shows, listeners are very capable of differentiating between Fújì and Jújú music, except when the voice is not present. This is largely because the instrumentation, particularly the percussion, of these two genres is almost identical, but as our study shows, the singing is not. Vidal (2012) refers to Fújì vocals as Islamicized singing, and it is true that Fújì is typically produced and consumed by Muslims. However, I instead
attribute the character of Fúji to the fact that it is not westernized, unlike Jùjú, which like so much African music, has been infiltrated by the Western tonal music system, largely through Christianity. I propose that Fúji is free from the conventions of Western music, and therefore able to be conc (concentrated) Yorùbá in a way that Jùjú is not. Fúji follows many of the same conventions of Yorùbá oral poetry that Ọlátunji (1984) and others have identified, revealing that language is a focal point of this music. While the claim that linguistic tone is a determinant of all melodies sung in Yorùbá cannot be verified, this study claims that tone is a determinant of Fúji melody.

In 1984, Ọlátunji outlined features of Yorùbá oral poetry. Fúji, a Yorùbá vocal music, exhibits similar features, including parallel structures, exact repetition, and a feature Ọlátunji calls tonal counterpoint, wherein a sequence is repeated with the same phonic content but with poetic changes in tone. Such wordplay does not alter lexical meaning in words where tone is non-contrastive. Alternately, tonal counterpoint is used to create ironic semantic contrasts in words where tone provides phonological contrast.

Fúji artists in particular use deep or conc Yorùbá, an abbreviation for concentrated. These considerations are present in analysis of three examples of voice-only introductions to Fúji recordings. The method includes detailed transcription with tone, translation, and computer-assisted analysis. These perspectives combine in the first formal analysis (to my knowledge) of Fúji, the most popular form of Yorùbá-language music since the 1990s (Waterman 2002).
14.3.1 RECORDINGS AND PROCESSING

Compact discs were purchased at Alaba and Yaba markets in Lagos state, Nigeria. A vendor at each location was asked for 10 Fúji and 10 Jùjú CDs from his current inventory. These were gathered for another study of genre identification and listener preference not addressed here, but with the intention of a secondary use for analysis. Unaccompanied vocal sections, appropriate for voice analysis, were culled from the Fúji discs. Three track introductions by three notable Fúji artists were selected: Barrister, K-One, and Saheed Oṣupa. These artists come from the first, second and third generation of Fúji artists, respectively. From these short excerpts, ranging from 14 to 31 seconds in length, transcriptions were made by University of Lagos undergraduate Lanre Ṣefunye, and verified by doctoral student Tola Ṣunnuga. I also checked them using Abrahams’ Dictionary of Modern Yorùbá (1962) and the University of Ibadan Yoruba Dictionary (1990). It is important to note that, much like Yorùbá oral poetry, Fúji uses poetic language and not conversational or conventional language. Similar to analysis in other chapters, the recordings were segmented in Melodyne according to the transcription and then exported for f0 and contour analysis in MATLAB.

14.3.2 FIRST GENERATION ARTIST: BARRISTER

Along with Alhaji Professor Ayinla Kollington, Alhaji Sikiru Ayinde Barrister are the two superstars of the first generation of Fúji artists (Barber and Waterman 1995:244). It is difficult to find credible biographical information on Barrister. Wikipedia has an entry for Barrister and cites Barber and Waterman (1995) for his basic biographical information including a birth year of 1948. However, that information is not included in
their chapter, which is one of very few scholarly writings that present primary research on Fújí. What I can verify from interviews is that Barrister died in 2010 and every Fújí singer my colleague David Àíná and I talked to mentioned Barrister as a primary influence. Barber and Waterman (1995) report that Barrister named his music Fújí after Mount Fuji based on seeing a poster in a travel bureau office. This story was also recounted by Saheed Ôṣùpá whom we interviewed in February of 2014.

<table>
<thead>
<tr>
<th>Artist</th>
<th>Fuji Commander Alhaji Sikiru Ayinde Barrister and his golden Fuji exponent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Album</td>
<td>Orisa Bi Iya Osi</td>
</tr>
<tr>
<td>Track</td>
<td>1. Ile Aiye Odun Pupo</td>
</tr>
<tr>
<td>Time</td>
<td>00:00 to 00:18</td>
</tr>
</tbody>
</table>

Table 14.2: Track Information for First Generation Artist

The track information in Table 14.2 does not include the date of recording or release because that information is not included on the sleeve.

![Image of album cover](image-url)
In general, CD sleeves in Nigeria include contact information for a marketer, but not much else. Copyright information is not included and piracy is rampant.

<table>
<thead>
<tr>
<th>Phr.</th>
<th>Yorùbá</th>
<th>Tones</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Oọrì mí màa jèn flì</td>
<td>MHHHMLL</td>
<td>My head (destiny) should not let me leave</td>
</tr>
<tr>
<td>2</td>
<td>ìyá mí s'áye lò.</td>
<td>LHMHML</td>
<td>my mother on earth.</td>
</tr>
<tr>
<td>3</td>
<td>ọm'abí ní w'ayé</td>
<td>MMHHMH</td>
<td>A child given on earth.</td>
</tr>
<tr>
<td>4</td>
<td>èdá mí màa jèn flì</td>
<td>LHHHMLL</td>
<td>My creator should not let me leave</td>
</tr>
<tr>
<td>5</td>
<td>ìyá mí s'áye lò</td>
<td>LHMHML</td>
<td>my mother on earth.</td>
</tr>
<tr>
<td>6</td>
<td>Ilé ayé ladún (pàpò jù)</td>
<td>MHMHML</td>
<td>This world is full of joy</td>
</tr>
</tbody>
</table>

Figure 14.3: Barrister

The lyrics are vaguely proverbial, but does not include specific proverbs that are usually associated with a distinct underlying tone that is realized clearly. The most notable aspect of the Barrister excerpt is that pitch may be largely flat for an entire phrase (see Figure 14.4). Phrase 5 spans four semitones while phrases the excerpts by K1 and Òṣùpá spanned 8–12 semitones. The monotone strikes one as tonally underspecified and uncharacteristic of Yorùbá speech, poetry and song.

Figure 14.4: Phrase 5

The observation of narrow range and underspecification in Phrase 5 may be generalized to the entire excerpt.
Mid and Low tones overlap significantly. High tones are more variable, which is often the case, in speech but not necessarily in song.

Table 14.3: Barrister

<table>
<thead>
<tr>
<th>Tone Level</th>
<th>Mean</th>
<th>St. Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>63.9</td>
<td>2.6</td>
</tr>
<tr>
<td>Mid</td>
<td>61.5</td>
<td>1.4</td>
</tr>
<tr>
<td>Low</td>
<td>61.0</td>
<td>1.8</td>
</tr>
</tbody>
</table>

Table 14.4: Comparison of CL2 (+1/-1) values for corresponding tones and realized pitch (1 = match, 0 = different)

Table 14.4 follows the format of Table 13.10. Contour levels based on 2–degree windows are already quite forgiving, so the contour similarity between the underlying tone and realized pitch is very low in all but the third phrase.

14.3.3 SECOND GENERATION ARTIST: K1

K1 shares similar appellations to the last artist. Colloquially, “Barrister” refers to the last artist and “Wasiu” or “K1” refers to this artist. Notice in Figure 14.5 that K1’s band is the “fuji commander” not the “GOLDEN FUJI EXPONENT” and that Wasiu precedes “Ayinde Barrister” whereas Sikiru preceded “Ayinde Barrister” for the previous artist.
They are differentiated biographically by their era of popularity, K1 is known as a second generation artist, popular from the 1990s and still active as an elder artist.

Despite some attempts, we were not able to interview K1 and verify our transcription of his lyric as we were able to do with the next artist, Saheed Ôṣùpá.

<table>
<thead>
<tr>
<th>Artist</th>
<th>Alhaji (Chief) Wasiu Ayinde Barrister &amp; his fuji commander (K1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Album</td>
<td>Isẹ́ Ògún Isẹ́</td>
</tr>
<tr>
<td>Track</td>
<td>1. Ire ni ìtèmi</td>
</tr>
<tr>
<td>Time</td>
<td>00:00 to 00:14</td>
</tr>
</tbody>
</table>

K1’s introduction is characterized by two instances of tonal counterpoint, in Phrases 1–2 and 3–4, accounting for four out of five phrases.
In phrases 1–2, kúbúrè (HHH) is mapped to kúbùrè (HLL), and in phrases 3–4 èfó (LH) is mapped to èfó (LL). Because there are not distinct homophones, the meaning is not ambiguous. Once the lexical tone is clearly stated in the first phrase of the pair, it may be modified as a “tonal wordplay” (Olátunji 1984:37).

![Figure 14.8: Tonal counterpoint in Phrases 1–2](image)

The greater dispersion of mean pitch for equivalent tones suggests a clearer realization of underlying tone.

<table>
<thead>
<tr>
<th>Phr.</th>
<th>Yorùbá</th>
<th>Tones</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Isán sán mó-o-ôn kúbúrè</td>
<td>LHHHMLHHH</td>
<td>The great harvester</td>
</tr>
<tr>
<td>2</td>
<td>sán sán mó-o-ôn kúbúrè</td>
<td>HHHMLLLL</td>
<td>The great harvester</td>
</tr>
<tr>
<td>3</td>
<td>Ai- gbin alubòsà kó wèfò</td>
<td>LMMHLHLH</td>
<td>We cannot plant onion and reap (green) vegetable</td>
</tr>
<tr>
<td>4</td>
<td>Ai- gbin alubòsà kó wèfò</td>
<td>LMMHLHLL</td>
<td>We cannot plant onion and reap (green) vegetable</td>
</tr>
<tr>
<td>5</td>
<td>O’un abá gbin lòòwù bó dòlòla</td>
<td>MMHHLMLHLL</td>
<td>Whatever you plant is what will grow when tomorrow comes</td>
</tr>
</tbody>
</table>

In Figure 14.7: K1

<table>
<thead>
<tr>
<th>Tone Level</th>
<th>Mean</th>
<th>St. Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>64.4</td>
<td>2.3</td>
</tr>
<tr>
<td>Mid</td>
<td>62.4</td>
<td>2.2</td>
</tr>
<tr>
<td>Low</td>
<td>60.4</td>
<td>1.8</td>
</tr>
</tbody>
</table>

Table 14.5: K1
However, contour level similarity (at 2 degrees) by segment and by phrase reveals that there are many discrepancies between the underlying and realized tone.

<table>
<thead>
<tr>
<th>Phr.</th>
<th>n</th>
<th>Initial Tones</th>
<th>Final Tones</th>
<th>μ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1 2 3 4 5</td>
<td>5 4 3 2 1</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>9</td>
<td>1 1 0 1 1</td>
<td>1 1 1 0 0</td>
<td>0.667</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>1 0 0 1 1</td>
<td>1 1 1 1 1</td>
<td>0.750</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>1 0 0 1 1</td>
<td>1 1 1 1 1</td>
<td>0.778</td>
</tr>
<tr>
<td>4</td>
<td>9</td>
<td>1 0 1 0 1</td>
<td>1 1 1 0 1</td>
<td>0.667</td>
</tr>
<tr>
<td>5</td>
<td>12</td>
<td>0 1 1 1 1</td>
<td>0 1 0 0 1</td>
<td>0.667</td>
</tr>
<tr>
<td>μ</td>
<td>9.4</td>
<td>0.8 0.4 0.4</td>
<td>0.8 1 0.8 0.4 0.8</td>
<td>0.706</td>
</tr>
</tbody>
</table>

Table 14.6: Comparison of CL2 (+1/-1) values for corresponding tones and realized pitch (1 = match, 0 = different)

14.3.4 Third generation artist: Òṣùpá

Saheed Òṣùpá is one of the leading third generation Fúji artists, along with Pasuma, Malaika and Ìfẹ̀ Alao. These artists have gained popularity since 2000.
“Tribute to my father” is from his 2009 album *Fuji Icon*. The albums is divided into two parts (A and B), like an LP or cassette, though it is only available on CD. Fuji Icon “B” follows a narrative trajectory from his father dying to celebrating his mother to acknowledging the celebrities (*gbajumon*) that came to his father’s wake. In our February 2014

<table>
<thead>
<tr>
<th>Artist</th>
<th>King Dr. Saheed Osupa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Album</td>
<td>Fuji Icon</td>
</tr>
<tr>
<td>Track</td>
<td>3. Tribute to my father</td>
</tr>
<tr>
<td>Time</td>
<td>00:00 to 00:31</td>
</tr>
</tbody>
</table>

Figure 14.10: Track Information for Third Generation Artist

The excerpt analyzed introduces a textual theme that runs throughout Fuji Icon “B” and that is close to oriki praise culture: people become words, their good and bad deeds is how they are remembered.
The first phrase sets Òṣùpá’s vocal style and tone realization apart from Barrister and K1 immediately. Although it is a four-semitone range, the pitch of each segment is clearly delimited and corresponds closely to the underlying tone. The relatively stable pitch of each segment may reflect broader trends in Nigerian commercial music, particularly the prevalent use of autotune. Òṣùpá does not use autotune, but the strong presence of it in Naija Hip-Hop and dance music since 2005 may explain why he makes use of straight tone instead of emulating the less stable pitch of Barrister or K1.
This first phrase could be frequency-banded. This is not the case for all phrases, but the dispersion of mean pitch for each tone level and standard deviations are more similar to the Ìjálá measurements than the other Fújì artists (see Chapter 13).

<table>
<thead>
<tr>
<th>Tone Level</th>
<th>Mean</th>
<th>St. Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>57.1</td>
<td>2.3</td>
</tr>
<tr>
<td>Mid</td>
<td>55.7</td>
<td>2.2</td>
</tr>
<tr>
<td>Low</td>
<td>53.3</td>
<td>2.6</td>
</tr>
</tbody>
</table>

Table 14.7: Òṣùpá

Half of the phrases (7 of 14) have contour level equivalence with a 2–degree window.

<table>
<thead>
<tr>
<th>Phr.</th>
<th>n</th>
<th>Initial Tones</th>
<th>Final Tones</th>
<th>μ</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6</td>
<td>1 1 1 1 1 1</td>
<td>1 1 1 1 1 1</td>
<td>1.000</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>1 1 1 1 1 0</td>
<td>1 1 1 1 0 0</td>
<td>0.667</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>1 1 1 1 1 1</td>
<td>1 1 1 1 1 1</td>
<td>1.000</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>1 1 1 1 1 1</td>
<td>1 1 1 1 1 1</td>
<td>1.000</td>
</tr>
<tr>
<td>5</td>
<td>9</td>
<td>1 1 1 1 1 1</td>
<td>1 1 1 1 1 1</td>
<td>1.000</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>1 1 1 1 1 0</td>
<td>1 0 1 0 1 1</td>
<td>0.667</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>1 1 1 0 1 1</td>
<td>1 0 1 1 1 1</td>
<td>0.833</td>
</tr>
<tr>
<td>8</td>
<td>6</td>
<td>1 1 1 1 1 1</td>
<td>1 1 1 1 1 1</td>
<td>1.000</td>
</tr>
<tr>
<td>9</td>
<td>7</td>
<td>1 1 1 1 0 0</td>
<td>1 1 1 0 1 1</td>
<td>0.714</td>
</tr>
<tr>
<td>10</td>
<td>6</td>
<td>1 1 1 1 1 0</td>
<td>1 1 1 0 0 0</td>
<td>0.667</td>
</tr>
<tr>
<td>11</td>
<td>7</td>
<td>1 1 1 1 1 1</td>
<td>1 1 1 1 1 1</td>
<td>1.000</td>
</tr>
<tr>
<td>12</td>
<td>9</td>
<td>1 1 0 1 1 0</td>
<td>1 1 1 0 1 1</td>
<td>0.778</td>
</tr>
<tr>
<td>13</td>
<td>9</td>
<td>1 1 0 1 1 0</td>
<td>1 1 1 0 1 1</td>
<td>0.778</td>
</tr>
<tr>
<td>14</td>
<td>6</td>
<td>1 1 1 0 1 0</td>
<td>1 0 1 0 1 1</td>
<td>0.667</td>
</tr>
<tr>
<td>μ</td>
<td>6.5</td>
<td>1.000 0.929 0.714 1.000 0.643</td>
<td>1.000 0.786 0.929 0.643 0.857 0.857</td>
<td></td>
</tr>
</tbody>
</table>

Table 14.8: Comparison of CL2 (+1/–1) values for corresponding tones and realized pitch (1 = match, 0 = different)

Phrases 3 and 11 are presentations of the same lyric using the same melody. They have contour level equivalence to the underlying tone as well as to each other despite some declination. They may be considered pitch equivalent.
Phrases 6 and 14 are not equivalent to the tone sequence because of the ascent between consecutive mid tones (elements 2–3). Likewise they are pitch equivalent.

14.3.5 SUMMARY

The Fújì excerpts examined in this chapter exhibit a tone-tune correspondence, but not as strong as expected. The extent to which Fújì is more representative of the Yorùbá
language than other genres is yet to be determined and will be the subject of further study. Only the singing of Saheed Òṣùpá reveals a very strong correlation between the underlying tones and the sung realization. The presence of similar features to Yorùbá poetry in Fújì, such as tonal counterpoint, is an interest finding. Generally, tones did not appear to correspond to fixed pitches (i.e. high tone is not consistently realized as the pitch F#4) within most phrases or between phrases. Contrary to my expectations, the most recent artist exhibited the clearest mapping of tone to tune. At the same time the sustained and stable pitch of the singing suggests the influence of global popular music trends.
15. The Ìgbò Glees of Laz Èkwúèmè

Introduction

My first exposure to Nigerian choral music was in August of 2010 in Běijing (北京), China. The Musical Society of Nigeria (MUSON) Diploma Choir performed at the International Society for Music Education's Biennial Conference. Their performance was full of energy and the reception of the standing-room only audience was indefatigably enthusiastic.

Figure 15.1: MUSON Diploma Choir at the ISME Conference, Beijing, China, August 2010
The choir presented a pan-African program of choral compositions and arranged folk songs marking the fiftieth anniversary of the wave of post-colonial independence that swept across Africa (circa 1960). A piece that caught my attention was Laz E. N. Èkwúèmé’s *Obi Dimkpa* (*Brotherhood of Youth*), a composition in call-and-response chorus form. The solo was delivered by Òméká Nwókêdi, a former chorister in the Èkwúèmé’s Chorale and now director of the MUSON Diploma Choir. A video of the performance is available at this link: https://youtu.be/N7KvlHO-qwM.

15.1 About Laz Èkwúèmé

Born in present-day Ánámbrâ state in 1936 during British colonial rule of Nigeria, Laz Èkwúèmé earned his doctorate in music in 1972 from Yale University. He published a series of articles soon after, including “Linguistic Determinants of Some Îgbô Musical Properties” in 1974. Èkwúèmé approached West African music through the discipline of music theory. His writings have had a profound impact on scholars researching African music with analytical methods, including African-born music theorists Kofi Agawu and Akin Eubá.

With “Linguistic Determinants” (1974c), Èkwúèmé set a mandate for composing vocal music with tone languages, particularly his own language, Îgbô of southeastern Nigeria. He went on to compose many choral works in the decades since, applying his theories in practice. The secular Îgbô *Glees*, “Eli Meli,” “Obi Dimkpa,” and “Ote Nkwu,” are unusual in the largely sacred choral repertoire of Nigeria (Ọmọjola 1994:534–35) and are popular among school choirs. They have great potential for teaching students about
Ìgbò culture abroad and are a synthesis of Òkwúèmé's disparate experiences and influences. These works juxtapose Òkwúèmé's enduring connection to his Ìgbò village Ókò, where he is now Ígwé (traditional ruler), and his education abroad. At Yale University, he was a member of and at times conducted the Yale Glee Club. Òkwúèmé returned to Nigeria to join the faculty of the University of Lagos in 1976, and his musical palette continued to include the glee club sound alongside Ìgbò and Nigerian music.

This chapter focuses on the realization of “linguistic determinants” in Òkwúèmé’s “Obi Dimkpa (Brotherhood of Youth)” composed during a cultural diplomacy visit to North Korea in 1980. Òkwúèmé has been a musical representative for Nigeria on many occasions. He was awarded a Fulbright fellowship to study in the United States, he was Nigeria’s first professor of music and was recently elected to the National Order of Merit. While staying in Pyong Yang, Òkwúèmé wrote “Obi Dimkpa” and it became one of his most often performed works. Òkwúèmé included two texts for the work, an Ìgbò lyric and an English lyric. If revealing of his mindset while in North Korea, two very different interpretations can be made depending on which text you read. The Ìgbò text indicates Òkwúèmé was finding strength and courage in the Ìgbò proverbs of his youth, while the English text indicates he was inspired by the collective spirit of the communist state: “Stretch out a right hand of brotherhood!”

15.2 Textual Analysis

The texts of Obi Dimkpa are by the composer, including an Ìgbò text typically performed in Nigeria and an English text. The English text is not a translation of the Ìgbò text. The
two texts address manhood, but they are contrasting perspectives reflecting different cultures.

The Ìgbò text is full of proverbs and life-lessons emphasizing courage and independence. The opening line, “Wèlú óbì dìmìkpà kpàghálìbá,” translates to English as “Move about with a brave heart.” The corresponding English lyric is “Stretch out a right hand of brotherhood!” An examination of a literal English translation (by Quintina Carter-Ényì) and the English lyrics written by Èkwúémé confirms the English text is not a translation.

The cumulative effect of the Ìgbò text is a complete picture of manhood from the rural Ìgbò perspective.

Move about with a brave heart,
“Kwango, kwango” goes the drumming of food and soup.
Any person who is overtaken hangs up his uniform.
Unless you wrestle, the backyard path will never close.
The snake pursues the lizard only to swallow it.
If one believes, then God will agree.
If a small boiling pot is ignored, it boils over and quenches the fire.
Move about with a brave heart.

The only line of the Ìgbò lyrics similar to a line in the English lyrics is:

Ónyé kwé, nà Chí yá èkwélú.
(Translation:) If one believes, then God will agree.
(English lyric:) With a firm faith so much can be done.

The English lyric version of the line is secularized, representing faith in the state as well as faith in God. Overall, the English text conveys a different sort of manhood. The wisdom of the Ìgbò proverbs is lost, conveying a Quixotic view of the world:

---

80 Lines 2–4 are the refrain, lines 5–8 are verses 2–5.
81 As Prof. A.V.E. Meréni points out, the Ìgbò proverb is similar to the English proverb “Heaven helps those who help themselves.”
Stretch out a right hand of brotherhood!
Caring, sharing, loving one another:
Making every one else your own true brother;
Malice towards none, good as every youth should,
Loving as nestlings of a single brood;
With a firm faith so much can be done;
A boiling pot can bring so much fun;
Stretch out a right hand of brotherhood!

A few of the Ìgbò proverbs are antithetical to the corresponding English text. Consider this line from the refrain:

Ágbà rò úzò òwèlè mbá ná ó dà èchí nú.
(Translation:) Unless you wrestle, the backyard path will never close.82
(English lyric:) Malice towards none, good as every youth should.

And Verse 2:
Ágwó nà àchú ñwèlè màkà ònúnó.
(Translation:) The snake pursues the lizard only to swallow it.
(English lyric:) Loving as nestlings of a single brood.

Both texts emphasize manhood, one in Ìgbò culture (self-reliance), the other in a communist society (belonging). Èkwúèmé did not create a metric translation of the Ìgbò text into English. Why? It is nearly impossible to produce an accurate line-for-line translation that preserves the number of syllables. However, even the composite meaning, which could be broadly transferred in a translation, seems opposed in the texts. Perhaps he doubted the receptivity of outsiders to Ìgbò proverbs. Or, maybe he was reluctant to produce a piece on a diplomatic trip in which choristers sing: “the snake pursues the lizard only to swallow it.” When I asked Èkwúèmé about the incongruence of the texts in 2014, he laughed but made no comment. Camaraderie is not so foreign to Ìgbò culture.

Once one has left the village, people of one’s village or ñdí-òbòdò become brethren and

82 Further explanation: If one wants the backyard to close with high grass (for privacy), he will need to wrestle with those who follow the path through the backyard. The path through the backyard will never close as long as people are passing there.
one is less concerned with high grass growing over the backyard path. The change of sentiment between the Ìgbò and English lyrics may represent a dichotomy between Nigerian and North Korean culture. It also parallels the transformation in rural-to-urban migration for Ìgbò people in Nigeria, which is also a linguistic migration.

15.3 Musical Analysis

While the English text is appropriate to a diplomatic trip, the music and the Ìgbò text are less related to the place of composition. The music is a unique marriage of the sounds of Ìgbò culture with North American glee-club harmony. There are manifold connections between all manner of sounds in Ìgbò culture, including language, music and the broader aural environment. Èkwùémé engages with multivalent sound through tone-to-tone mapping and onomatopoeic vocables.

![Figure 15.2: Sound relationships in Èkwùémé’s Glee](image-url)
15.3.1 VOCABLES

In Niger-Congo languages, many utterances in speech and singing originate in the aural environment (onomatopoeia) or are external expressions of internal non-verbal ideas and emotions (ideophones) (Mphande 1992:127). In Ìgbò, many of the names for instruments are vocal imitations of the sound they produce. For example, the Ìdù is a large pot that is gently struck at the opening with a pad to create a low, resonant release of air pressure. Preliminary research on vocal mimesis reveals a few cross-cultural trends:

1. Voiced stops [b, d, g] are associated with low resonance and voiceless stops [p, t, k] are associated with high resonance in drums (membranophones);
2. Fricatives [f, s, ʃ] represent shakers, rattles, cymbals, etc. (idiophones);
3. Stop-fricative sequences [tʃ, ts, pf] represent combinations like drum and rattle. Many of these correspond to descriptions in Stowell’s “Beatbox Alphabet” (2012). Phonological constraints account for variation between languages. In Yorùbá, the gourd-shaker is Ọkèrè, while in Ìgbò it is Ìchàkà. Both cognates have fricatives, low tones and vowel harmony, but the fricative [ʃ] is common in Yorùbá not Ìgbò, and the stop-fricative [tʃ] is common in Ìgbò not Yorùbá.

In Èkwùèmé’s Glees, sounds of the aural environment become text and thematic material. This is most prominent in “Ote Nkwu” (Sweet Palm Wine) from 1992, where the sound of the palm-wine tapper’s tool becomes the primary lyrical and musical motive of the piece:
In “Obi Dimkpa” the preparing of food becomes lyrical expression:

![Image of woman pounding yam]

Figure 15.4: “Kwango, kwango” (the sound of pounding yam)

15.3.2 **Text Setting**

The most crucial exchange between categories of sound in Èkwélemé’s art is the preservation of tonemes in language and music. In “Linguistic determinants of some Ìgbò musical properties” Èkwélemé states that the “melodic line must be controlled by the contour (tonal rise and fall) of the words” (1974c:337). The Ìgbò language has three
tonemes: high (’ )\textsuperscript{83}, step (¯) and low (‘ ). Step tone is a slightly lower pitch from the previous tone and can only follow high tone, not low tone (Williamson 1972:x). Thus, there are five possible disyllable tone combinations in the Ígbò language.

<table>
<thead>
<tr>
<th>Ígbò</th>
<th>Tone</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>ákwá</td>
<td>HH</td>
<td>cry</td>
</tr>
<tr>
<td>ákwá</td>
<td>HS</td>
<td>cry</td>
</tr>
<tr>
<td>ákwá</td>
<td>HL</td>
<td>cloth</td>
</tr>
<tr>
<td>ákwá</td>
<td>LH</td>
<td>egg</td>
</tr>
<tr>
<td>ákwá</td>
<td>LL</td>
<td>bed</td>
</tr>
</tbody>
</table>

Table 15.1: /akwa/ homophone group

According to Èkwúèmé, there is generally one relatively flat (not sloped) tone per syllable in Ígbò speech, so each syllable is sung on one pitch. Melismas are used occasionally, but not when lexical tone is present in a word. Usage of melismas is “almost invariably on exclamations” (1974c:343). Èkwúèmé also notes the rhythms of Ígbò speech and proposes that text setting follow this as well. He notes that “often the rhythm is dependent on the tone” (1974c:346). However, I have not found other evidence for this position and there seems to be no correspondence between specific tone levels and note duration in the piece.

15.3.3 STROPHISM

In improvised singing, when a “solo lead sings several verses in a song... he adapts the melodic line of his music to accommodate the new words of every verse” (1974c:349).

\textsuperscript{83} In some orthographies (including Williamson 1972), high-tone is not marked.
Because of the “linguistic determinants,” vocal music in tone languages cannot be strictly strophic across verses, even if the verses have the same number of syllables. For example, Verse 1 and Verse 3 of “Obi Dimkpa” both have nine syllables, but the two texts cannot be set to the same melodic contour.

If the text of verse 1 is placed on the melody of verse 3, or vice-versa, the meaning may not be completely lost, but the words are garbled, the diction less intelligible. The effect would be reminiscent of the tone-tune mismatch addressed in Chapter 1.

The melodies of verse 1 and verse 3 are strophes because of their analogous locations in the piece, however the melodic similarity is quite low. Strophic correspondence is more keenly observed in the soprano part of verse 1 and verse 2, which have a much clearer modified strophic relationship.
In contrast to Figure 15.5, in Figure 15.6 the ordering of pitch content is nearly identical, verse 2 is simply expanded. In both, the soprano outlines: D to Bb, C to G, and G to Bb. Modified strophes accommodate the differences in lexical tone between the verses, contours vary while the harmonic underpinning remains the same.

The second and fourth verses are of the same cardinality, and their contour similarity is quite apparent. Comparing them using C+ matrices and C+SIM yields a rather high value of 80% similarity.
15.3.4 SIMILARITY BETWEEN CALLS AND RESPONSES

Call and response between soloist and ensemble is common in both Ìgbò music and university glee-club singing, and as such is perfect for a piece that integrates the two. In Ìgbòland, male praise singing groups use this texture and are accompanied by a variety of percussion, including ògéné (metal gong), ichàkà (gourd shaker) and ùdù (metal pot).

Female groups have a similar make-up, but mixed voice ensembles are a more recent phenomenon associated with Christianity (Mereni 2004a:16). “Obi Dimkpa” was originally composed for male voices (TTBB), but four-part SATB choirs are now more common and I have typically heard the piece sung in this arrangement (as in the MUSON Diploma Choir performance linked above).
Each verse of “Obi Dimkpa” is a different Igbo adage. The text of each verse is introduced in the Baritone Solo part and then repeated in response by the choir.

<table>
<thead>
<tr>
<th>Verses:</th>
<th>1 (n=9)</th>
<th>2 (n=10)</th>
<th>4 (n=10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soprano</td>
<td>1</td>
<td>1</td>
<td>.990</td>
</tr>
<tr>
<td>Alto</td>
<td>1</td>
<td>1</td>
<td>.990</td>
</tr>
<tr>
<td>Tenor</td>
<td>.988</td>
<td>1</td>
<td>.990</td>
</tr>
</tbody>
</table>

Table 15.2: C+SIM between Call (Baritone) and Response by Voice Part

Table 15.2 includes C+SIM values between the call and each voice part of the response for verses 1, 2 and 4. All are identical or nearly identical in terms of contour. The third verse is not included because it does not have the same cardinality in all voices.

15.3.5 **Obi Dimkpa: Verse 1**

The first verse of the piece means “Move about with a brave heart.” In Figure 15.8, the tone levels of the text are indicated with the colored noteheads, yellow for low and blue for high.
An aggregate matrix of the setting in all four voice parts includes one fuzzy value because the tenor changes notes from “-kpa” to “kpa-.” A single CSEG would represent all four iterations of the text without this discrepancy.

The pitch changes from Bb3 to C4 while the underlying tone remains the same (L). Is this a violation of lexical tone? Probably not. It is a change of one step and all changes of tone...
level are a melodic interval of a third or fourth in the excerpt, so there is no contradiction between stepwise motion and static tone. However, it does mean that the four settings of the text do not have full CSEG equivalence. If the piece were put into the segmentation algorithm, it would not recognize settings of verse 1 as a single recursive, but as a compound segment, indicated by the bold boxes in Figure 15.10, extracted and reoriented in Figure 15.11.

Within each segment in the compound, the slices can be summed into a contour level series (though in fact these are contour pitches because everything is compared to everything else in the segment). These values then need to consecutized because there are recurring pitches. To extract the phonological tone levels, declination and downstep needs to be accounted for. The second segment within the compound is not effected by the consecutization or the intonation adjustment.
15.3.6 *Obi Dimkpa*: Verse 2

The second verse is a colorful one: “The snake pursues the lizard only to swallow it.”

This verse includes an ossia for the solo to be used in repetition of the verse (the text is above for the first time, though the notes are stem down).

In this case, a single recursive segment describes the entire text setting in all verses. The consecutization and intonation adjustment handily produce the tone levels.
This verse does present an ambiguity however. In verse 2, a step up is a change in tone level in all voice where as a step up was static tone in the tenor of verse 1.

15.3.7 OBI DINKPA: VERSE 4

In verse 4, the ambiguity between verse 1 and verse 2 in terms of the tone-level signification of a step up appears in the proximity of a measure. In the solo voice, F3 to G3 is L to H on beat 2 and on beat 4 D3 to Eb3 is state tone (L to L). The latter is a m2, but one wonders whether the distinction matters in diatonic space.
The study results in Chapter 5 indicated that a rise of 2 semitones was necessary to be a clear contrast in speech, but in singing, the distance may be smaller. Whether a step is representative of a tone level change is not consistent among composers. Verse 4 also includes a step tone (indicated in green).

Like the first verse, the fourth verse has a slightly different contour present, but this time it is in the solo. Because of the fuzzy value, it is a compound of recursive segments. The
second segment includes the step tone, so the end contour level series has three levels, 2, 1 and 0, for high, step and low.

15.3.8 HARMONY, TEXTURE AND FORM

The high C+ SIM (at n-1 degrees) between the four settings of each verse to the four voice parts indicates a strong presence of semblant (including similar and parallel) motion between contours that are vertically aligned. Table 15.3 has measures of similarity between pairs of parts for all verses and the refrain.

<table>
<thead>
<tr>
<th></th>
<th>Soprano</th>
<th>Alto</th>
<th>Tenor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alto</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tenor</td>
<td>.974</td>
<td>.974</td>
<td></td>
</tr>
<tr>
<td>Bass</td>
<td>1</td>
<td>1</td>
<td>.974</td>
</tr>
</tbody>
</table>

Table 15.3: CAS Similarity (phrase-bounded and only where there are complete verticalities, n=39)

A value of 1 indicates all harmonic intervals formed by two voices move in the same direction. In Table 15.3, Soprano, Alto and Bass comparisons have identical C+AS, the tenor deviates slightly from the others. This is not motivated by the verticalities themselves, but because “melodic shape has to be retained by any parts (singing the same words)” (1974:344).

Parallel harmony… seems to be a regular African system for the harmonization of tunes. The larger intervals—octaves, fifths, and fourths—tend to be favored, although quite a number of African peoples, including the Igbo, employ thirds and sixths either in addition to, or instead of, the perfect intervals. (1974:344)

Èkwúémé faces a challenge in reconciling his diverse influences. Formally composed choral music in the west tends to avoid parallel harmony so that there is independence
between voices, however in this case, there would also be independence in the text the choristers are singing. Glee-club and barbershop-style harmony is sometimes less formal, but is not predominantly parallel in three out of four voices. In Glee-Club singing, the Tenor 2 often has the melody, tenor 1 and bass 1 often have oblique motion and bass 2 is governed strongly by the harmonic changes. In English, there is no constraint against the Bass 2 voice singing in contrary motion on the same text, contrasting the speech contour.

Recall “Obi Dimkpa” was composed initially for male voices, let us consider that version to have the original vertical ordering of parts. In the male voice version, harmonic intervals between vertically-adjacent voices are almost exclusively m3, M3, P4. M3 is the most common harmonic interval, accounting for 44.6% of the harmonic events between Tenor 1 and Tenor 2. The average of the vertical pitch-distance between adjacent voices is a M3, 4.188 semitones for Tenor 1 to Tenor 2 and 3.877 semitones for Tenor 2 to Bass 1.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Deviation</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tenor 1 and Tenor 2</td>
<td>4.188</td>
<td>1.345</td>
<td>4 (M3)</td>
</tr>
<tr>
<td>Tenor 2 and Bass 1</td>
<td>3.877</td>
<td>1.973</td>
<td>4 (M3)</td>
</tr>
</tbody>
</table>

Table 15.4: Statistical analysis of harmonic intervals

There are a few notable exceptions to the predominance of 3rds and 4ths, including measure 9 when the Tenor 2 voice crosses below the Bass 1 (Baritone) voice, creating a m7 and M6 against the Tenor 1 voice. In mm. 5–6, the Bass 2 voice leaps from F3 to Bb3 instead of stepping to the adjacent tone of G3 (see Figure 15.17), producing a same direction movement to a unison with the bass 1. This avoids a parallel fifth between the tenor 1 (top) and bass 2 (bottom).
Èkwúèmé does not freely compose the subordinate voices with only speech-tones in mind, he incorporates rules of western voice-leading, including avoiding parallel fifths. In addition, Èkwúèmé makes uses of, vocables and staggering of text, which make contrary and oblique motion permissible.

Three devices that provide alternative to parallel motion are highlighted in Figure 15.17: similar motion to avoid parallel fifths (orange boxes), vocables to allow free movement.
with in a voice part (blue box) and staggering text through use of elision also to allow one
voice to move independently (green box). Comparing Figure 15.16 with the first two
measures in Figure 15.17 indicates the alto voice of the SATB version has what was
formerly the Bass 1 part, but the Soprano and Tenor have a mix of horizontal lines from
the TTBB version.

As proposed in “Linguistic Determinants,” the nature of the Ìgbò language has a
profound impact on the compositional process and product. There are also outside
influences at work. *Obi Dimkpa* is termed an Ìgbò Glee for a reason. Èkwúèmé’s time at
Yale included active involvement with the University’s famous Glee Club. The glee-club
style is reflected in the texture and the form of *Obi Dimkpa*. A male vocal ensemble
singing in unison or harmony in response to a soloist’s call is not foreign (Mereni
2004:3), but the precise texture is derived from the glee-club style. Like the North
American glees, the melody is found in the Tenor 2 voice in the TTBB version of “Obi
Dimkpa”, a common feature of glee-club and barbershop singing and is characteristic of
the sound.

The formal structure is similar to a perennial favorite of glee clubs, “Viva
L’Amour”, in which brief call-and-response verses are juxtaposed with a lengthy refrain.
In the male praise-singing ensembles of contemporary Ìgbòland, the leader typically
improvises extensively, and if the praise is well-received, the leader’s call will be
extended. In *Obi Dimkpa*, each four-measure verse is followed by a choral refrain of
eight measures. The form is quite compact, true to glee-club style, lasting 2–3 minutes
depending on the tempo. This is in contrast to Ìgbò folk-singing which is longer and more
flexible formally, adaptable to the occasion. The concise piece is capped off by a memorable *allargando* ending, precipitated by a caesura (sudden pause).

The glee-club character of the piece is strongest in this coda. In the last measure, the Bass voice splits with the upper portion creating a passing chromatic dissonance: G-Gb-F (with an intervening C to accommodate the tone contour of the text). It is this moment when the signature sound of glee clubs, that had so far been suppressed, is revealed: a dash of saccarin-sweet chromaticism.

15.3.9 SUMMARY

“Obi Dimkpa” is a very successful piece, enjoyable for performers and audiences, but it is also a clear manifestation of the principles laid out in Èkwúèmé's article “Linguistic Determinants” (1974c). This can be quantified by calculating similarity (the number of
matching) values between the directions between the underlying tone of the text and the
directions of the melodic intervals in each part.\textsuperscript{84}

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|}
\hline
Soprano & Alto & Tenor & Bass \\
\hline
Tone & 1 & 1 & .974 & 1 \\
\hline
\end{tabular}
\caption{CAS Similarity (phrase-bounded and only where there are complete verticalities, n=39)}
\end{table}

Values at phrase-boundaries are excluded because of declination and pitch reset, which is
ehibited in his setting. Declination is reflected in the phrases Èkwúémé composed for
each Ìgbò adage.

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|}
\hline
Verse 1 & Verse 2 & Verse 3 & Verse 4 \\
\hline
Initial tone & Low & High & High & High \\
Initial pitch & F3 & Bb3 & Bb3 & Bb3 \\
Final tone & High & High & High & Step \\
Final pitch & F3 & F3 & F3 & F3 \\
\hline
\end{tabular}
\caption{Phrase-initial and phrase-final tone and pitches (in Baritone Solo)}
\end{table}

According to the data in Table 15.6, tonemes at the beginning of two-measure musical
phrases are realized at higher pitch than tonemes at the end of the phrase. For example, in
verse 2 the phrase-initial high toneme is realized with the musical pitch Bb3 and the
phrase-final toneme is realized with the musical pitch F3. I also calculated a correlation
between the two-level tones and the melody, which returned $r=0.739$. If phrase-slope
were reflected in the written language, e.g. phrase-initial high tones were differentiated
from phrase-final high tones, then the correlation between tonemes and melodic pitch
would be greater. Tones are quite reductive of phonetic complexity, abstractions of the

\textsuperscript{84} Because of the antiphonal structure of the work, the melody is passed from the soloist to the Tenor 2
voice, thus the melodic data is a combination of the Solo and Tenor 2 voices.
necessary contrast to differentiate lexical items, and are not fully representative of the contours of speech. The system of marking speech-tones is simplistic in comparison with the detailed inventory developed for vowels and consonants. Èkwúèmé himself tells me that the proper pronunciation of his name is do-fa-do-re, even though the written tones are LHLH.

15.4 Conclusion

The potential of Èkwúèmé’s Ìgbò Glees for educating non-Ìgbòs about Ìgbò ethno-linguistic culture and disseminating Ìgbò language is immense. There is growing interest in linguistic diversity, particularly languages classified as Less Commonly Taught Languages (LCTLs), and music is an effective medium for learning about cultural diversity (Nketia 1970:48, Brown and Lamb 2004, Goetze 2010:21). Although the English text is intended to make the piece more approachable for non-Ìgbòs, it is through singing the Ìgbò text that the composers’ artistic vision is realized. In 2011, I taught the piece to secondary and tertiary students in Texas (see Figure 15.19).

85 Ìgbò is included in a list of 80 LCTLs published by the United States Dept. of Education and Title VI of the Civil Rights Act.
Based on that experience, I recognize why Èkwúèmé offers alternative texts for the glee for non-Ígbò singers. In Ote Nkwu (another popular piece in Nigeria) “Ta-kam chi-ki-li, chi-ki-li, chi-ki-li” may be substituted for “I-gbam chi-ki-li, chi-ki-li, chi-ki-li.” Specifically intended for non-Africans, this illustrates the phonetic challenges for singers new to Ígbò and other Niger-Congo languages. As Èkwúèmé indicates with the substitution of “takam” for “igbam,” coarticulated labiovelar consonants are among the most challenging sounds in Ígbò for non-speakers. The consonant stops “gb” and “kp” are articulated with the lips and the uvula at the same time, “gb” or /ɓ/ is the voiced and “kp” or /p/ is the voiceless (Williamson 1972:x). The challenge of learning this new sound is worth overcoming because it is enriching for the singers, students especially.

The essentialist stance of some ethnomusicologists might contest the “authenticity,” “indigeneity” or “traditional-ness” of the Ígbò Glees, and therefore their validity as bearers of African culture (see Kidula 2006:100). Let’s not dismiss the argument. “Obi Dimkpa” conforms to the Western diatonic scale and harmonic conventions, and it is recorded in Western notation. However, because it is literature, it is communicable to and replicable by literate musicians around the world. Through an
experimental choir at Indiana University, Goetze has shown it is possible to learn world
music through mimesis as in the Kodály method (2010:22), but this also conforms to
Western musical practice based in solmization (i.e. do, re, mi). At this time, there is no
standard method of teaching non-western styles outside of cultural immersion. In 1974,
Èkwúèmé wrote “we are yet to see a notational system which appears to be more
satisfactory overall in the representation of African music than a modified version of the
staff notation” (“Concepts of African Musical Theory” 1974b:46). Over forty years later,
this is still the case and in his 2015 keynote address to the Society for Music Theory,
Agawu urged that the staff is useful for transcription, especially when we are not tied to
its representation to equal temperament. As this article attests, Èkwúèmé’s Ìgbò Glees are
not indigenous Ìgbò music and should not be presented as such. Instead, these works are
Ìgbò art music. Formalized music literature that strongly possesses an ethnic identity is
still effective as a bearer of culture (Agawu 2001:8, Euba 2001:120). Èkwúèmé’s glees
contain Ìgbò proverbs and representations of the aural environment in Ìgbòland. Most of
all, they contain the features of Ìgbò language and traditional music practice as described
by Èkwúèmé in “Linguistic Determinants.” As Èkwúèmé has shown in his theory and
practice, these linguistic determinants do not always but perhaps should transcend
musical style, be it High-life, praise-singing, or choral music.
16. AESTHETICS IN NIGER-CONGO CULTURES

16.1 Mapping versus Aesthetic Influence

Broadly, there are two ways to look at music and language connections. The first is the cross-domain mapping of one to the other, like the setting of text to music. The setting of tone language texts to music is interesting because there is an objective mapping across domains, of implied speech contour to realized melody. Because the speech tones are contrastive and necessary for intelligible speech, we expect them to surface in the music. However, if we take a deterministic view of the mapping of tones to notes in music, we may be disappointed because the search may fail. There is often some tone-tune correspondence, but just like the realization of tone in speech, realization of tone in music is sometimes obfuscated.

In European vocal music there is often a representation of speech prosody, including intonation and timing, but it is less objective because these are not lexically contrastive features. Similarity between general features of language with features of music does not imply one-to-one correspondence as in a mapping. Instead, it is possible to look for shared characteristics in domains of pitch (tone), rhythm (timing), dynamics (stress), and timbre (phonic). This second type of music-language connection, I call it aesthetic influence, is a bit less frustrating to study because we are looking for less objective results. The Sapir-Whorf hypothesis suggests our perception and cognition is
shaped by our language and culture. How we hear language cannot be divorced from how we hear music and vice-versa.

A number of studies have addressed similar features in language and music within cultures. Specifically, speech timing and musical rhythm has been studied using nPVI (normalized pairwise variability index) in studies by Patel and Daniele (2003), Huron and Ollen (2003) and VanHandel (2006). Both Patel’s and Huron’s studies found that English speech and instrumental music has higher pairwise variability than French speech and instrumental music. The fact that instrumental music had similar features to the language of the composer was a novel discovery.

There are many possible directions for research on cross-domain aesthetic influence. By and large it is best not to make a strong interpretation of the Sapir-Whorf hypothesis. This is the problem with Diana Deutsch’s work on tone languages and absolute pitch. Instead of suggesting that there is higher pitch sensitivity among tone language speakers, she has posited that there is “perfect” pitch sensitivity among tone language speakers. While her findings are no doubt accurate, the conclusions she makes do not follow logically, particularly given counter evidence. Many linguists have demonstrated there is intonation in tone languages, I have presented even more evidence for that. Likewise, Udo Will and Catherine Ellis have shown that Australian Western Desert vocal music makes use of absolute pitch in the absence of tone languages and Western instruments (1994, 1996, 1997).

The better route is that of Patel and Daniele (2003), in which prosodic features in the language seem to nudge instrumental music, not all the way, but towards greater or lesser durational variability. Part of the strength of the comparison in the study was the
narrow-range of music studied, French and English instrumental themes by late-nineteenth century composers appearing in the 1983 edition of Barlow and Morgenstern’s *A Dictionary of Musical Themes* (B38). This narrow range of music contributes to a well-controlled correlational study (B38).

Huron (2008a) proposes that musical homogeneity is coming and culture-specific musical features and aesthetics are on their way out. Western scales and harmony are a big part of this, but technology, including equal-temperament instruments, may be the driving force. In our transition to a global monoculture there are ample opportunities to study how different cultures use the same technology differently. French and English musicians used similar instruments in the nineteenth century. Now much of the world uses keyboards and guitars and digital recording technology. Bode Omojola writes about the nuances of African Pianism (2001). Likewise, Èkwúèmé has written about “African music retentions in the new world” (1974a). Instruments and technology may not only be used to develop novel styles, as in African-American gospel and jazz, but entirely repurposed. Hip-Hop turned a passive technology (the turntable) into an active technology (Katz 2010). Cultures around the world are being shocked by new instruments, technologies and mass media. However, we know there is the possibility of cross-generational and cross-continental retentions. This is evidenced by black arts in the Americas. But what are these traits? Antiphonal singing and polyrhythmic drumming are often cited as features of African music. In this chapter, melodic angularity and semblant motion in vocal harmony are proposed as musical features that extend from the presence of lexical tone.
16.2 Contrastive versus expressive temporalities

How does a tone language speaker access her lexicon when listening? At some point during the expansion of the Niger-Congo language family east and west, some languages crossed a threshold where tone became non-contrastive at the lexical order. However among many prototypical Niger-Congo tone languages, tone contours are the same across dialects even with significant allophonic variance. Consider the data in Table 16.1 from Clark (1990).

<table>
<thead>
<tr>
<th>Gloss</th>
<th>Òniça Dialect</th>
<th>Central Dialect</th>
<th>Tone Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reach the farm</td>
<td>Lùà ọ̀ lùà</td>
<td>Rùò ọ́ ũùì</td>
<td>LH H H</td>
</tr>
<tr>
<td>He has a bicycle</td>
<td>Ò-ŋwèlù ́iğwè</td>
<td>Ọ-ŋwèrè ́iğwè</td>
<td>HLL LH</td>
</tr>
<tr>
<td>Uzo has eaten yams</td>
<td>Ìù ́ Ố ̀ èlgó ji</td>
<td>Ìù ́ Ố ̀ èrìèlè ji</td>
<td>HL LHH H</td>
</tr>
<tr>
<td>They didn’t buy anything</td>
<td>Fá ̀ ègòrò ́ ifè</td>
<td>Há ́ ̀ à ̀ ́súyì ́ iyè</td>
<td>H H H H H</td>
</tr>
</tbody>
</table>

Table 16.1: Allophones in Ìgbò dialects (Clark 1990:17)

It is cherry-picked but represents a very real phenomenon. For our tone interval study, we used Òniça dialect but more speakers of Central dialect and Enugu dialect participated. A few participants reported some moments of hesitation at the beginning of the task. Once they realized it was not their dialect, they were able to perform the task.

<table>
<thead>
<tr>
<th>Gloss</th>
<th>Òyó</th>
<th>Ibápá</th>
<th>Ibólló</th>
</tr>
</thead>
<tbody>
<tr>
<td>To work</td>
<td>šíşé</td>
<td>tsítsé</td>
<td>şişé</td>
</tr>
</tbody>
</table>

Table 16.2: Allophones in Yorùbá dialects (Crowther 1852:1)

Crowther also pointed out the cross-dialect consistency of tonemes despite the presence of significant allophonic variation. I can imagine an entirely different orthography for these languages not governed by the alphabet. In which detailed transcription of tone is
primary and phonic content is secondary. It will probably never happen, but a new approach to writing these languages might resolve standardization issues, which have challenged Igbo language in particular. By decentering the focus on phonic content that varies so greatly from dialect to dialect, and recentering on tonemes that are invariant across dialects, Igbo orthography might enter a new era. Toneme consistency across dialects and the mapping of tone to speech surrogates like the talking drum indicate some linguistic signals are intelligible when stripped of phonic content. Tone may not only be important when it is an essential contrast, like in a homophone group, but may be even more important to all lexical processing, syntax and semantics than presently thought. In Yorùbá, the metalanguage for tone diacritics is ami ohun, or sign of voice. It is not clear when this term originated, but some time between Crowther’s 1852 grammar and Abraham’s 1962 dictionary, it appears in the latter not the former.

While I disagree with Deutsch that there is attention to absolute pitch required for tone language comprehension, there is clearly special attention paid to the pitch dimension in tone languages. In Chapter 6, the stratification of the signal into different temporal scales within a single dimension was addressed. Lexical tone and intonation are both present at different temporal scales in the pitch-dimension of Niger-Congo tone languages. Because of the contrastive nature of the lexical-order scale of the pitch dimension, there are more strictures on the use of pitch within vocal music primarily and to a lesser extent instrumental music. Composers and vocalists do not freely manipulate lexical-order contour because of the significance of such changes. There is more flexibility on the larger-scale phrase order in music, which is the analog to paralinguistic intonation. Although Èkwúèmé suggests that tone is tied to rhythm, I disagree. I think the
strictures of the pitch dimension draw attention away from the rhythmic domain, allowing for greater freedom and complexity as found in West African polyrhythms. The complement is cultures of Europe with stress- and syllable-timed languages where there is a freedom and complexity of pitch. What other cultures have music such as chorales and hymns that are so incredibly dull in terms of rhythm but often complex in terms of harmony and voice-leading? I argue the linguistic and paralinguistic scale-dimension temporalities of a language (see Chapter 6) correlate to constrained and expressive temporalities in music of the same culture. Musical aesthetics are influenced by our linguistic attentions and inattentions. The harmonic complexity of Northern European music and rhythmic complexity of West African music are juxtaposed in African-American music, which is without doubt some of the most compelling music humans have produced. Both Nketia and Èkwúèmé reject missionary-ethnomusicologist A. M. Jones’ assertion about the lack of pitch organization in African music. There are unique approaches to pitch in Africa, and they are often quite systematic.

16.3 Niger-Congo Aesthetics

Both linguist Will Leben (1983) and musicologist Kofi Agawu (1988) point out that realizing speech contours in music is an art, not a science. One aspect that frees composers and improvisers of song to some extent is that tone is not always contrastive unit. On the other hand, vowels often play a small role in lexical processing and allophones are common. Even when a word is not part of a homophone group tone may contribute to lexical processing. Clear mapping of tone-to-tune may be observed in chant
and song that is not conformed to Western tonality (such as much Yorùbá poetry and praise-singing). Schellenburg (2012) includes a meta-study of various tone-tune correlations from Africa and Asia that suggests traditional music generally has stronger faithfulness to linguistic tone than popular music (that is presumably more influenced by encroaching Western tonality). Many aesthetics associated with Western tonality and harmony are incongruent with the aesthetics of ethnolinguistic cultures with tone. One of the most apparent incompatibilities is contour-tone languages in which speech has pitch trajectories on the syllable and the pitch stability of Western instruments and singing. Peking opera is rife with pitch glissandi that are quite foreign to Western ears. The practice of singing on stable pitches and not sliding between them (given so much prestige in the European tradition) seems to be unique to ethno-linguistic cultures with stable pitch instruments (like the piano) and may be a relatively recent phenomenon (Schubert and Wolfe 2014). In recent years, a fondness for affective (not corrective) auto-tune has developed among West African recording musicians. However, many other aesthetic preferences, such as step-wise motion and contrary harmony, associated with Western tonality distort Niger-Congo linguistic tonality. The presence of two or three tone-levels is common from Ghana to Cameroon and these ethnolinguistic cultures also tend to share some musical aesthetics.

16.3.1 ANGULARITY

The general angularity of Niger-Congo tone-language speech is reflected in the melodic character of the music. Kolinski describes a song from Dahomey, a neighboring kingdom
to Ōyó, as “bold and ragged” (1965: 116). Kolinski’s Western-acculturated ears, which are accustomed to small intervals between pitches (steps and skips) and only rarely large intervals (leaps), underlie his somewhat pejorative assessment. Different language typologies correspond to distinctive musical aesthetics.

Figure 16.1: transcription and contour analysis of a Dohomean song (Kolinski 1965)

EX. 114

EX. 115. Melodic Structure of the Song

Figure 16.1: transcription and contour analysis of a Dohomean song (Kolinski 1965)

Angularity in modernist music (such as the Schoenberg piece analyzed in Chapters 9 and 10) goes against popular conceptions of musicality. In the case of Schoenberg and the
Second Viennese School, it is a sign of sophistication to befuddle perception and common sensibilities. Such a sanguine view is not afforded the Dahomeans.

According to a posthumously published chapter by Nick Clements, a prominent theoretical linguist who died in 2009, African tone systems form scales. This is manifest in speech surrogates in which tone levels are mapped to musical pitches. This mapping can be made to variable pitch or fixed pitch instruments and if the mapping is to fixed pitches, the mapping may not be one-to-one. Neither Clements statement nor the Do-Re-Mi folk heuristic imply that superparticular ratios as in the diatonic scale are at play. It also does not mean that these are the same as musical scales (as Ladd 2008 is careful to note) or that these cultures do not also have musical scales (as Èkwúèmé notes in multiple writings). Instead, as we saw with the tuned drum rows in Chapter 10 (based on Nzewi 2001) and Èkwúèmé’s Igbo Glees in Chapter 15, the tone-level scale imposes itself on the notes of a musical scale in manifold ways, exerting its presence through contour.

In tone-level scales, steps between scale degrees are bigger than in diatonic or even pentatonic scales. This explains Kolinski’s account of the Dahomey melody as “bold and ragged.” What is disjunct to Kolinski may be conjunct to a Dahomean. The language of the Dahomey Kingdom is Fon, a two-level language. I cannot know without a text with diacritics, but drawing on Èkwúèmé’s writing and text setting of Igbo (also two-levels), the segment in the blue box of Figure 16.2 in may be at a single level.
The boxed figure in Figure 16.3 might be an alternation of HLHLH.

As Huron (2001) proposes, aspects of smoothness in voice-leading in Western music, may be attributed to the Pitch Proximity principle. Motivations for breaking through the “fission boundary” are fewer in common-practice European music, and therefore leaps are very “marked.” There are more motivations for fission in both instrumental and vocal music in tone language cultures. The tone perception study in Chapter 5 provides useful evidence for this. The lack of leaping began to change in Western music with nineteenth-century gospel hymns and Scott Joplin’s rags. The angularity in this music may be a retention from African music in the new world, much like syncopation (Ékwuèmé 1974a). The motivation for frequent fission is not the same because tone languages are not spoken by African-Americans, but there may be influence, just the same. Yorùbá dialects are spoken in Brazil, Cuba and Haiti.
16.3.2 SEMBLANT MOTION

In the socio-cognitive realm of Niger-Congo cultures melodic inversions do not share a common Gestalt. It is not certain that they are in Indo-European cultures either, but I think this is a fairly strong premise within the 60% of the world’s languages which are tonal. Inversions are used, but it is for contrastive variation, not likeness. In Yorùbá poetry, the poetic device Olátunji calls tonal counterpoint involves an inversion of tone at the end of a phrase (not including the entire phrase, see Chapters 11, 13 and 14). This is a very amusing “fish-out-of-water” type device and does not produce melodic gestalts. Additional evidence for this position comes from the use of similar and parallel (or collectively, semblant) motion in Christian praise-singing in Nigeria. Even in Westernized music, even when the melody itself does not strictly follow tone, singers use similar motion to harmonize Ìgbò and Yorùbá melodies. They are harmonized almost exclusively with semblant motion even when there are violations of lexical tone. This makes hymns with contrary motion even more problematic within tone language cultures.

In a 2013 study, we found that when alto and tenor singers harmonized a soprano melody in an indigenous language they are fluent in, they restrict harmonic movement to semblant motion. More oblique motion is present in the harmonization of English-language choruses. The singers that participated were multilingual and sang in professional-level Western-style choirs that performed repertoire in English, Latin and Nigerian languages. The assistant choirmaster selected tunes for the choristers to harmonize, three hymns and choruses for each language: English, Igbo and Yoruba. Three ethnically-Ìgbò singers harmonized Igbo and English pieces, three ethnically-Yorùbá singers harmonized Yoruba and English pieces. One of the Igbo hymns was “Oha
Kele Ike Jesu” (All hail the power of Jesus’ name), introduced in Chapter 1. Recall that when sung to the preferred British hymntune Coronation an unfortunate obfuscation of the tone occurred, changing the meaning to “Trees hail the buttocks of Jesus.” This melody was not used. Instead, an original melody composed for the translation of the hymn-text was used (the composer is unknown). This appears with a transcription of the improvised harmony in Figure 16.4.

Figure 16.4: Improvised harmony for “Oha Kele”
The harmony in chorus singing from the study of semblant motion in improvised choruses lies far from the centers of distributions for each type of harmonic motion based on a massive corpus of 7,000 hymns. There is far more parallel and semblant motion and far less oblique and contrary motion (see Figure 16.5).

Figure 16.5: Distributions of harmonic motion
Bass voices of the hymns, which are more likely to move in contrary motion, are excluded from the distributions. The plots only include data for upper-voice harmonic motion. The plot for semblant motion includes both similar and parallel movements. Roughly 80% of harmonic movements are in the same direction for the improvised choruses (labeled “AllNigeria” in the legend). Much of this is strict parallel motion (nearly 70%). The mean similar motion for the hymn corpus is roughly half (39%) of the mean for the improvised choruses recorded in Nigeria (~80%). Although it was not tested, I predict that much more semblant and oblique motion is more prevalent in improvised singing in all cultures. And, that contrary motion is a rarefied practice specific to composed homophony in European cultures.

16.3.3 VARIATION WITHIN NIGER-CONGO CULTURES

In accord with a weak version of the Sapir-Whorf hypothesis, I believe the presence of some linguistic determinism (that Èkwúèmé supports in Ìgbò music, but Agawu found little evidence for in Ewe music) is triggered not by whether a language is tonal or not, but whether tonemes continue to play an active role in language comprehension, or are simply a residual trait from a language’s ancestors. If tone is strongly contrastive within an ethnolinguistic culture, this may inform small and large structures and broadly the aesthetics of music with words, surrogate speech and even instrumental music, extending beyond the more obvious cross-domain mapping of tone and tune.

Interviews suggest Yorùbá is in some ways more strongly tonal than Ìgbò, that Yorùbá has a “lock to it” while Ìgbò does not. Analysis of two dictionaries, one Ìgbò (Williamson 1972) and one Yorùbá (A Dictionary of the Yorùbá Language), indicates
that tone is the sole contrast in 57% of Ìgbò disyllables but only 40% of Yorùbá disyllables.

<table>
<thead>
<tr>
<th>Entries</th>
<th>Yorùbá</th>
<th>Ìgbò</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contrastive Tone</td>
<td>1299</td>
<td>461</td>
</tr>
<tr>
<td>Percentage</td>
<td>40.0%</td>
<td>57.2%</td>
</tr>
</tbody>
</table>

Table 16.3: Disyllables that form tone-varied minimal pairs

The “lock” may refer to the greater number of tone levels in Yoruba, which has undoubtedly influenced the linguistic theory that frequency banding is necessary for this language.

While counting homophones in a dictionary supports the importance of tone in Ìgbò, a two-level system does allow for more flexibility in tone realization and thereby paralinguistic effects. In Èkwúèmé’s settings, tones at the beginning of an Ìgbò utterance have higher fundamental frequency (f0) or pitch than equivalent tones at the end of an utterance. Èkwúèmé’s setting intentionally represents declination. Consecutive equivalent tones may have decreasing f0 (downstep), contributing to overall phrase declination. Conversely, Adeyemo’s Yoruba poetry controlled declination. While all languages exhibit declination, likely because of the universal effect of diminishing air pressure in the course of a phrase, the more tone levels are present, the more declination may be controlled.
16.4 Ethnographic Case Study: F.I.L.M.

The Forum for the Inculturation of Liturgical music (FILM) is a biennial choral competition between choirs formed of college-age Catholic students from across Nigeria. Each choir presents an original setting of the mass in one of the indigenous languages spoken in the geographical region of their university. Language counts for Nigeria are above 200 languages at the low end. Some counts are as high as 500. This creates a lot of potential for new settings of the mass. The settings are required to have tone-tune correspondence. According to the competition coordinator:

The tonal inflection is not explicitly judged during the festival as all the judges would not be fluent in all the languages presented. It is left to the diocesan liturgical/music commissions who are expected to have experts in the languages spoken within their dioceses. Thus if a song does not respect the tonal inflections of the language (which would lead to perceived meaning being different from the intended) the committee will not approve the song for liturgy (correspondence with Anthony Okoro, October 2015).

Many of the languages do not have grammars or dictionaries, so a comparison of contour adjacency series of the underlying speech tones and their musical realization is not possible. The experts may be making the assessment more holistically.

Although, tone-and-tune is screened prior to the competition, there is much discussion of it at the competition, which I found delightful, enlightening and fascinating. There is constant informal assessment and debate by fluent and/or non-fluent speakers of a piece’s tone-tune acumen, which may attribute some specific ethnic associations to the music, but basically boils down to whether a piece sounds “African”.

474
I discussed the breadth of tone-tune assessment that goes into the competition with a colleague at the University of Lagos who attended some of the competition with me. He insisted that if you speak one Niger-Congo tone language and are familiar with the musical aesthetics of the region, it is not hard to make the judgment about similarity between the underlying speech tones and melody, even for a language you do not speak. Angularity and semblant motion may signal tone-tune correspondence, even if a more objective comparison of tone and tune fails.

16.5 Analytical Case Study: Òjúkwú’s “Tètá”

Òjúkwú’s “Tètá” (date of composition unknown) is a setting of an Ìgbò translation of Ephesians 5:14, “Wake up, O sleeper, rise up from the dead...” “Tètá” means “wake up” and “bìlíe” means “rise up.” The lexical tone contour of ascent, common to [tètá] and [bìlíe], serves as a micro-motive in the piece which has frequent leaps upward.
A performance of the piece by the Lagos City Chorale is available at this link (<https://youtu.be/yZcawAHg3to?t=2m30s>). Their Pan-African approach is indicated by the use of both the Yorùbá dundún and the Ìgbò údù to accompany an Ìgbò piece. This analysis corresponds to 2:30 (where the link starts). There are several repetitions of the ostinato (Figure 16.7) before the soprano melody starts (Figure 16.8). In the score I have indicated the refined contour implied by the lexical tone with arrows between adjacent notes, except at phrase boundaries. Violations of lexical and interlexical tone within phrases are indicated with red boxes.

**OSTINATO CHORUS & XYLOPHONE**


*Alto, Tenor, Bass & Soprano are to start from the starting point*

Figure 16.7: “Tètè” ostinato section, p. 1. (xylophone omitted)
A key difference in Òjúkwú’s music from Èkwúèmé’s Igbo glee (analyzed in Chapter 15), is that after initial presentation, there is less faithfulness to tone, both in terms of direction and magnitude. This is not uncommon. It is Èkwúèmé that is quite exceptional.

Analyses of both speech and musical settings indicate lexical content demands less
faithful realization in repetition than in initial presentation. Once the meaning has been established this may generate interesting variation. A prime example of this is tonal counterpoint that intentionally obfuscates tone in repetition (see Chapter 11). Despite having more flexibility after initial presentation of text, Òjúkwū still uses a high degree of angularity: 87 out of 204 notes qualify as “pivots” (changes in direction). He also relies heavily on semblant motion, not only when text aligned, but also in staggered text. This indicates that angularity and semblant motion may be initially motivated by lexical tone, but these features persist even in the absence of tone-tune mapping.

<table>
<thead>
<tr>
<th>Disyllable</th>
<th>Symbol</th>
<th>Semitones</th>
</tr>
</thead>
<tbody>
<tr>
<td>#tè.tá#</td>
<td>↑</td>
<td>12</td>
</tr>
<tr>
<td>#sì.kwá#</td>
<td>↑</td>
<td>5</td>
</tr>
<tr>
<td>#n’à.rá.</td>
<td>↑</td>
<td>5</td>
</tr>
<tr>
<td>#bí.líe#</td>
<td>↑</td>
<td>3</td>
</tr>
<tr>
<td>#g’à.mú</td>
<td>➞</td>
<td>3</td>
</tr>
<tr>
<td>#g’ó.nyé#</td>
<td>➞</td>
<td>0</td>
</tr>
<tr>
<td>.rá.r’ú.</td>
<td>➞</td>
<td>0</td>
</tr>
<tr>
<td>#nwú.r’á.</td>
<td>➞</td>
<td>0</td>
</tr>
<tr>
<td>.r’á.nwú#</td>
<td>➞</td>
<td>0</td>
</tr>
<tr>
<td>.mú.kwá#</td>
<td>➞</td>
<td>–3</td>
</tr>
<tr>
<td>.r’ú.râ#</td>
<td>➞</td>
<td>–3</td>
</tr>
<tr>
<td>#á.hù#</td>
<td>▼</td>
<td>–4</td>
</tr>
</tbody>
</table>

Table 16.4: Initial Presentation of Disyllables in Òjúkwū’s “Tè tá”

In the analyzed ostinato section of the piece, 90% (C+SIM (1 degrees) = 124/138 = 89.9%) of adjacent contour comparisons excluding phrase boundaries (||) match. Several of the mismatches are repeated tones not realized at the same pitch. This is to some extent consistent with speech, and if taken into account, would increase the contour similarity calculation. Ranges of refined contour categories meet, but do not overlap. These can be looked at as ambiguities but not contradictions (see Rahn 1991). This suggests Òjúkwū
conceives of discrete interval ranges for each interlevel. There are many different possible mappings of interlevel to melodic interval. Based on interviews, some composers intentionally represent magnitude, others do not. Among those that do, they disagree about key aspects, such as whether a semitone or whole-tone up or down is still at the same tone level.

Ọjúkwū and Èkwúèmé are both widely performed Ìgbò composers who attend to lexical tone in text setting, but there are differences in their approaches. First of all, Èkwúèmé often writes secular works while Ọjúkwū does not. This makes a big difference in whether lexical tone is important because translated Christian texts have more loan words in which tone is merely conventional while in Ìgbò proverbs tone is highly important. The proverbs in Èkwúèmé’s “Obi Dimkpa” could be played on speech surrogate instruments without any phonic content. Ọjúkwū’s music also tends to have more repetition of single words than Èkwúèmé. In repetition, Ọjúkwū is flexible with magnitude he gives to interlevels (particularly LH) while Èkwúèmé remains consistent, even in repetition. Èkwúèmé’s text setting is strictly imitative between voices, and presentations of the same word are largely consistent within the work. In general, Ọjúkwū is liberal with the setting of a particular word after its initial presentation.

Regarding texture and form, Èkwúèmé’s “Obi Dimkpa” uses a call-and-response texture with alternations of verses and refrain. Ọjúkwū’s “Teta” uses a mixed texture ostinato with a descant above it. These textures and structures do not describe all of their music, but do correspond to strong Western influences that define the style of each composer: Èkwúèmé sang and at times conducted the Yale Glee Club during his doctoral work. Handel’s oratorios (and not just Messiah) are widely performed in protestant
churches in Nigeria. Another one of Òjúkwú’s pieces, “Jehova Me Wo,” has been described to me as the “African Hallelujah Chorus.”

16.6 Influence of music on language

Language exerts an influence on music, particularly the vocal arts, but what about the reverse? In the West African oral tradition, poet-singers act as curators of language: repositories of tales, proverbs and riddles, and sources for deep language. In contemporary sound culture, recording artists have taken on this role to some extent. In the 1970s, Fẹ́lá Kuti took a radical step towards pan-Africanism by using Pidgin English in his music almost exclusively. Kuti’s use of language, including caustic portmanteaus like “colomentumality” for colonial mentality, continue to influence both colloquial speech and popular music, including Naija Hip-Hop and “Afrobeats” in the twenty-first century. Faze’s “Kolomental” (2006) is just one example of popular culture perpetuating Kuti’s influence.

A more specific instance of language change articulated, and perhaps influenced, by music is the song “Igwe,” a multi-lingual gospel hit from 2009 (the music video may be viewed here: <https://youtu.be/93zmgKydbml>). Like Òjúkwú’s “Teta” the same word is repeated over and over. Also like “Teta,” the direction between syllables in each setting is consistent but interval magnitude varies. The setting of the word [tètá] (HL) varied from an ascending melodic 2nd to an entire octave. In Midnight Crew’s “Igwe” 2009, the word “Ígwě” (H!H or HS) is sung or spoken 57 times in the recording and set to a variety of musical intervals by lead vocalist and in choral responses including diatonic 2nds, 3rd
and 4ths. In effect, [ígwē] becomes the inverse of [tète], a musical micro-motive of
descent. However, /te.ta/ which does not have homophones, while /i.g"e/ has three:
[ígwē] king, [ígwè] iron or bicycle, and [ígwè] crowd. In the popular gospel song, the
intended meaning is “king,” referring to the Abrahamic god. A downward contour on the
homophone can also mean bicycle, so there is potential ambiguity.

\[
\text{Figure 16.9: Backup singers’ refrain}
\]

The refrain of the backup singers cycles through intervals: m3 (3 semitones), M2 (2 s.t.),
m2 (1 s.t.), M2 (2 s.t.), and repeats. The soloist elaborates further. The video also
reinforces a single meaning for the homophone: the word appears without diacritics on
shirts and signs throughout the video.

\[
\text{Figure 16.10: A still from the music video for “Igwe” (2009)}
\]
In 2012, I conducted 20 interviews in which I gave participants a list of twenty homophones in Ìgbò and Yorùbá without tone marks and asked them to teach me all the meanings associated with each homophone, /igwe/ among them. The respondents were of various ethnicities from southern Nigeria and all had some knowledge of both languages. Very few, however, knew multiple meanings for the homophone /igwe/, most knew only one: king. The emerging dominance of “king” in the homophone group and the popularity of the song is merely correlational. Yet, the high word frequency of ìgwē (king) within multi-ethnic gospel music, and contemporary Nigerian Christian culture in general, likely catalyzes language change.
17. CONCLUSION

17.1 African Tone Systems (Reprise)

In African tone systems, a pitch segment (or tone) may be equivalent to another pitch segment through relative pitch height within a local environment, regardless of absolute pitch height. The notion of phonological equivalence of tones has been applied to non-African languages since the 1970s. It may also be applied to music as a novel form of melodic invariance, distinct from transpositional and complex-contour equivalence.

Many of the elaborate tools developed for studying pitch in music theory are relevant and may be fruitfully applied to speech analysis. Dilley (2005) and Ladd (2008a) have demonstrated this, but with caveats. Many tone-language speakers in Nigeria already do this informally through the Do-Re-Mi folk heuristic. There was not a strong formal concept corresponding to phonological equivalence of tones in western culture, so when British culture injected itself into Niger-Congo culture, solmization was adapted as a heuristic. So, where do tone levels in African Tone Systems and their unique form of phonological equivalence fit in to other concepts of pitch organization? Existing models from music theory include:

- melodic intervals: vectors with a direction and magnitude component;
- contour pitches: heights ranked low (0) to high (n-1) within a melodic segment;
- scale degrees (do-re-mi);
- notes (absolute pitch);
The following is a comparison of tone levels to pitch models from Western music theory, particularly to scale degrees. This comparison is implied by the do-re-mi heuristic, most associated with Yorùbá but also present in other Niger-Congo cultures. Throughout this section, the tone system of Yorùbá is used as a model, with some mention of Ìgbò’s two-level system. These languages are representative of other Niger-Congo languages to some extent.

17.1.1 ABSOLUTE PITCH

The do-re-mi heuristic (see Chapter 2) implies tone levels are not tied to absolute pitches. Ladd (2008a) refers “speaker-specific” scales or Dilley (2005) “speaker-specific” referents implying tones must adjust to speaker range. In addition, paralinguistic intonation may also modify the pitch height of equivalent tones within a speaker’s range (see Chapter 4). Tone levels may adhere to frequency bands, but it is not necessary for speech to be intelligible.

17.1.2 STABLE PITCH

Schubert and Wolfe (2014, Wolfe and Schubert 2010) suggest stable pitch in vocal production arises from the presence of stable pitch instruments in a culture. Some Niger-Congo cultures have instruments or sets of instruments that produce a collection of stable pitches which serve as surrogates for speech tones, others do not. The Yorùbá talking drum is variable-tension and can produce stable or sloped pitches on instruments of varying size. Tones played on the talking drum are analogous to Yorùbá speech tones.
Figure 17.1: The same phrase with one tone altered “Mo mú [igbá/igbá] wá fún e”

<table>
<thead>
<tr>
<th>Syllable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text</td>
<td>Mo</td>
<td>mú</td>
<td>i-</td>
<td>gbá</td>
<td>wá</td>
<td>fún</td>
<td>e</td>
</tr>
<tr>
<td>Tone Level</td>
<td>Mid</td>
<td>High</td>
<td>Mid</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>Mid</td>
</tr>
<tr>
<td>Mean Pitch (ST)</td>
<td>47.3</td>
<td>48.5</td>
<td>46.9</td>
<td>49.2</td>
<td>48.4</td>
<td>48.0</td>
<td>46.3</td>
</tr>
<tr>
<td>Interval Direction</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Pitch Difference (ST)</td>
<td>1.2</td>
<td>-1.6</td>
<td>2.2</td>
<td>-0.8</td>
<td>-0.4</td>
<td>-1.7</td>
<td></td>
</tr>
<tr>
<td>Duration (sec)</td>
<td>0.403</td>
<td>0.283</td>
<td>0.467</td>
<td>0.398</td>
<td>0.381</td>
<td>0.413</td>
<td>0.396</td>
</tr>
<tr>
<td>Power (normalized)</td>
<td>0.06</td>
<td>0.36</td>
<td>0.12</td>
<td>0.77</td>
<td>0.46</td>
<td>0.12</td>
<td>0.11</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Syllable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text</td>
<td>Mo</td>
<td>mú</td>
<td>i-</td>
<td>gbá</td>
<td>wá</td>
<td>fún</td>
<td>e</td>
</tr>
<tr>
<td>Tone Level</td>
<td>Mid</td>
<td>High</td>
<td>Mid</td>
<td>Low</td>
<td>High</td>
<td>High</td>
<td>Mid</td>
</tr>
<tr>
<td>Mean Pitch (ST)</td>
<td>49.2</td>
<td>50.2</td>
<td>46.8</td>
<td>43.5</td>
<td>44.4</td>
<td>47.9</td>
<td>46.0</td>
</tr>
<tr>
<td>Interval Direction</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>0</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Pitch Difference (ST)</td>
<td>1.1</td>
<td>-3.5</td>
<td>-3.3</td>
<td>0.9</td>
<td>3.5</td>
<td>-1.9</td>
<td></td>
</tr>
<tr>
<td>Segment Slope (ST/sec)</td>
<td>2.0</td>
<td>-2.6</td>
<td>-8.0</td>
<td>-2.3</td>
<td>4.6</td>
<td>-4.0</td>
<td>-0.1</td>
</tr>
<tr>
<td>Duration (sec)</td>
<td>0.465</td>
<td>0.320</td>
<td>0.405</td>
<td>0.277</td>
<td>0.510</td>
<td>0.555</td>
<td>0.450</td>
</tr>
<tr>
<td>Power (normalized)</td>
<td>0.38</td>
<td>1.00</td>
<td>0.33</td>
<td>0.16</td>
<td>0.12</td>
<td>0.13</td>
<td>0.15</td>
</tr>
</tbody>
</table>

Table 17.1: Measurements for each syllable segmentation of the two homophonous 7-syllable phrases

86 For a description of the measurements in the table, see Chapter 1 where it first appears.

485
Depending on the syntagmatic context, speech tones may be stable or sloped and they adapt to different voice ranges (Ladd 2008a:198). In Figure 17.1 and Table 17.1, the upper plot shows a fundamental frequency (f0) extraction using the YIN algorithm in MATLAB. The y-axis is log frequency and the x-axis is time.

Solmization (do-re-mi) also adapts to range, but in principle does not describe sloped (rising or falling) pitch. The ideal of solmization is stable and discrete pitches, but in reality “scooping” between pitches often occurs in singing in the west and elsewhere. Despite the implied prohibition of sloped pitches, Tonic Sol-Fa fits with enculturated pitch perception among bilingual Yorùbá speakers even if it is not representational of the pre-colonial epistemology. The lack of Sol-Fa syllables to describe the sloped pitch trajectories that occur in Yorùbá is less problematic under Bamgbose’s orthography (first published 1966), which divides sloped syllables into two segments.

17.1.3 LOCAL FREQUENCY BANDS

In music, ideally 'do', 're' or 'mi' remain close to the same absolute frequency each time they occur within the same octave while one is singing in a key. This is usually not true of 'do', 're' or 'mi’ speech tones. In the Yorùbá sermons in Chapter 4 the phrase-order speaking range gradually rose by as much as an octave. It is a bit like a singer who does not hold pitch (maintain a key) but preserves the contour of a melody quite precisely. Speech is less periodic (and less precise with pitch) and significant paralinguistic modification is possible in Niger-Congo languages (see Chapter 4), so frequency bands often fail to describe tone realization even on a local level.
In Ìgbò, frequency-banding may characterize Low tone level to some extent. Low in non-elevated, relaxed speech tends to sit at the bottom of the speaking range. High tone level tends to descend over the course of an utterance, reflecting declination, and then resetting to a higher pitch once subglottal air pressure has been renewed. Laniran and Clements (2003) describe a similar phenomenon for Mid tone level in Yorùbá. However, if more effort is used, in excitement or anger, all of the tones may rise towards the top of the vocal range. Tone adapts to different speaker ranges and their arousal level (see Chapter 4). I once heard a man with an extraordinarily high and narrow range speaking Yorùbá. He could manipulate the tones very effectively, though his vowels were a bit unusual because he was speaking within the first formant range.

17.1.4 RELATIONAL (SYNTAGMATIC)

Instead of conceiving of tones as strictly paradigmatic “atomic units” with segmental features, the do-re-mi heuristic implies some relationality. I might be thinking ‘mi’ when I sing my first note, but unless the listener knows where ‘do’ is, this is meaningless. I suggest that tone may be produced paradigmatically, but it is largely perceived in context (syntagmatically) and requires initialization for intelligibility. With pitch reset, which usually occurs with each new utterance, a new “key” is established. Single-syllable words may be understood in isolation, extremes of range or exaggerated pitch trajectory may be used to realize a single high-tone or low-tone, but in longer utterances this very marked use of range is not necessary because a different mechanism is possible: syntagmatic direction (contour).
17.1.5 DISCRETE INTERVALS

Following from the variation in tone level location, there is also no set interlevel distance in either frequency or logarithmic frequency (semitone) scales. Ladd (2008a) suggests that tone spacing normalizes to speaker range. Chapter 4 demonstrated that speakers tend to maintain a roughly similar interlevel distance in semitones whether they are using the upper or lower register of their voice. This is evidenced by fairly constant standard deviation of mean pitch even as the mean pitch varied significantly. Based on experimental data in Chapter 5, there does seem to be a minimum threshold for interlevel distance, a “fission boundary” (after Van Noorden 1975). Our study also suggested that clear interlevel contrast requires spacing between levels greater than the two-semitone distance suggested by do, re and mi. A more likely tone dispersion is roughly do-mi-so or do-fa/so-do’ (the second do an octave higher). This suggests that Welmers’ description of spacing was accurate, though the association of tone levels to notes, even on the local level, is problematic. This is also a fault with the do-re-mi heuristic. Demany, Semal and Pressnitzer (2011) provide evidence that perception of pitch as higher or lower than the previous pitch is automatic among WEIRD\(^{87}\) participants. Also based on findings from relatively WEIRD participants (but ones that can speak Mandarin) Bidelman and Chung (2015) suggest that there is lateralization of direction (contour) and magnitude (interval) processing in tone language speakers. Directional cues are much stronger than distance cues.

\(^{87}\) Western, educated, industrialized, rich and democratic.
17.1.6 Tendencies and Qualities

Of the three conventional Yorùbá tones (low-mid-high), mid is not marked by a diacritic and fits comfortably within the range, with the speaker neither expending the effort to “elevate” or “depress” the tone true to Crowther’s original description (1852). As overall effort in speaking increases or decreases (perhaps in line with subglottal air pressure or vocal fold adduction), so too does the baseline frequency of the neutral (unmarked) mid tone. High and low on the other hand are marked both orthographically and physically. From the baseline mid level, high and low are elevated or depressed in pitch respectively. If this is extended to the do-re-mi heuristic, re is the unmarked category while 'do' and 'mi' form marked poles around 're'.

The scale degrees do-re-mi also have tendencies and qualities that bear some similarity to the linguistic concept of markedness. John Curwen (see Chapter 2) developed a handsign system (chironomy) for Tonic Sol-Fa to reflect this (see Figure 17.2).
The chironomy was not widely used by missionaries, unlike Tonic Sol-Fa notation was, but was spread through Europe and America through the Kodály method. Each sign physically and visually indicated the melodic energy of each scale degree and had an impact in the expansion of tendency tones in nineteenth-century theory (Day-O’Connell

---

88 Scanned and enhanced by Matthew D. Thibeault [Public domain or Public domain], via Wikimedia Commons <https://commons.wikimedia.org/wiki/File%3ACurwen_Hand_Signs_MT.jpg>
2002:38). Pertinent to this discussion, Curwen describes: Do as “the STRONG or firm tone”; Ray as “the ROUSING or hopeful tone”; Me as “the STEADY or calm tone”; and Soh as “the GRAND or bright tone” (1880:iv). The neutral center flanked by marked features formed by the three tone levels, low-mid-high, resembles Curwen’s attributes for do-mi-so not do-re-mi. However, recent music theory tends towards ‘mi’ as the marked element of the triad. In a 2006 presentation, David Clampitt argued for a binary (marked or unmarked) status for each of the notes of the triad. In this model, the third (middle note of the triad) is marked as ‘agent’ and the two outer ‘pillars’ (root and fifth) are both unmarked. This deviates from the trichotomy of base, agent, and associate in Harrison’s model (1994). In both however, the third is the most marked element.

Between the three tone levels there are no other tone levels, just as there are no other scale degrees between do, re and mi, only chromatic alterations. A scale-step from do to re or re to mi, is like a step between adjacent tone levels. This is the aspect of the tone levels and scale degrees that is most salient to each other. Thus, although do-mi-so is more similar in terms of extent and in terms of Curwen's attributes, do-re-mi has more similarity to the tone levels in the most important regard: adjacency in a scale.

17.1.7 MARKEDNESS AND DIATONIC SYMMETRY

Do-re-mi also has semblance to the symmetry of the scale: low-mid-high=marked-unmarked-marked=do-re-mi. Re is the only diatonic scale degree which forms identical interval patterns in both directions (presented here in semitones):
If the symmetrically-balanced center of the diatonic pattern (re) is classified as the least marked element, than a three-tone system in which the middle level is unmarked is aptly modeled with do-re-mi. This can be extrapolated to other discrete tone-level systems:

- A two-level system with the lower level unmarked as Re-Mi;
- A two-level system with the upper level unmarked as Do-Re;
- A five-level system with the middle level unmarked as Ti-Do-Re-Mi-Fa, where Ti is relatively very low and Fa is relatively very high, a doubly-marked extension of Do and Mi with strong inward-pointing tendencies.

The use of ‘high’ and ‘low’ for two-level systems (like Ògbò) has no indication of which belongs to the marked and unmarked category, and which tends towards being the “null tone” in surface underspecification (see Myers 1998). Distinguishing between “Do-Re” and “Re-Mi” systems indicates something about the dynamics of realization.

17.2 Contour Levels (Reprise)

African tone systems use “speaker-specific normalized scales” (Ladd 2008a) that have been shown to change and transpose over time. Whatever tone levels are, thus far, they have failed to be formalized in the way we can define vowels in terms of formant space or scale degrees in terms of a network of superparticular ratios (intervals). Tone levels are very messy. Are contour levels a valid model for the scales formed by tone levels?
In Crowther’s orthography, initial high or low tones are marked, but repeated high and low tones are not. It makes sense that the initial move to the high level or low level is more marked productively (as Crowther’s method suggests) and may be how it is perceived, but is this reflected in the physical signal? In Figure 17.1, the rise from M to H (in the upper plot) is discrete, but the rise from L to H (in the lower plot) appears more like L to M then continues the ascent to another H, as if LHH is LMH. Thus, the change of level as markedness suggested by Crowther’s early (1852!) description is very sympathetic to the contour level formulation here, but the gradual rise confounds his description. An orthographic problem is how to indicate returning to M level if repeated H and L tones are unmarked. Laniran and Clements (2003) also note realization of LHHH as <0123> (in contour pitches). Phenomena like this confound the do-re-mi heuristic, frequency banding and contour level reduction. Rules, or alternatively constraints, are a method to explain such aberrations advanced by generative phonology and optimality theory respectively. Ajibóyè, Déchane, Gick and Pulleyblank (2007) give a thorough accounting of phonosyntactic rules for Yorùbá. Akinlabi’s forthcoming phonological grammar will likely also be a good source for this. Elsewhere, Pulleyblank has argued applying a constraints-based approach to Yorùbá tonology requires MAX (markedness) constraints but not IDENT (faithfulness) constraints (2004:417). Contour level analysis is more sympathetic to a constraints-based approach.

As our understanding of the tonology of Niger-Congo languages increases, models should incorporate the knowledge that tone production is not restricted to frequency bands and the tone perception does not rely heavily on interlevel magnitude

---

89 This has been reported as forthcoming from Harvard University Press since last year (2015).
(also suggested by Demany et al 2011 and Bidelman and Chung 2015). Contour level reduction of a signal is not challenged by paralinguistic modification, however tonal underspecification and phrase declination remain a problem.

To illustrate this, consider Figure 17.3, which previously appeared in Chapter 4 and Chapter 10. The plot includes three repetitions of the same phrase. Each consequent phrase is compressed and raised from its antecedent. In each, there is pitch reset, full initialization is delayed to the fifth syllable (the first L tone) and declination sets in almost immediately after initialization is completed. In Table 17.3, contour-level values are included calculated by using a “−1/+2” window (CL2 reduction).

![Figure 17.3: “L’órúkọ Jèsù ẹ́ ré́n a yọ́” repeated three times](image)

<table>
<thead>
<tr>
<th>Segment</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text</td>
<td>L’ó-</td>
<td>rú-</td>
<td>kọ</td>
<td>Jé-</td>
<td>sù</td>
<td>ẹ</td>
<td>ó</td>
<td>rẹ-</td>
<td>rín</td>
<td>a-</td>
<td>yọ</td>
</tr>
<tr>
<td>Speech Tone</td>
<td>High</td>
<td>High</td>
<td>Mid</td>
<td>High</td>
<td>Low</td>
<td>Mid</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>Mid</td>
<td>Low</td>
</tr>
<tr>
<td>Tone Level</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Implied Direction</td>
<td>0</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CL2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Mean Pitch (ST)</td>
<td>58.2</td>
<td>58.4</td>
<td>56.7</td>
<td>58.7</td>
<td>48.0</td>
<td>54.6</td>
<td>54.6</td>
<td>51.8</td>
<td>50.3</td>
<td>49.0</td>
<td>45.1</td>
</tr>
<tr>
<td>CL2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Mean Pitch (ST)</td>
<td>58.7</td>
<td>58.4</td>
<td>56.5</td>
<td>59.6</td>
<td>53.5</td>
<td>56.4</td>
<td>56.9</td>
<td>55.6</td>
<td>54.6</td>
<td>53.9</td>
<td>48.2</td>
</tr>
<tr>
<td>CL2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Mean Pitch (ST)</td>
<td>60.6</td>
<td>60.5</td>
<td>58.9</td>
<td>61.8</td>
<td>56.3</td>
<td>58.6</td>
<td>59.0</td>
<td>58.2</td>
<td>57.4</td>
<td>57.1</td>
<td>52.4</td>
</tr>
<tr>
<td>CL2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 17.3: Values for Figure 17.3
The strength of the CL$_2$ analysis is that each phrase produces the same reduction without prior segmentation. Furthermore, the CL$_2$ reductions for the realizations are nearly identical to the CL$_2$ reduction of the tone levels. With horizontal reduction before CL$_2$ analysis, they would be equivalent. The undercontextualized values, at the beginning (segment 1 of phrase 1) and end (segment 11 of phrase 3) of the excerpt, are indicated by italics. Based on context outside of the excerpt, these are the CL$_2$ values. CL$_2$ reduction could be a basis for segmentation through pattern-finding (as in Chapter 9). Furthermore, by adding one to values 1–3 and 7–9, the underlying tone sequence is produced, so the edit distance between the underlying tone levels and contour level reduction is not large. Contour-level space is not equivalent to underlying tone levels, but it reflects equivalence between underlying tone sequences. This needs to be tested further, but in analysis so far, realizations of equivalent tone sequences usually produce equivalent contour-level reductions.

Using a degree window of 2 with a radius of 1 around the focus produces three contour levels: minima (0), medials (1) and maxima (2). In contour level reduction, maximum and minimum values are extremes within their window. If a “−1/+1” window is used, these are equivalent to Thomassen’s pivots (1982). Ladd points out that tone is not so simple as turning point (2008a:134). It is with symmetrical windows larger than two degrees that more nuance is added, with maxima and minima that are not merely turning points but extreme within a larger window. Medial categories of super-minima and sub-maxima are added in a 5-level reduction and super-medial and sub-medial categories are added in a 7-level reduction.
Table 17.4: Relative height categories for 3-, 5- and 7-level reductions

<table>
<thead>
<tr>
<th>Level</th>
<th>3-levels (0–2)</th>
<th>5-levels (0–4)</th>
<th>7-levels (0–6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximal</td>
<td>2</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Sub-Maximal</td>
<td></td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Super-Medial</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Medial</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Sub-Medial</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Super-Minimal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimal</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

The extremeness of the contour level correlates with markedness or what Huron calls melodic accent.

17.2.1 Fuzzy-Gestalt Classification of Contour Level Equivalence

The CL2 reductions (as in Table 17.3) are based on the same information included in a contour adjacency series. In fact, a CL2 reduction can be produced from a CAS. Both use adjacent pairwise comparisons solely, but these comparisons are interpreted differently. Binary representation (C+ contour) of adjacent pairwise comparisons makes it possible to sum the comparisons associated with a single event, producing a metaparadigmatic value describing relative pitch height for each event. Events at the beginning or end of the signal are undercontextualized if the window is held in a constant orientation to the event and its referents. If this orientation is allowed to shift, non-adjacent comparisons may be used to give each event the same amount of context. Once the window includes non-adjacent comparisons, either through contextualization or using a larger window, a contour level series (CLS) is clearly distinguished from a contour adjacency series (CAS). However, a CLS does not include all possible comparisons within a phrase-order or lexical-order segmentation of the signal, as a contour pitch series.
(a CSEG) does. If comparing to melodic segments, a CLS may include as many degrees of adjacency as produce crisp values in a conventional \((n \times n)\) contour matrix. For the Bach subject and answers, this was 10 degrees \((-5/+5)\). In this way, the best categorization of a CLS is as a fuzzy contour equivalence, though it only uses crisp values.

Figure 17.4 includes five stratifications of melodic equivalence. P represents pitch equivalence, T is transpositional equivalence, C is contour pitch equivalence, L is contour level equivalence and S is syntagmatic direction equivalence (or contour adjacency series equivalence).

\[\text{Figure 17.4: Crisp, Fuzzy and Gestalt Equivalences}\]
Pitch equivalence (P) suggests that each event has the same absolute pitch height, rounded to some reasonable figure, perhaps semitones along the logarithmic pitch scale (C4=60). Transpositional equivalence (T) in the realm of western music comes in two shades: chromatic or diatonic. Chromatic transpositional equivalence (C) implies that between two melodic segments, S1 and S2, the same factor can be added to each element to get from S1 to S2. Diatonic transpositional equivalence implies that the factor is not the same in logarithmic frequency, but in terms of generic steps. In this way, it is more similar to the next strata, contour pitch equivalence, which ranks all events from low to high. Events with absolute pitch equivalence are given the same contour-pitch value which produces vertical gaps. The method is better-suited for music without recurring pitches within the segment. Contour-level equivalence is less stringent than contour pitch equivalence. In two contour-level equivalent segments, analogous events in each have the same relative pitch height within a window, not within the entire segment. Finally, syntagmatic direction equivalence implies that analogous adjacent pairs of events have the same direction of pitch change between them. In Figure 17.4, equivalence strata to the left are more rigid than equivalence strata to the right. If two melodic segments have P, they also have T, C, L and S. If two melodic segments have T, they also have C, L and S and so on. In the other direction, two segments with S equivalence will have considerable L similarity (depending on the number of degrees used), and two segments with L equivalence will have considerable C similarity, also depending on the number of degrees used for L equivalence. This similarity can be measured by CSIM or C+SIM using matrices. These similarity measures were degree-limited, indicated by a subscript, in Chapters 8 and 9. The strata of contour equivalence (C, L and S) do not necessarily
correspond to any amount of T or P similarity. Yet, I group C with T and P as crisp
equivalence because they all imply something about the relationship between the 1st
element to the nth element of the segment, L and S do not. However, L is grouped with P, 
T and C as constituting perceptually gestalt melodies if and only if L includes non-
adjacent comparisons (e.g. a four-degree symmetrical window). The subject and answers
in Bach’s c-minor Fugue (BWV 847) are gestalt not because they have S equivalence
(which they do) or P, T or C equivalence (which they do not), but because they have the
same L equivalence (see Figure 17.5).

Figure 17.5: a tonal answer of the C-minor Fugue subject with its CAS below

For example, the 4th, 10th, 15th and 20th elements are local minima within a symmetrical
window up to 10 degrees. The lack of P, T or even C equivalence does not disrupt this.
At the same time, S equivalence does not describe the similarity.
17.3 Disciplinary and Relativistic Comparisons

The disciplinary view is that language is one thing and music is another. A common, and to some extent valid, argument used to temper language and music analogies is that language and music rely on some of the same perceptual mechanisms, but are cognitively different. A relativistic perspective might argue cultures of sound rely on the same perceptual mechanisms, but are cognitively different. Consider two cultures: A and B.

<table>
<thead>
<tr>
<th></th>
<th>Culture A</th>
<th>Culture B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language</td>
<td>Lang-A</td>
<td>Lang-B</td>
</tr>
<tr>
<td>Music</td>
<td>Mus-A</td>
<td>Mus-B</td>
</tr>
</tbody>
</table>

Table 17.5: Sound categories in two cultures

Lang-A and Mus-A are found in Culture A and Lang-B and Mus-B are found in Culture B. A disciplinary comparative study would compare only language, or only music, across cultures. A relativistic study is interdisciplinary, emphasizing understanding of Mus-A is increased by studying Lang-A, and vice-versa. Through studying Lang-A and Mus-A we gain a more complete picture of Culture A. Both disciplinary and relativistic studies have the potential for hidden biases and insularity. Comparative musicology may strive for a level plane, but whether the researcher comes from Culture A, Culture B or from another culture altogether, the researcher’s language cannot help but tilt the plane of comparison. How one parses sound cannot help but make some features more apparent than others regardless of the nature of the signal. To a greater or lesser extent, each culture prioritizes sound dimensions and temporal scales differently.

According to Henrich, Heine and Norenzayan, there are no “obvious a priori grounds for claiming that a particular behavioral phenomenon is universal based on
The prioritization of sound dimensions in Western culture and Western culture’s role in globalization explain why linguistic documentation has focused on segmental phonemes (vowels and consonants), and musical transcription on notes. Focus on phones and notes is Eurocentric, but not without merits. Phones are important in most languages, even in whistle languages (as found in the Canary Islands) which do not have consonants but do have vowels in the use the second formant (Meyer and Gautheron 2006). But there are also speech surrogates, such as drum languages, where phones have no presence (Stern 1957). Many linguists have questioned Chomsky’s notion of a universal grammar. If anything, empirical cross-cultural study of language reveals tendencies not universal structure (Evans and Levinson 2009). In music research, we are appreciating more and more the impact of other features, particularly rhythm and timbre. I had a reawakening to the sound world of non-musicians and non-tone language speakers recently when I gave a presentation to a general audience and received a question that revealed the person did not know the difference between loudness and pitch. Apparently, the assumption I made that my audience would be privy to a high-low conception of pitch (if nothing more) was errant. Musical preference in media-saturated cultures may have more to do with timbral and rhythmic characteristics, which often define boundaries between musical genres than melody or harmony. WEIRD cultures and societies have fooled us, so have WEIRD languages (Majid and Levinson 2010), so has WEIRD music.

A WEIRD vision of the world has often been counteracted with other parochial lenses. As an area studies specialist there is potential for digging so far into a culture that it is hard to see outside of one’s perspective. Rejection of outsider views of a culture by
insiders causes stagnation, as does a failure to consider how one’s discoveries relate to broader humanistic questions. As we move forward, we must consider the types of cross-disciplinary and cross-cultural comparisons that have been made, and those yet to happen.

Comparative study, both disciplinary and relativistic, has been happening for millennia. Classical thought, by Aristotle and others, often drew comparisons between music and language. Such analogies are popular notions in many societies, certainly Niger-Congo cultures. The bi-directional nature of cross-disciplinary relativistic thought has not always been shared by cross-cultural disciplinary research, which has tended to see the language or music of the other in terms of the language or music of one’s own culture. Kolinski’s comparative analyses of contour (see Chapter 16) were highly original and important contributions but suffered from unqualified evaluations. A Kwakiutl (Amerindian) melody was “smooth” while a Dahomean (Niger-Congo) melody was “bold and ragged” (1965:114–6). This assessment would have been well-tempered if Kolinski (1) examined other aspects of the cultures and of the music itself (such as the lyrics and language of the songs); and (2) laid bare his own aesthetic biases and considered how that influenced his reaction to the melodies. This problem in comparative musicology has been overcompensated by antiformalism in ethnographic musicology.

Autosegmental phonology, emerging in the 1970s, is a great example of knowledge about another culture impacting the study of one’s own culture. The tone levels of Niger-Congo languages inspired a new phonological method for all languages. This was not only a major step forward in linguistic theory but generally in cross-cultural
disciplinary studies of any sort. Similarly, documentation of non-western musical scales has informed scale theory, a subfield of music theory.

Figure 17.6: Cross-disciplinary and cross-cultural research

Drawing on cross-disciplinary knowledge has also shaped cross-cultural studies. The easier path for such an endeavor is from well-tread cultures (represented by Culture A in Figure 17.6) to lesser-known cultures (represented by Culture B). The green arrows in Figure 17.6 represent using knowledge about the Language and Music of Culture A to study either the language or music of another culture. In some respects, this is the only way to conduct research on another culture’s music until one knows the language. In linguistics, researchers have also combined models from music theory with linguistic
models to study non-western languages. As already noted, features of a Lang-B have been used for new models of a Lang-A, but it rarely happens that a composite picture of sound in Culture B is used to study Culture A.

If Culture A and Culture B are Western Europe and a non-Western culture (such as the Niger-Congo A family), studies that are both cross-disciplinary and cross-cultural have been rare. But they have become more common if Culture A and Culture B are a bit closer together. Patel and Daniele (2003) make an empirical comparison of rhythm in English and French speech and music. There are many other comparisons that could be made in terms of cultures and domains of sound, including broader groupings of cultures. Research that makes a complex of disciplinary and relativistic comparisons may be very fruitful. There is always potential to draw inappropriate conclusions from any line of inquiry. Diana Deutsch’s research on absolute pitch perception and tone languages is an example. But there is also great potential for collating cross-disciplinary and cross-cultural comparisons in studies like Patel and Daniele (2003).

This dissertation applied an understanding of Lang-B (Niger-Congo tone languages) to the analysis of Mus-A (Western European and North American music). What does Lang-B have to do with Mus-A? If the former is Yorùbá and the latter Bach’s C-minor Fugue, the answer is not much. One could even question the relevance of studying Niger-Congo cultures to understanding African-American Vernacular English (AAVE) or Black American music (e.g. blues, jazz, spirituals). In the case of Afro-Brazilian or Afro-Cuban culture, Yorùbá language, religion and music have a very tangible presence, but the presence of African cultures is not always so tangible in North American culture. Can AAVE be called Lang-A/B or jazz called Mus-A/B even after
Lang-B disappears? My answer is yes. Does the Yorùbá language need to have a clear connection to Bach’s music for an understanding of one to aid in the understanding of the other? My answer is no, there does not have to be any historical or ethnographic connection. To the extent that there is overlap between the language and music within a culture, there are also key differences (as many have pointed out). Language may monopolize a contrastive temporality leaving other temporalities to be used expressively in speech and music. This makes the comparative study of a Lang-A and a Mus-B or a Lang-B and a Mus-A quite useful. West African drumming might be compared with stress-timing in English, or as done here, African tone systems may inform analysis of melodic contour in Western European and North American music. Applying a model like African tone systems to Western music does not necessarily imply any universals or underlying motivations to sound organization other than that we all have ears and minds to listen and voices and hands to make sounds.

Emphasizing research that is cross-disciplinary and cross-cultural smacks more than a bit of affirmative action in scholarship. Are such endeavors for the progress of human knowledge or are they intellectual reparations, or both? The benefits lie so strongly in the first, in the increasing of human knowledge, that further justification in terms of fairness may not need to be emphasized (to those who have deaf ears). This is why (in my mind) Kofi Agawu’s 2015 keynote to the Society for Music Theory was titled “Rethinking music theory with African aid” not “Rethinking music theory to be more inclusive.” Does it benefit a society to privilege the same paradigm repeatedly in education and research? I do not think it does.
BIBLIOGRAPHY

*Iwe Orin Mimo Fun Ijo Enia Olorun Ni Ilẹ Yoruba.* Aberdeen, UK: Church of the Province of West Africa / The University Press, 1923.


Boynton, Susan. “Orality, Literacy, and the Early Notation of the Office Hymns.” 


Brown, Nyssa, and Deborah Lamb. “Parallels between Music Learning and Language Acquisition: From Fluency to Literacy.” American Council on Immersion Education (ACIE),


Carter-Cohn, Aaron, and Lazarus Èkwúèmé. Interview, November 15 2011.


513


Miller, George A. “The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information.” *Psychological Review* 63, no. 2 (1956): 81.


Oreoluwa, Ogunbowale Mopelolade. “‘In the Ghetto, Life No Easy for We’: The Construction and Negotiation of Identity in Ajegunle Raga.” The University of Guelph, 2012.


-----. *Psychopy - Psychology Software for Python, Release 1.80.04.* Nottingham, UK: University of Nottingham, 2014.


Ross, Will. “Nigeria's Thriving Art and Music Scene.” BBC.


527


Tovey, Donald Francis. *The Forms of Music: Musical Articles from the Encyclopaedia Britannica*: Oxford University Press, 1957.


Williams, Rowan. “General Synod Eucharist.” youtube.com.


