COGNITIVE CONNECTIONS BETWEEN LINGUISTIC AND MUSICAL SYNTAX

AN OPTIMALITY THEORETIC APPROACH

by

LAURA BREWER

(Under the Direction of Keith Langston)

ABSTRACT

Language and music share empirically validated syntactic similarities that suggest these modalities may have important cognitive connections. These findings challenge the current view of the human language faculty as encapsulated from other cognitive domains. Patel (2008) emphasizes the distinction between the representational knowledge base of a domain (competence) and the syntactic processes that are employed during access and integration of this knowledge, proposing that the processing resources are shared between the two domains while the knowledge bases remain separate. I propose that the shared processing resources for linguistic and musical syntax are characterized predominantly by the constraint evaluation and optimization mechanisms of Optimality Theory. The goal of this endeavor will be to explore the OT character of musical syntax in support of the claim that a unified theory can reflect cognitive overlap in the language and music modalities and to encourage future refinement of this theoretical framework.

INDEX WORDS: Optimality Theory (OT), syntax, Western tonal music. modularity, weighted constraint, Generative Theory of Tonal Music (GTTM), competence, performance, processing

COGNITIVE CONNECTIONS BETWEEN LINGUISTIC AND MUSICAL SYNTAX

AN OPTIMALITY THEORETIC APPROACH

by

LAURA BREWER

B.A., The University of Georgia, 2007

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment

of the Requirements for the Degree

MASTER OF ARTS

ATHENS, GA

2014

COGNITIVE CONNECTIONS BETWEEN LINGUISTIC AND MUSICAL SYNTAX

AN OPTIMALITY THEORETIC APPROACH

by

LAURA BREWER

Major Professor:		Keith Langston
Committee:		Jonathan Evans
		Jared Klein

Electronic Version Approved:

Julie Coffield
Interim Dean of the Graduate School
The University of Georgia
December 2014

ACKNOWLEDGEMENTS

I am extremely grateful to my committee, Dr. Keith Langston, Dr. Jared Klein, and Dr. Jonathan Evans, for taking a chance on an interdisciplinary topic. I owe a special 'thank you' to my advisor, Dr. Langston, for never giving up on me, even when I thought I had hit a dead end. Without your insightful guidance and enthusiastic support, I could not have pursued such a challenging and interesting topic.

I would also like to thank two individuals who happened into my life by mere circumstance but have had an enormous impact on my writing experience: Dr. Nancy Thompson and Dr. Sally Kennedy. Your encouragement has meant more to me than you will ever know.

I am deeply appreciative of my friends and family who have tirelessly provided support and encouragement throughout this process. I am especially grateful to Whitney Holley, Cathy Kilroe-Smith, Paul Kim, Meaghan Nappo, and Kate Pavich for believing in me and helping me to push through some particularly challenging times. Your company, humor, and perspective have been invaluable.

My deepest gratitude is to my family: Mom, Dad, Patrick, Britney, and Connor. Without your unfailing spiritual and emotional support, I never would have made it.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1:

INTRODUCTION

1.1 INTRODUCTION

Researchers have been fascinated by analogies between language and music for hundreds of years, and

many believe language and music are connected in the mind. Proponents of the hypothesis that language

and music have cognitive connections cite several significant similarities between these domains.

Language and music are among only a few cognitive domains that are uniquely human; some researchers

have even hypothesized that language and music evolved from the same primitive cognitive capacity

(e.g., Brown, 2000; Darwin, 1871; Fitch, 2002). Language and music also both rely on the auditory

domain, specifically the production and perception of organized sound, for the purpose of communication

and expression. While music may not be capable of expressing meaningful propositions with the same

precision and fecundity as language, these modalities share the property of being primarily

communicative, social behaviors that foster connection and cohesion among humans by facilitating

mutual understanding.

Perhaps the most interesting comparison between language and music is that they both rely on

rich syntactic cognitive systems. The human language and music modalities embody remarkably similar

capacities for combining acoustically discrete segments in complex and meaningful ways. We are born

without any preference for or ability in a specific language or musical system, but virtually all humans are

able to acquire implicit knowledge of a language or musical system fairly easily, not necessarily through

explicit instruction, but merely through exposure. Numerous studies (e.g., Trainor & Trehub, 1994;

Dowling, 1982; Sloboda, 1985), show that the syntactic nuances of a musical system are acquired in

stages throughout childhood similarly to language. Given the syntactic similarities between language and

music, it is reasonable to suspect that these modalities share cognitive connections that run deeper than

just auditory perception and processing. Like language, members of a culture that are exposed to a musical system seem to become "fluent" in it, even if they are not skilled performers.

What does it mean to be "fluent" in a musical system? Sloboda (2005) outlines several criteria, in alignment with conventional psychological methodology, that indicate that an individual can "make sense" of the syntactic structures of music in a way analogous to fluency in a language. The first criterion is that an individual can better remember music that conforms to the structure of his or her own musical system than music that follows a system with which he or she is unfamiliar. Secondly, when recalling a musical sequence, individuals who have "made sense" of the music seem to have stored something more abstract than the note-for-note memory of the sequence, and they often make plausible rather than random or surface-level substitutions. This phenomenon is akin to language recall, in which individuals may not remember information word for word but do remember the meaning of what they have heard. The third and fourth criteria refer to an individual's abilities to correctly judge a musical sequence as acceptable according to implicit syntactic rules and to correctly identify the mood or emotion of a musical passage, both of which begin to come into place for even musically-untrained children between the ages of 5 and 10 (Sloboda and Pinchot-Kastner & Crowder, as cited in Sloboda, 2005: 265-6). Reimer and Evans also eloquently analogize language and the richly syntactic musical system of Western culture (often referred to as "tonal music"):

> The listener need not know the functional technicalities, for they are so much a part of our lives, like the language we learn as babies, that they can be regarded as given, as one of the basic ways our minds operate. Tonal music makes sense to us because we have internalized the system. Particular pieces of tonal music may be more or less understandable, as particular books in English are more or less understandable, depending on their complexity and the experience of the reader. But the basis for understanding is present, in tonal music or in a book of English; neither requires that a new language be learned. (1972: 76)

Despite the alluring similarities between language and music, serious theoretical comparisons of the two domains have been met with considerable skepticism. Lerdahl and Jackendoff caution the researcher against focusing on superficial analogies between language and music, calling this "an old and largely futile game" (1983: 5). Pinker also points to a cognitive rift between the two domains, stating, "Compared with language, vision, social reasoning, and physical know-how, music could vanish from our

species and the rest of our lifestyle would be virtually unchanged" (1997: 3). Among such criticisms is the argument that musical and linguistic knowledge are not even comparable, because musical knowledge is not ubiquitous but is rather only possessed by those who have received explicit musical training. Further resistance to these comparisons is rooted in the claim heralded by leading linguists (e.g., Fodor, 1983; Hauser, Chomsky & Fitch, 2002) that the core cognitive mechanisms implicated in linguistic behavior are domain-specific and isolated from the rest of the mind or brain.

While it is true that erroneous comparisons of language and music should be avoided, this paper will investigate the syntactic similarities between these modalities and the implications these similarities have for the integrity of linguistic theory. I will argue that evidence for cognitive overlap between these domains cannot be ignored and should be appropriately integrated into a model of syntax, and I will propose a way of modeling this overlap using a contemporary linguistic theory. The discussions and analyses presented here will focus on the musical system of Western tonal music, although ideally any conclusions made about music as a cognitive modality should be applicable to any musical system. The decision to focus on Western tonal music is made here both out of convenience and necessity, as it will allow me to better incorporate my own musical training and instincts, as well as those of the intended reader.

1.2 SUMMARY

The purpose of this paper is to contribute to the discussion of how language and music might be cognitively connected by exploring whether a unified theoretical model (Optimality Theory) can accommodate both linguistic and musical syntactic phenomena. The paper will begin with a more detailed description of the major structural similarities between language and music that point to a possible connection between linguistic and musical syntax. This possible connection will be substantiated in the following section, which will present a summary of evidence from neuropsychological research suggesting overlap in the parts of the brain and cognitive resources used for specifically syntactic linguistic and musical tasks. I will then discuss the implications of such an overlap for our theoretical

approaches to linguistic and musical syntax, making the argument that a theoretical model should be consonant with empirical evidence. The remainder of the paper will explore how a current model for linguistic phenomena, Optimality Theory, may be expanded to account for musical syntactic phenomena as well. The goals of this endeavor will be to support the claim that one unified theory can serve as a characterization of the cognitive overlap across the language and music modalities and to encourage future refinement of this theoretical framework.

CHAPTER 2:

EVIDENCE FOR COGNITIVE CONNECTIONS BETWEEN LANGUAGE AND MUSIC

2.1   INTRODUCTION

There are many reasons comparisons between language and music constitute an attractive area of inquiry,

but few are more alluring to the cognitive scientist than connections between linguistic and musical

syntax. This chapter will discuss some of the justifications for the claim that linguistic and musical syntax

have more significant cognitive connections than is generally thought. Specifically the goal will be to

reveal that our mental representations of music, like language, involve more abstract syntactic principles

than simply the perception of acoustic signals. The first of these justifications concerns the following

structural similarities between these two domains: 1) Language and music are both built from discrete and

arbitrary units of sound; 2) These units take on meaning when they are combined in principled and rule-

based ways; and 3) These principles of combination are applied to hierarchical levels, both abstractly and

in actual utterances/compositions. In exploring each of these points of similarity, this chapter will briefly

touch upon some of the fundamental concepts of Western tonal music that will be implicated in the

discussions and analyses to follow. The second form of justification for cognitive connections between

language and music will be a review of neuropsychological evidence that reveals unprecedented neural

and cognitive overlap between linguistic and musical syntax.

2.2   STRUCTURAL SIMILARITIES BETWEEN LANGUAGE AND MUSIC

2.2.1   Discrete and Arbitrary Units of Sound

It is well-known that every language uses only a subset of all the sounds that can possibly be produced by

the human articulatory mechanisms, and these subsets differ from language to language. Every language

comprises an inventory of phonemes that represent the smallest perceptual units of sound in that

language. These units of sound are intrinsically arbitrary and meaningless. For example, the word *cat*

contains the sounds [k], [æ], and [t], but none of these sounds independently contributes any meaning to

the word *cat;* if they did, we might expect *cat* and *cap* to have very similar meanings since the words

share two out of three component sounds. The influential linguist Ferdinand de Saussure described this

arbitrary sound-meaning relationship in language as indicative of a semiological, or sign-based, system

(Holdcroft, 1991). According to Saussure, a sign contains two essential and mutually exclusive parts: the

signifier, which represents an object or concept without actually *being* that object or concept, and the

signified, which is the actual object or concept that is represented. Without an object or concept to be

signified, a signifier has no inherent meaning and is unmotivated to enter into the signifier-signified

relationship.[1] It is arguable that at some point in the development of a word and its incorporation into a

language, the connection between the word and its meaning loses its essential arbitrariness, because the

word and its meaning become so integrated that the existence of one without the other seems impossible.

Lévi-Strauss (1958/1963) makes the distinction that a sign is arbitrary *a priori,* but is not arbitrary *a*

posteriori.

Like language, musical systems are built from an inventory of sound units. Musical systems

almost universally build these inventories by systematically dividing the entire continuum of sound

frequencies into discrete units, called notes. The collection of notes available in a musical system is

conventionally referred to as a *scale*.[2] One notable feature about the concept of a scale is that, in most

musical systems, the precise acoustic frequencies of the notes in a scale are not as important as the *ratios*

between those frequencies, called *intervals*.[3] For example, the "Happy Birthday" song is not always sung

beginning on any specific pitch, but most individuals automatically maintain the proper interval distance

---

[1] Exceptions to the notion that a signifier has no inherent meaning and is arbitrary may include onomatopoeia and iconic language.

[2] The only musical idioms that do not seem to draw upon a repertoire of discrete intervallic relationships are found in primitive, tribal cultures and are few and far between (Sachs, 1962, Malm, 1967, and McPhee, 1966, as cited in Burns & Ward, 1982).

[3] It is the *ratios* that remain constant, rather than the exact distance in acoustic frequency between the notes because the change in frequency throughout the sound continuum is nonlinear.

between each of the sung pitches. In this sense, the precise acoustic qualities of a note are essentially arbitrary in that they do not independently contribute to the musical structure; instead, the characteristic quality of Western tonal music emerges from the relationships between the pitches.

The intervals (and resulting notes) that characterize Western tonal music are created by dividing the octave into 12 equal segments. In most musical systems, the octave is recognized as the most natural and fundamental division of the sound continuum, because frequencies of pitches that are an octave apart are related by a factor of two: A pitch one octave higher than another pitch has a frequency that is exactly double that of the original pitch. This simple mathematical relationship between octave frequencies makes octave pitches sound exceptionally similar to the human ear and mind, as if they are the same pitch in many respects but differ only in the sonic quality of pitch height. The division of the octave into 12 equal intervals (called *semitones* or *chromatic steps*) produces 12 notes, which comprise what is conventionally called the *chromatic scale*. While all of the notes of a chromatic scale are available to be used in Western tonal music, a smaller subset of notes, called the diatonic scale, is selected from the chromatic scale to form the primary collection of notes used in musical composition. The diatonic scale comprises seven notes that are successively either one or two chromatic steps apart in the chromatic scale, following a pattern of 2, 2, 1, 2, 2, 2, 1:[4]
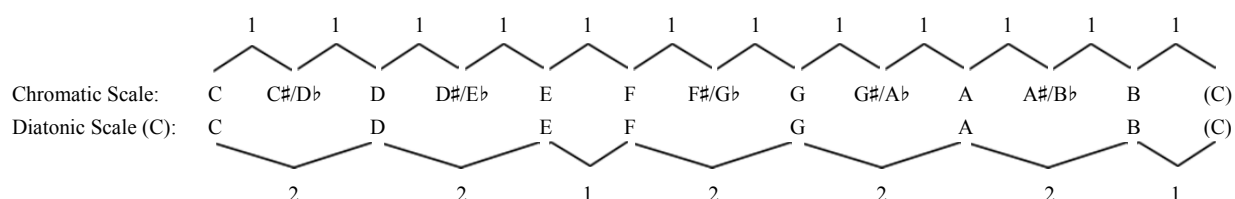
Figure 1: Chromatic and diatonic scales (beginning on C)

---

[4] This pattern corresponds to the *major scale*. Another type of diatonic scale, the natural *minor scale*, uses the same pattern of half and whole steps but begins at a different point in the pattern. Still other variations of the minor scale use different patterns. References to diatonic scale in this paper will always pertain to the major scale for analytical simplicity and uniformity.

Of all the domains of human cognition, language and music are the only ones that are constructed of minimal, basic acoustic segments that are selected from the entire range of possible human- or instrument-created sound. While it is clear that language and music use categorically different inventories of sound as their basic building blocks, the inherent arbitrariness of their compositional components is just one of many facets of this point of similarity between these two domains. For instance, research shows that listeners of Western tonal music, like the speakers of a language, engage in *categorical perception*, automatically and unconsciously attempting to hear musical stimuli as fitting into the diatonic/chromatic system even if the ratios between notes do not precisely fit the Western tonal schematic (Ball, 2010; Burns & Ward, 1982; Patel, 2008). The use of scales as the fundamental organizational framework for tonal music can also lead to such enculturating in this musical system that listeners familiar with tonal music often hear music that uses other scales or tunings as sounding inherently wrong or odd.

2.2.2 <u>Principled and Rule-Based Combinations of Constituents</u>

At the core of what allows large groups of people to create and understand an infinite number of utterances in a language is the notion that the combination of linguistic units is governed by systematic, rule-based principles. Like the diatonic and chromatic scales used in Western tonal music, the inventory of words available to a language user is fairly static; most of the words we use have been established in the lexicon with meanings that are generally agreed upon among fellow language-users. The order and grouping of those words, however, is much more variable, and we often create sentences that have never been uttered before yet are easily understood by those who know the language. Of course the same is true for music; new compositions are written every day using essentially the same set of notes that has been used for hundreds of years. Novel utterances and musical compositions are possible in part because meaning in each of these domains is closely tied to the variable word order of their constituents. Consider how different the utterance "The man with the tall hat saw the girl" is from "The tall girl with the hat saw the man." Likewise, in the popular nursery tune "Twinkle, Twinkle Little Star," consider how a child

might respond if the second half of the song ("how I wonder what you are") were changed to ascend from C to D to E to F instead of descending from F to E to D to C.

Another manifestation of the complex principles governing the combination of linguistic constituents is the syntactic relationships that obtain between words as determined by their context. Words can take on grammatical functions, such as subject or direct object, that are not inherent to the words themselves but emerge as a result of their order in a clause and other syntactic (and morphological) cues. Similarly, musical elements also take on contextual meaning and function based on syntactic combinatorial principles. The creation of a mutually-understandable system of tonal music requires more than just selecting notes from an inventory of sounds; these notes must also be combined and organized in ways that are consistent with the rules and norms of the musical system. In this sense, "rules" are not necessarily prescribed instructions that composers consciously strive to follow; rather, they are syntactic principles that constrain the organization of notes into a pattern that can be interpreted by those with an implicit knowledge of the musical system. Like language, this implicit knowledge of Western tonality's combinatorial principles is developed merely through exposure to the idiom, not by explicit training or instruction. This phenomenon is observable even in young children, who may scrunch up their face or laugh when they hear a chord progression with a syntactic violation (Sloboda, 1985), or when an individual without any musical training presses clusters of keys on the piano and knows it does not sound like music.

Chapter 4 will discuss in more detail the syntactic functions at work in Western tonal music, but it is important to introduce a few concepts here. Combinatorial principles in Western tonal music apply to both melody and harmony, concepts that are theoretically distinct but in actuality are so interlaced that they are often inseparable. *Melody* refers to a single line of music in which notes are linearly concatenated and heard as a single entity. While other notes may be sounding simultaneously with the melody, the melody is perceived more or less as an independent entity. The melody may be a line of music that is singable or easy to remember (sometimes to the point of becoming "stuck" in one's head). *Harmony*, on the other hand, refers to the sonorities that are created when multiple notes are sounded simultaneously.

Harmonic progression, or the movement from one harmonic sonority to another (effected by changing

some or all of the notes that are sounded at any given moment), is highly constrained and guided by a

number of principles of combinatorial well-formedness.

Perhaps the more richly syntactic combinatorial principles that guide musical intuitions in

Western tonal music involve how notes may combine vertically to create harmony. Acceptable harmonies

in Western tonal music are built by stacking two, three, or sometimes four notes in intervals of a third. A

third is an interval spanning three or four chromatic steps in the chromatic scale so that the selected notes

are separated by one note in the diatonic scale. Since the basic form of these harmonies includes three

notes, stacked in thirds, these harmonies are called *triads*. Figure 2 shows the triads, one built on each

scale degree of the diatonic scale, that comprise the allowable harmonies in Western tonal music:
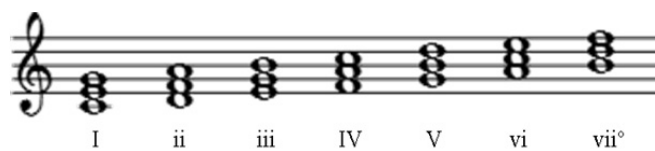
Figure 2: Chords built on each note of the C diatonic scale

The triads in Figure 2 are built on the notes of the C scale, or the diatonic scale that begins the interval

pattern of 2 2 1 2 2 2 1 beginning on C. A new set of chords can be formed by selecting a different

diatonic scale and building triads on each of its notes. The note that begins the scale is referred to as the

tonic, and the chord built on the tonic is called the *tonic chord.* Any note of the chromatic scale can serve

as the tonic note for a key by building the diatonic scale and its chords beginning on that note.

Adherence to one diatonic scale and its harmonies contributes to a listener's concept of the

musical *key*. A key is often described as a "tonal region," because once it is established, the listener has a

sense of orientation to a tonic note (and chord) and the other notes and chords in relation to it. The

establishment of a key is achieved through adherence to a diatonic collection and the use of functional

(i.e., constrained and not random) harmonic progressions. Tonal music feels like it is constantly moving

10

away from and being pulled back to the tonic, and a musical composition may leave one key altogether

and enter another one by shifting the entire collection of notes and harmonies from one diatonic collection

to another, called *modulation*. While modulations are common,[5] the principles underlying a well-formed

key change are highly constrained and rigorously taught in music theory and composition classes. Most

notably, modulations often happen between keys that are perceived to be *nearby* rather than *distant*.

Nearby keys are those that share many common tones. For example, they key of C contains the notes C,

D, E, F, G, A, and B, so it shares several common tones with the key of G (which contains G, A, B, C, D,

E, and F-sharp, differing by only one note), but very few common tones with the key of B major (which

contains B, C-sharp, D-sharp, E, F-sharp, G-sharp, and A-sharp, differing by five notes).

The adherence to a system of harmonic relationships that help establish key and constitute our

knowledge of tonality is often referred to as "functional harmony." Functional harmony is sensitive to and

achieved through a number of factors, including the function and position of the notes in the context of

the diatonic scale, the resulting quality and character of the specific combination of notes in a chord

(commonly referred to as a *sonority*), and the function and character of a specific sonority within the

context of a larger chord sequence. What is notable about functional harmony is that, while its principles

are far too complex and multidimensional to attribute to template learning, they are also easily and readily

acquired as implicit knowledge by young children who have adequate exposure to the tonal idiom without

necessitating formal training or instruction. This paradox, often referred to in the context of language

acquisition as "poverty of the stimulus," indicates that this knowledge of the tonal system is highly

cognitive and syntactic in nature.

2.2.3   Hierarchical Organization

In addition to word order and other combinatorial principles, language and music are characterized by the

organization of elements into higher and more abstract levels of constituency. One manifestation of this

hierarchy in language is the grouping of sounds to make morphemes, morphemes to make words, words

---

[5] For example, sonata form requires a modulation to a specific key (or keys).

11

to make phrases, phrases to make sentences, and so on. Additionally, constituents at each of these levels may enter into complex and sometimes long-distance hierarchical relationships with each other that, while not overtly discernible in the linear structure of speech or text, may be subconsciously intuited by the speaker or listener. For example, in the sentence *That girl who kicked the brown dog has hurt her foot*, lower levels of the hierarchy might group together nouns and their determiners, creating constituents out of [that girl], [the brown dog], and [her foot], each of which is embedded in higher levels of organization. A higher level of organization would create a group out of [that girl who kicked the brown dog], since this entire phrase is the subject of the sentence, and another group out of [has hurt her foot], since this group contains all of the information about what happened to the subject. Another type of syntactic hierarchy involves relationships between words that are not necessarily side-by-side but do exhibit dependency. In our example sentence, a higher-level dependency between *girl* and *hurt* informs the hearer of this sentence that, although the sentence contains the string of words *the brown dog has hurt her foot*, it was the girl whose foot was hurt and not the dog's. The formal representations used by traditional generative grammar use tree structures to model the hierarchies of constituent grouping and dependencies, as exhibited in Figure 3:
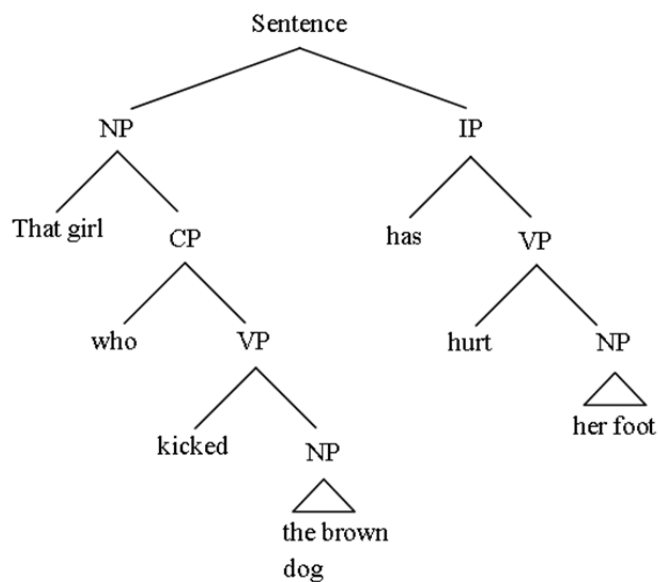
Figure 3: Tree representation of *That girl who kicked the brown dog has hurt her foot.*

Hierarchical structure in music is central to perceptions of meaning and patterns of tension and relaxation, and this hierarchical structure is manifested in music in essentially two separate but interrelated ways. The first type of hierarchical structure, which Patel refers to as *pitch hierarchy*, comprises an "atemporal schema derived from experience with musical melodic patterns" (2008: 201).[6] Pitch hierarchies reside in the abstract knowledge and understanding of the relationships between sonorities that are part of and can be built from the diatonic scale. These hierarchical relationships exist both among the individual notes of the diatonic scale and among the chordal sonorities that are formed from combinations of notes to create functional harmony. The second type of hierarchical structure, *event hierarchy*, pertains to the actual structure of a temporal musical sequence, both the melodic and, perhaps more importantly, harmonic successions of sonorities.

At one level of organization, atemporal pitch hierarchies involve the relationships among the individual notes of the diatonic scale, specifically their relationship to the first note of the scale, the tonic. As was previously discussed, the tonic is the most perceptually stable note, acting as the "center of gravity" for the entire scale and providing an abstract sense of a tonal center. The next most stable notes are the fifth and third notes of the diatonic scale, which, together with the tonic, make up the most stable diatonic chord, the tonic triad. The fourth and sixth notes of the scale fall somewhat lower in the hierarchy of stability, and the least stable notes of the diatonic scale are those that are on either side of the tonic note: the second note of the scale and the seventh note of the scale. Finally, the notes that are perceived as the least stable of all are those that are in the chromatic scale but are *not* part of the diatonic scale. The perceptual salience of this hierarchy of stability is substantiated by an experiment by Krumhansl and Kessler (1982), in which participants listened to a harmonically functional sequence, followed by a brief silence, and then a note (the "probe tone") from the chromatic scale. Participants were asked to rate how well the probe tone fit with the musical sequence they had just heard, and these judgments, shown in Figure 4, coincided with the hierarchy of notes described above:

---

[6] It is unclear why Patel omits chordal hierarchies from this category, but I will include them.

13
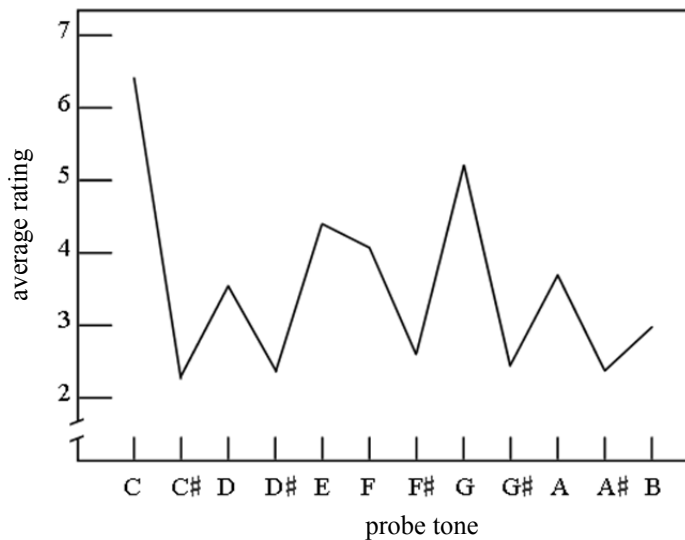
Figure 4: Profile of probe tone relatedness judgments in a major key (C Major) (Krumhansl & Kessler, 1982: 343)

Atemporal pitch hierarchies also exist as harmonic relationships among the chords that can be formed from the notes of the diatonic scale. The seven chords that serve as the basic inventory of harmonic material in Western tonal music have been presented in section 2.2.2. Like the individual pitches of the diatonic scale, these chords are in a hierarchy of stability in relation to the most stable harmony, the tonic chord (which is the triad built on the first degree of diatonic scale). Music theorists, supported by rigorous empirical evidence (e.g. Bharucha & Krumhansl, 1983; Krumhansl, Barucha, & Kessler, 1982), give the following stability hierarchy for major-key harmonies in Western tonal music: The tonic chord is most stable, followed in order by the V, IV, vi, ii, iii and vii° chords. Chords that are not part of this basic set are sometimes used but are highly unstable. It is important to note that knowledge of these chords is such a fundamental part of our musical listening that all of the notes of a chord do not even need to be present in the music for a listener to interpret their structural significance. Although the details of this phenomenon, often referred to as *implied harmony*, are not entirely understood from a cognitive perspective, the interrelatedness of melodic and harmonic structure is a central component of music theory.

Event hierarchies involve the actual, temporal instantiation of a musical sequence and the extent to which the different sonorities in the sequence provide structure to the musical phrase. A musical phrase is intuitively perceived as having sonorities that are more or less important to its structural identity, and these sonorities enter into a perceptually salient hierarchy of importance that exists abstractly outside of the linear succession of musical events. Notes that are low in the structural hierarchy may be ornamental, and depending on their position in the hierarchy, they may be easily changed or omitted without disturbing the integrity of the music. As an example of a common and well-known practice that exploits the structural event hierarchy, consider the many performance variations of "The Star Spangled Banner". Non-structural notes may and often are added, omitted, or changed to create artistic flair, but the core structural skeleton must remain intact in order for the music to be recognizable. Even notes that are part of the basic melody only have semi-structural importance, for instance, the second note of the opening "Oh…", or "the" in the phrase "By the dawn's early light," might be omitted without the same degree of eyebrow-raising that might follow from omitting the note on "say" or "see."

The perceptual reality of a structural hierarchy in Western tonal music has been a central topic in recent music theory. Heinrich Schenker is a music theorist recognized for introducing analytical techniques in the early 20th century that were helpful in understanding and depicting the prolongational and structural hierarchical relationships in musical compositions. Schenkerian analysis uses an adapted system of musical notation in graphic representations of how notes that may not be right next to each other are heard as being on the same structural level and therefore may be perceived as forming a higher-level melody or harmony, providing complexity to the musical experience. Figure 5 shows an example of this type of grouping in the first phrase of a familiar tune, "Greensleeves". The beams that extend upward above the staff (a notation that would never be used in a regular musical score but here serves the purpose of demonstrating the structural level) indicate that the descending line from A to G to F to E may be perceived as forming a melodic group that forms the skeleton of the phrase. The notes that are not included in this melodic grouping, while important for melody recognition, occupy a lower level of

structural importance and are essentially embellishments of the underlying harmonies implied by the melodic structure delineated here:
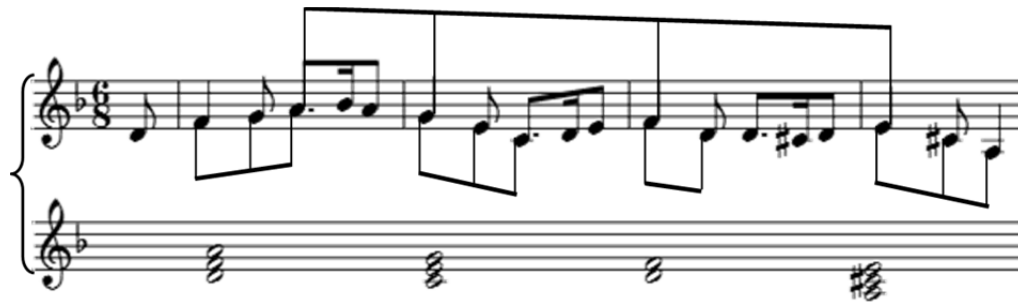
Figure 5: Schenkerian notation of "Greensleeves" (Traditional; Cadwallader & Gagne, 1998: 23)

In developing a Schenkerian analysis of a musical composition, this process of identifying levels of structural importance becomes increasingly complex, because longer, more intricate compositions will inevitably comprise more and more tiers. For reasons of both analytic efficacy and practicality, entire compositions are reduced to only the top several levels of structural importance, so that entire phrases of music may be represented in a graphic reduction as a single note.

While the notation of Schenkerian analysis may be somewhat esoteric, a musician-linguist duo by the names of Fred Lerdahl and Ray Jackendoff recognized the abstract, hierarchical organization of Western tonal music as strikingly similar to that of language. In *A Generative Theory of Tonal Music* (1983), the researchers acknowledge that this hierarchical organization is effected through several different components of structure, all of which interact in a complex system that can be successfully analyzed using the tools of the linguistic theoretical approach of generative grammar. GTTM describes the hierarchies of grouping structure, in which successions of tones are perceived as belonging to linear groups according to gestalt-type principles, and metrical structure, which refers to layers of organization of the underlying rhythmic beat patterns that help drive the music. Grouping and metrical structures are then analyzed as influencing the time-span reduction, which describes the hierarchy of structural importance of musical events, and the prolongational reduction, which describes the hierarchical patterns

of tension and relaxation created by the use of sonorities of varying stability throughout a piece of music. Figure 6 shows a time-span reduction in which the height of a chord's branch corresponds to the degree of the chord's structural importance:
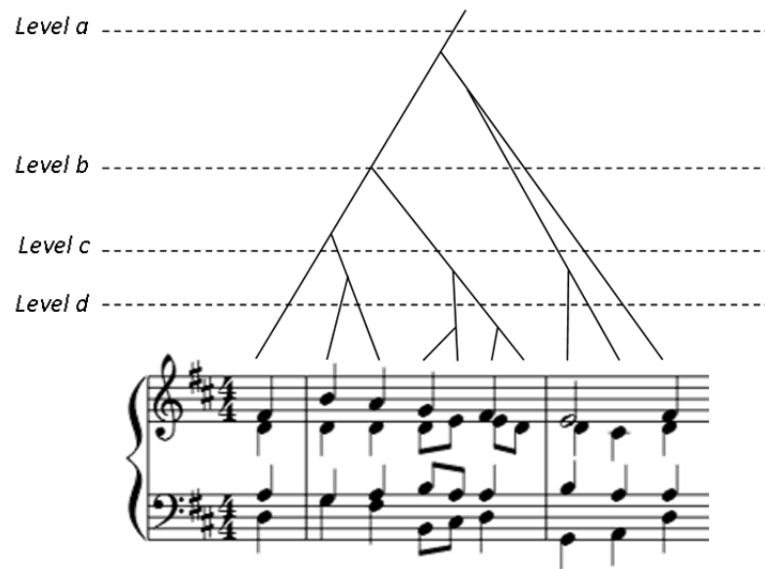
Figure 6: Hierarchical structure of *O Haupt voll Blut und Wunden* (Lerdahl & Jackendoff, 1983: 115)

      The observation that language and music exhibit hierarchical relationships among constituents has significant implications for our understanding of these two domains as cognitive systems. Schenker's methodology elucidates the fact that the hierarchical structure of music bears similarity to the principles of systematic organization promoted by generativism. Western tonal music is certainly not just an arbitrary aggregation of sounds combined for artistic purposes, but is highly cognitive, even for those who have had no formal musical training. Lerdahl and Jackendoff exploit this similarity in their approach to modeling our perception of structure in music by successfully applying the general principles of generativism to analyses of musical compositions. Their approach, Generative Theory of Tonal Music, will become the basis of an analysis to follow in Chapter 4.

2.3 EMPIRICAL STUDIES REVEALING COGNITIVE OVERLAP BETWEEN LANGUAGE AND

MUSIC

The following sections will review some of the empirical research that has investigated cognitive

connections between linguistic and musical syntax. Some of these studies focus on neural activations and

networks while others use behavioral methods to simulate interference paradigms in linguistic and

musical activities. These studies will reveal significant overlap during tasks that are specifically syntactic

(and not just related to similarities in the processing of aural stimuli).

2.3.1 Neuroimaging Studies

In one of the earliest studies examining the link between linguistic and musical neural activity, Patel,

Gibson, Ratner, Besson, and Holcomb (1998) focus on the P600, an event-related potential (ERP)

component that is generally thought to play a major role in grammatical processing. While there is some

debate over whether this brain potential, which is commonly known as a "syntactic positive shift," is

language-specific, its distinction from P300, an earlier non-language-specific positive component elicited

when there is a sudden change in a sequence, suggests that the P600 is specific to linguistic processing

(Picton, 1992; Osterhout, McKinnon, Bersick, & Corey, 1996). The P600 is elicited by a word that

obstructs parsing in an otherwise well-formed and meaningful sentence, such as the word *to* in *The broker*

persuaded to sell the stock was sent to jail (Osterhout & Holcomb, 1992, 1993). Most speakers expect the

verb *persuaded* to be followed by a noun phrase acting as the object of the action. The presentation of a

sentential complement is unexpected and difficult to integrate into the sentence.

In this study, researchers aimed to determine whether the P600 component elicited by difficult-to-

integrate words would also be elicited by a chord in a musical sequence that does not fit the established

tonal center, or diatonic scale. In the language trials, participants heard sentences with one target word or

phrase that was either easy to integrate (such the target phrase _an old idea_ in *Some of the senators had*

promoted _an old idea_ of justice.), difficult to integrate (*Some of the senators endorsed promoted _an old_*

idea of justice.), or impossible to integrate (*Some of the senators endorsed the promoted _an old idea_ of*

18

justice.) Likewise, the music trials presented a musical phrase with full three- or four-note chords; a target chord was either easy to integrate (in key), difficult to integrate (altered to fit a nearby key), or impossible to integrate (altered to fit a distant key).

ERPs were measured for each of the participants as they listened to the sentences and musical phrases. The results showed that the P600 components elicited by linguistic and musical anomalies were statistically indistinguishable in amplitude and distribution across the scalp. The moderate and severe incongruities in both the language and music trials produced a difference in amplitude. For all trials, severe incongruities produced significantly higher amplitude than the moderate incongruities. The authors conclude that the significant activation of the P600 in response to both linguistic and musical anomalies suggests that the syntactic integration processes that cause the P600 are not language-specific.

In a similar study, Maess, Koelsch, Gunter, and Friederici (2001) use magnetoencephalography (MEG) measurements to determine the localization of harmonic anomalies in a musical phrase that uses otherwise standard tonal progressions. Participants were presented with a sequence of chords that were organized to establish a tonal key center. In 25% of the phrases, the third chord of the sequence was changed to a harmonically unexpected chord, and in another 25% of the chords, the fifth and final chord of the sequence was changed to the same harmonically unexpected chord. The results of the MEG showed that magnetic early right anterior negativity (mERAN) was elicited by the chord in and around Broca's area and its right-hemisphere counterpart. The unexpected chord in the fifth position elicited a significantly larger mERAN than the same chord in the third position, likely because an unexpected chord is even more unacceptable in the final position of the chord sequence than in the middle of the sequence. Broca's area is commonly associated with syntactic processing in language and was previously thought to be language-specific, but this evidence of activation during a harmonic integration task suggests that the role of Broca's area in syntactic processing may be more generalized.

Koelsch, Gunter, von Cramon, Zysset, Lohmann, and Friederici (2002) further explore how language-specific domains may be implicated in musical syntactic processing and find that attempts to integrate harmonic incongruities activate an entire cortical network that is strikingly similar to a network

attributed to language processing. In this study, functional magnetic resonance imaging (fMRI) was used to measure the brain activity of participants, all non-musicians, as they performed a behavioral button task while listening to functionally tonal chord sequences played on a piano. Some of these functional sequences contained an unexpected chord, which was either harmonically deviant (e.g. a dissonant tone cluster or out-of-key chord) or timbrally deviant (i.e., played on another instrument, such as guitar). While listening to the progressions, the participants focused their attention on the harmonic and timbral elements of the chord sequences by pressing one of two buttons when they heard the timbrally deviant instruments. The right button was to be pressed when the participant heard a deviant instrument only if no tone cluster had been played since the last occurrence of a deviant instrument; otherwise the left button was to be pressed.

The fMRI results showed that all unexpected chords (i.e., both harmonic and timbral deviants) activated a number of structures used in syntactic processing of language, including the pars opercularis (part of Broca's area, which has been shown in previous studies to be implicated in music-syntactic processing) and the planum polare. Most significantly, the researchers observed interdependence between Broca's area and Wernicke's area that was previously only observed in language-syntactic processing. Other activated areas that are also implicated in language processing are the superior temporal sulcus, Heschl's gyrus, planum temporale, and the anterior superior insular cortices. While the researchers expected the harmonic deviations to elicit activations in language-syntactic areas, the elicitation of the same activations in response to timbral deviations contradicts the notion of syntax-specific processing mechanisms. If the network activated by harmonically deviant chords is detecting a true syntactic violation, then it should be unaffected by chords that are harmonically functional but are played by different instruments.

The researchers attribute the confounding activation of a syntactic processing network in response to timbrally deviant chords to the fact that the deviant-instrument condition was targeted by the behavioral task. By focusing a participant's attention concurrently on syntax and timbre, the experiment created a scenario in which "the decoding of the syntactic structure of deviant-instrument sequences was

more difficult because the different sounds presumably interfered with the harmonic analysis of the chord functions" (2002: 961). While more conclusive study is needed to determine the extent to which syntactic processing may be influenced by non-syntactic elements (such as timbre of the sound source), this study provides a valuable account of a language-syntactic processing network that also seems to be activated by music-syntactic processing tasks.

A study by Koelsch, Gunter, Wittforth, and Sammler (2005) tests the hypothesis that, if language and music share general syntactic processing mechanisms, then activations elicited by linguistic and musical syntactic incongruities will be affected when these incongruities are presented simultaneously. EEG measurements were taken while participants heard harmonically functional chord sequences, some of which contained a harmonically incongruous chord, while reading sentences, some of which contained a syntactic violation.[7] The EEG measurements showed that linguistic and harmonic syntactic violations that occurred simultaneously produced significantly different levels of activation than each of these violations produced independently. In the conditions with a harmonically incongruous chord with no linguistic violations, an early right anterior negativity (ERAN) was elicited maximally around 190 milliseconds, and in the conditions with a syntactically incongruous word but no harmonic violations, a left anterior negativity (LAN) was elicited maximally around 390 msec. The amplitude of the LAN is significantly reduced when the violation occurs simultaneously with a harmonically incongruous chord.

To ensure that the difference in LAN amplitude was actually due to syntactic processing and not simply any deviance-related negativity, the researchers conducted a second experiment using tones that were physically deviant rather than harmonically deviant. The physically deviant tones were manipulated to have a significantly different frequency, volume, or timbre from the other tones. The EEG measurements showed that the amplitude of the LAN elicited by a syntactically incongruous word was not affected by the simultaneous occurrence of a physically deviant tone. The researchers conclude that there must be some overlap in the neural resources used for processing linguistic and musical syntax.

---

[7] Some of the sentences used in this study contained semantic deviations, but this element of the study did not affect the syntactic elements and is therefore irrelevant to this paper.

A neuroimaging study by Jentschke, Koelsch, Sallat, and Friederici (2008) examines music-syntactic processing in children with a language impairment that is known to affect processing of (linguistic) syntax. Specific Language Impairment (SLI) in children is characterized by difficulties with syntactic comprehension and complexity that cannot be explained by other factors typically accompanying language impairment. The researchers use EEG measurements to compare the activations of two ERP components, the ERAN and the N5, in children with SLI and children with no language impairment in response to a harmonically functional musical sequence containing one harmonically incongruous chord. Both the ERAN and the N5 are elicited in 5-year old children in response to harmonic incongruities (Koelsch, Grossmann, Gunter, Hahne, Schröger, & Friederici, 2003). The researchers predicted that, because language- and music-syntactic processing use shared resources, children with difficulties processing linguistic syntax will also have problems processing musical syntax.

As predicted, the onset of a harmonically incongruous chord resulted in different ERP activations for children with SLI than for children who had no apparent language impairments. The deviant chord elicited both an ERAN and N5 in children with no language impairment, but neither of these components was observed in children with SLI. The researchers confirmed that these differences were not due to differences in intelligence or hearing abilities, which were controlled for in a pre-experiment participant screen, nor were they due to a difference in processing of acoustic features, because ERP responses to the first chord of each sequence were the same for both groups of children. The researchers conclude that the elicitation of the ERAN and N5 in children with no language impairment and the absence of these components in children with SLI strongly suggest that processing of language and music share cognitive resources.

2.3.2 Behavioral Studies

In more recent years, a number of behavioral studies have reinforced the evidence from neuroimaging research that language and music seem to share some overlapping cognitive resources. A study by Slevc, Rosenberg, and Patel (2008) uses an experimental setup very similar to many of the neuroimaging

studies, in which participants are presented with syntactically complex sentences while hearing chord sequences that may contain a harmonic violation. The sentences contained one of the following: 1) a reduced sentence complement that was predicted to involve "garden path processing" in which the subject of the embedded clause is initially interpreted as the direct object of the verb (such as *The scientist confirmed the hypothesis was false*); 2) a syntactically unambiguous full sentence complement (with *that*, such as *The scientist confirmed that the hypothesis was being studied in his lab*); 3) a sentence in which a target word is semantically expected; or 4) a sentence in which a target word is semantically unexpected. Additional "filler" sentences without syntactically or semantically unexpected elements were also included. The purpose of including sentence types (3) and (4) was to determine whether manipulation of the harmonic syntax of the chord sequences interfered with only syntactic processing or with more general language processing.

Participants read one constituent at a time on a computer screen, pressing a button to advance to the next constituent. Each press of the button also played one chord; the chords were organized to advance in a predetermined order, so that they formed a harmonic sequence establishing a tonal center. Each type of sentence listed above was presented with a chord sequence, and some of these chord sequences contained one harmonically "out-of-key" chord. In the syntactically and semantically complex sentences, the out-of-key chord aligned with the constituent that would be expected to produce a processing difficulty. After each sentence presentation, the participant answered a timed comprehension question.

An analysis of the reaction times showed that the presentation of an out-of-key chord does interfere with syntactic processing of a "garden path" sentence. The overall reaction times for the embedded verb were significantly larger in the reduced sentence complement trials than in the full sentence complement trials, showing the expected "garden path" effect. The "garden path" effect was significantly larger in trials that included a harmonically incongruous chord; i.e., the reaction times for the embedded verb in the reduced sentence complement trials were significantly slower when a harmonically incongruous chord was presented than when a harmonically functional chord was used. Furthermore,

analysis of the reaction times for the semantically manipulated sentences showed that, while reaction times for semantically unexpected items were significantly slower than those for semantically expected items, the presence of a harmonically incongruous chord did not significantly influence this effect. This distinction from the effects of the harmonically incongruous chord on the syntactically manipulated sentences is evidence for a specific connection between processing of linguistic and musical syntax.

Slevc et al. (2008) conducted a second study to confirm that the influence of a harmonically unexpected chord on syntactic sentence processing can be attributed specifically to the syntactic nature of the musical violation rather than the more general attentional effects of acoustic unexpectancy. This experiment used the same sentences and chord progressions as the previous experiment, but the harmonically unexpected chords were changed to harmonically expected chords with a dramatically different timbre (pipe-organ, as opposed to piano, a change significant enough to be equally or more distracting than the harmonically unexpected chords). An analysis of the reaction times showed that processing of the syntactically manipulated sentences, as well as the semantically manipulated sentences, was not significantly impacted by the presentation of a chord presented in the pipe-organ timbre.

2.4   CONCLUSIONS

From their basic building blocks to highly syntactic hierarchical representations, it seems that language and music share striking similarities that suggest these uniquely human modalities may be closely connected. What is most interesting about these similarities is that they represent *implicit* knowledge that a language user or music listener has acquired merely from exposure to, not even training in, the language or musical system. Of course, observations of structural similarities between language and music do not necessarily mean that these domains are connected in the brain or mind in any significant way, but they have inspired researchers to take a closer look at each of these domains and our understanding of their functions in a broader cognitive context. Neuroimaging studies reveal that music-syntactic tasks activate neural functions that were previously thought to be dedicated to linguistic-syntactic processing. Similarly, behavioral studies using interference paradigms suggest that music-syntactic tasks and linguistic-syntactic

tasks use overlapping resources. This research challenges the notion that language and music are separate

cognitive domains and invites further examination of how language and music are connected in the mind.

CHAPTER 3:

THEORETICAL IMPLICATIONS

3.1 DESIDERATA OF LINGUISTIC THEORY

The aims of linguistics as both a scientific and theoretical field of study have become as complex and

multidimensional as the very phenomena it seeks to explain. With the introduction of generative theory by

Noam Chomsky, linguistics, specifically syntactic theory, has become increasingly concerned with

distinguishing language use from linguistic knowledge as an abstract mental construct described using

deductive formalisms. On the first page of his influential *Aspects of the Theory of Syntax,* Chomsky

focuses on an "ideal speaker-listener, in a completely homogeneous speech community, who knows its

language perfectly" (1965: 3). This reference to an idealized form of language as distinctive from

language use represents unprecedented headway for linguistics as a theoretical field, because it allows the

researcher to get at the true nature of language isolated from interfering variables. Investigations of the

more cognitive aspects of language, such as language acquisition, comprehension, and production, have

come to encompass essentially autonomous fields of research independent from the abstract models of

linguistic knowledge proposed by theorists.

An oft-invoked justification for isolating language theory from any other cognitive phenomenon

is the idea that the language faculty is uniquely human and unlike any other cognitive capacity. By

assuming that there is a unique, dedicated language faculty, theories of linguistic knowledge are given

license to abstract away from principles of general cognition. The existence of dedicated special-purpose

cognitive mechanisms is a well-established biological phenomenon, from a human infant's natural ability

to recognize an unfamiliar face to a spider's competence in creating an intricate web without having been

taught. However, as Jackendoff (2011) contends, evolutionary principles require all special-purpose

mechanisms to have developed from the same neurons and cognitive machinery that inform general

cognition and must therefore be relevant enough to general cognition to be able to be encoded on the genome. Jackendoff proposes that linguistic study be subject to a standard of *graceful integration* with what we know about other human cognitive capacities: "[To] the extent that a theory of language permits a graceful integration with a plausible picture of the structure and function of the rest of the mind/brain, it places fewer demands on the genome, and therefore it is a better theory" (2011: 590).

In Chapter 2, we saw that another component of cognition, musical tonality, has syntactic structure and functions that bear striking resemblance to linguistic syntax. Considering the criterion of graceful integration, these similarities raise an important question about the desideratum of linguistic theory: If a study of language should be sensitive to what we know about the rest of the mind/brain, to what extent should our theoretical accounts of language accommodate and reflect a cognitive similarity between linguistic and musical syntax? Currently, we know very little about the neuronal activations and computations involved in syntactic processing, and we know even less about how the brain stores and accesses knowledge of linguistic or musical structures. But as we learn more about human cognition and the language faculty, maintaining a priority of graceful integration of abstract theory with what we know about language use will ensure that the field of linguistics remains objective, relevant, and meaningfully descriptive. Using graceful integration as a standard of scientific integrity, I will explore how a theoretical approach to linguistic and musical syntax can be consistent with empirical evidence of a cognitive connection between these two domains.

3.2   PROPOSING A SHARED SYNTACTIC PROCESSING MECHANISM

The first step in conceptualizing the link between linguistic and musical syntax is to identify the point at which the two domains diverge. The structural similarities between language and music described in Chapter 2 are interesting in that they show music to be analogous to language in much more sophisticated ways than is commonly thought. But the significance of this comparison should not overshadow the vastly intricate and complex ways in which linguistic and musical structure differ. Music has no analogue for language's parts of speech, morphology, or lexemes. The complex systems proposed by traditional

generative grammar and the Minimalist Program that involve abstract movement or transformations of deep, underlying structures do not have a translation or practical application for music (that we know of), suggesting that the similarities in the characters of linguistic and musical structure remain at a more conceptual rather than pragmatic level of comparison. Furthermore, evidence for double dissociations between language and music in patients with aphasia or amusia point to a categorical cognitive distinction between the two domains (Peretz; Peretz et al.; Ayotte et al; and Tzortzis et al., as cited in Patel, 2008: 268-270).

Patel (2008) reconciles these problems by proposing a "conceptual distinction" between representation and processing of linguistic and musical syntax. His "Shared Syntactic Integration Resource Hypothesis" (SSIRH) purports that language and music have separate stores of knowledge, such as information about words and how they can combine in language, or harmonic relationships and hierarchies in music, but they share the cognitive processes and resources that are required to access and integrate this knowledge. The SSIRH reconciles the apparent dissociations between linguistic and musical syntactic processing by purporting that cases of amusia are actually due to impairment or difficulties in the development of musical syntactic representations. Furthermore, Patel asserts that studies of individuals with aphasia without amusia are either significantly out-of-date or focus on individuals with extraordinary musical abilities, rendering them tangential to a study of general music cognition (2008: 284-285).

If we adopt the hypothesis that language and music share a syntactic processing component, then the task at hand involves identifying (or developing) a syntactic model that gracefully integrates linguistic and musical syntactic processing into a single shared mechanism. This body of research on linguistic theory makes a somewhat contentious distinction between language processing and production, often referred to as *performance,* and the abstract, stored knowledge a language user has of the language structure, called *competence.* Chomsky makes this distinction by describing competence as "the speaker-hearer's knowledge of his language" and performance as "the actual use of the language in concrete situations" (1965: 4). He and subsequent leading theorists deem performance subsidiary to considerations

of competence for several reasons, not the least of which being that idealized competence seems to have been the focus of the founders of modern general linguistics. The isolation of competence from performance is sound scientific methodology and should not be abandoned, but it is equally important that they be integrated into a broader context of the language faculty. During performance, processing mechanisms must access and interface with linguistic stores of knowledge; as such, models of performance and competence must somehow be relevant to and compatible with one another. Insofar as theories of syntactic processing are integral to a comprehensive understanding of the language faculty, a theory that elegantly models the processes of syntactic integration that are shared between language and music will constitute an important contribution to our understanding of how linguistic behavior fits into a greater cognitive context.

3.3 OPTIMALITY THEORY

Recent work in syntactic theory has advanced the application of Optimality Theory (OT), originally posited by Smolensky and Prince (1993) as a model of phonology, to linguistic syntax. According to the OT model, the forms of a language are created through the interaction and optimization of a set of violable constraints. The notion that constraints on syntactic structure are part of linguistic competence allows the model to be felicitously implicated in performance as well, namely in language processing. Given the success of OT approaches to linguistic syntax, this thesis will explore whether the cognitive machinery responsible for musical syntax also operates as a constraint optimization system. The analysis that follows will discuss the viability of an optimality theoretic approach to linguistic and musical syntactic processing as a step toward understanding the cognitive connection between the two domains.

Optimality Theory in its purest form as a theory of linguistic competence purports the existence of a universal set of linguistic tendencies, expressed in the form of violable constraints. Different languages may adhere to or violate these universal tendencies to different degrees, a concept traditionally related to *markedness*. Forms or patterns that are ubiquitous across different languages are referred to as *unmarked*, while patterns that are specific to a language are considered *marked*. Under the OT model, the

29

set of universal constraints serves to impose specific criteria on a surface form; however, these criteria often conflict with one another. Optimality Theory resolves these constraint conflicts by proposing that a language ranks some constraints higher in importance than other constraints; a language is differentiated from other languages by its precise ordering of the universal set of constraints.

The basic architecture of the OT model of phonology, given in (1), facilitates an optimal input-output relationship that violates the fewest highly-ranked constraints:

(1)   Input → *GEN* → Candidate set → *EVAL* → Output

The input refers to the underlying representation of a linguistic combinatorial unit. The generator, *GEN*, produces a candidate set of possible output structures based on the input. The evaluator, *EVAL*, evaluates each output candidate in terms of the ranked set of constraints. When evaluating which candidates to eliminate, *EVAL* considers the number of constraints each candidate violates, as well as the rank and type of each constraint. OT posits two types of constraints: faithfulness and well-formedness constraints. Faithfulness constraints reflect the linguistic preference that an output form matches its underlying representation, i.e., elements are not present in the output that are not in the input, and vice versa. Well-formedness constraints, on the other hand, concern the notion of *markedness* and are responsible for forms that differ from the underlying input but adhere to formulations required by the language.

Although Optimality Theory is typically understood to be a theory of linguistic competence, an OT approach to syntax has a distinct advantage over other competence models in that it can be readily extended to the domain of processing. Many competence theories, such as Principles and Parameters and the Minimalist Program, propose inviolable rules and transformations that often apply at the sentence level and simply cannot provide a derivation during online parsing and processing of sentence fragments. In contrast, a model that uses constraints that are violable to different extents offers a simple and elegant explanation for real-time processing phenomena such as ambiguity resolution and grammaticality judgments, even if the input is incomplete. This link between the grammar and parser provides a degree of efficacy to the OT approach to syntax, whereas previous competence theories either ignore

considerations of performance or require a completely separate framework to explain online processing (Stevenson & Smolensky, 2006).

3.4 OT SYNTAX MODELS FOR COMPETENCE AND PROCESSING

The competence model of OT for syntax uses the same architecture as the OT originally posited for phonology to account for the observed forms of a language, but OT syntax operates on the level of the phrase or sentence rather than the morpheme or word. Under this model, the *GEN* function is responsible for generating a candidate set of possible structures expressing a given proposition. These candidates represent both the articulated surface form of a phrase and also the relationships that are thought to exist among the constituents of the proposition. In the body of literature on optimality-theoretic syntax, these relationships are framed in terms of x-bar theory.[8] The candidate structures are evaluated by the *EVAL* function against the set of ranked syntactic constraints for a given language. The structure that emerges as grammatical is the one that is most harmonic[9] or optimal, meaning that the structure violates the fewest highly-ranked constraints.

Smolensky, Legendre, and Tesar (2006) give the following analysis (summarized here) as an example of an OT model of syntactic phenomena. Consider the sentence *He has sung*, which has an underlying interpretation that may be represented as follows using somewhat standard syntactic notation where *x* indicates topic and T indicates tense: /sing(x), x=he; T=present perfect/. A number of syntactic constraints are implicated in the process of constructing possible output forms for the expression of this underlying interpretation. These constraints, proposed as syntactic constraints by Grimshaw and Samek-Lodovici (as cited in Smolensky et al., 2006) are summarized as follows:[10]

(2) SUBJ (Subject) - The subject position of a clause must be filled.

(3) PARSE (Parse) - A unit of the interpretation has a corresponding unit in the expression.

---

[8] A concise but thorough explanation of x-bar theory is given in Smolensky, Legendre, and Tesar, 2006: 466-472.

[9] Not to be confused with musical harmony, harmony in this context refers to linguistic well-formedness.

[10] I have excluded **ALFOC Align Focus** because it is not relevant to this example.

(4) FULLINT (Full Interpretation) - A unit of the expression has a corresponding unit in the interpretation.

(5) DROPTOP[11] (Drop Topic) - Arguments referring to the topic in the interpretation have no corresponding units in the expression.

These constraints are listed as column headers in the tableau given in Table 1, and in the left-hand column are four possible structures that spell out the underlying interpretation I = /sing(x), x=he; T=present perfect/:

Table 1: Tableau for [I = /sing(x), x=he; T=present perfect/], English

I = /sing(x), x=he; T=present perfect/	PARSE	SUBJ	FULLINT	DROPTOP
☞ a. [$_{IP}$ he$_i$ has [t$_i$ sung]]				*
b. [$_{IP}$ it has [[t$_i$ sung] he$_i$]]			*!	*
c. [$_{IP}$ has [[t$_i$ sung] he$_i$]]		*!		*
d. [$_{IP}$ has [sung]]	*!	*		

Constraint violations for each of the candidate structures are indicated by an asterisk in the column of the constraint. SUBJ is violated by candidates (c) and (d), because both candidates are lacking a filled subject position. All candidates except for (d) express every element of the interpretation (*sing*, topic *he*, and present perfect tense), so (d) is the only candidate to violate PARSE. Similarly, (b) is the only candidate to have a unit that does not have a corresponding unit in the interpretation (expletive *it*), so (b) is the only candidate to violate FULLINT. The topic *he* is expressed in all candidates except for (d), so all candidates except (d) violate the DROPTOP constraint.

The constraints are listed in the tableau in order from left to right of decreasing dominance, often also notated as PARSE >> SUBJ >> FULLINT >> DROPTOP. This dominance hierarchy is derived by examining numerous different grammatical formations and determining patterns of constraint strength.

---

[11] DROPTOP is not highly ranked in English (i.e., violation is acceptable), but it is included here for comparison to Italian constraint rankings.

All constraints are potentially violable, but a highly ranked constraint will only be violated in a grammatical formation if all other candidates violate an even more highly ranked constraint, creating a situation in which the given formation is optimal in terms of violation in comparison to other candidate formations. The candidate in Table 1 with the least severe constraint violations given the constraint ranking PARSE >> SUBJ >> FULLINT >> DROPTOP is candidate (a), indicated with the ☞ symbol. Not only does this candidate only incur one constraint violation, but the constraint it does violate, DROPTOP, is ranked lower than the other three. Languages differ by their constraint rankings, and a different constraint ranking would result in a different candidate chosen as optimal (resulting in syntactic variation across languages). For example, Table 2 gives the tableau for the same proposition with the rankings adjusted for Italian (given in English here), a language that gives greater importance to the DROPTOP constraint. This ranking results in the selection of candidate (a) (which was candidate [d] in Table 1) as optimal and most grammatical.

Table 2: Tableau for [I = /sing(x), x=he; T=present perfect/], Italian

I = /sing(x), x=he; T=present perfect/	FULLINT	DROPTOP	PARSE	SUBJ
☞ [$_{IP}$ has [sung]]			*	*
[$_{IP}$ he$_i$ has [t$_i$ sung]]		*!		
[$_{IP}$ has [[t$_i$ sung] he$_i$]]		*!		*
* [$_{IP}$ it has [[t$_i$ sung] he$_i$]]	*!	*!		

This account of Optimality Theory as adapted for linguistic syntax is, like its phonological counterpart, a theory of linguistic competence. It is a descriptive theory in that it elucidates the language-specific relationship between an underlying interpretation of a proposition and the form of the actual expression that articulates that interpretation in a given language. A controversial extension of this conception of Optimality Theory (and presumably of all competence theories) is the claim that it can model the actual cognitive processes through which the language faculty arrives at a correct expression of

a proposition. The extent to which the mechanisms employed by the OT competence framework are cognitively instantiated is not an issue I will examine here, but such concerns may be an interesting avenue of future exploration in light of the current discussion.

The counterpart to linguistic competence is the notion of linguistic performance, and a subcategory of performance is language processing, or how language is received and understood in real time. The division of these components within the field of linguistic theory has resulted in highly disparate theories of each. Typically theories of competence have little connection with or ability to account for aspects of performance, making these theories difficult to incorporate into a larger context of language as a cognitive activity. By contrast, the architecture of Optimality Theory, while traditionally understood in terms of linguistic competence, is not at all at odds with aspects of linguistic performance. Assuming that hearing and understanding propositions in a given language involves the constant integration and evaluation of incoming elements in terms of the grammaticality of an overall structure, and given that the constraint hierarchy of a given language constitutes (at least in part) its grammar, the concept of constraint satisfaction can be soundly adapted to linguistic processing as a function of the parser. This interpretation of Optimality Theory as a theory of both competence and processing lends the approach an unprecedented degree of robustness and explanatory power.

How does an OT model of linguistic processing differ from an OT model of linguistic competence? The framework of a processing OT model uses the same optimization mechanism, the same constraints, and the generative and evaluative functions but reverses the components that are generated and evaluated. In competence OT models, the input is an underlying interpretation, the candidates are possible structural descriptions given the interpretation, and the output is an optimal expression of the interpretation. In processing OT, the input is a given surface form, the candidates are possible interpretations of the surface form, and the output is an optimal interpretation and structural description of the given expression. While the framework of the theory remains the same, the component that generates the candidate structures for OT competence is performing a different function than the component that generates structures for OT performance. For this reason, I will follow a convention proposed by

Stevenson and Smolensky (2006) to refer to the generation component of OT competence as *Gen* and that of OT processing as *Int* (because it generates a candidate set of interpretations).

Stevenson and Smolensky demonstrate the OT processing model by exploring how the parser handles points of syntactic ambiguity that arise after only a portion of a sentence has been presented. Intuition, as well as theoretical reflection and empirical study, tells us that the parser does not wait until the end of a sentence to extrapolate an interpretation of the sentence but rather incrementally integrates incoming units into a posited structural interpretation. The OT model hypothesizes that the structural interpretation assigned by the parser at different points in the presentation of an utterance is shaped by the language's syntactic constraints, which reside in the language user's linguistic knowledge, or competence.

Following methodological procedures of OT for phonology and syntactic competence, Stevenson and Smolensky work through a number of analyses to derive a ranking of syntactic constraints. To show that constraint interaction plays a part in determining parsing preferences during online processing, the analyses compare the possible interpretations of a sentence fragment at the moment the parser encounters a syntactic ambiguity, or a point at which more than one sentential interpretation is possible depending on how the input proceeds.

Consider the following sentence fragment, where *[x]* represents material that has not yet been pronounced: *John put the candy on [x]*. When the parser receives this fragment, the presentation of *on* projects a prepositional phrase that may serve as the locative argument of *put*, as in *John put the candy on the table*; alternately, the prepositional phrase may serve as an optional modifier of *candy*, as in *John put the candy on the table into his mouth*. At this point in the parse, each of these candidate structures violates a syntactic constraint. Candidate (a) violates LOCALITY, which requires that a phrase XP that is c-commanded by phrase YP to be preceded by YP. In the given fragment, the NP *the candy* is c-commanded by the PP *on [x],* because the PP is higher in the tree than the NP, and neither dominates the other; however, the PP does not precede the NP, so LOCALITY is violated. Candidate (b) violates ASSIGN-θ, which requires a predicate to assign all of its thematic roles. In the given fragment, this parse has not

yet projected a phrase to fill the locative thematic role of the verb *put*. These violations are represented in the following tableau:

Table 3: Tableau of candidate interpretations for *John put the candy __on__ [x]*

John put the candy __on__ [x]	ASSIGN-θ	LOCALITY
☞ a. … the table		*
b. … the table into his mouth	*	

Stevenson and Smolensky argue that candidate (a) is the preferred parse (i.e., this candidate is how most individuals expect the sentence to continue and is the parse that presents the least amount of processing difficulty). For this reason, candidate (a) is the optimal parse, and we can conclude that, since violation of ASSIGN-θ results in a suboptimal parse, the ASSIGN-θ constraint is ranked higher than the LOCALITY constraint.

In a second analysis, Stevenson and Smolensky incorporate a third constraint, OBHD (Obligatory Head), which requires the heads of all projected phrases to be filled.[12] When presented with the fragment *I told the department __committees__ [x]*, the word *committees* might be interpreted as the head of an NP (containing *department* as a modifier) that serves as an argument of *told*. While this parse is the preferred interpretation, it violates ASSIGN-θ, because the sentential complement of *told* has not yet been presented. Another candidate interpretation parses *committees* as the NP subject of the sentential complement of *told,* which satisfies ASSIGN-θ because it allows the projection of the required argument but violates the lower-ranked constraint LOCALITY because the sentential complement c-commands the preceding NP *the department*. However, this candidate also violates the OBHD constraint, because the projected sentential complement does not have a head, only a specifier (under the analysis that the NP occupies the specifier position of the VP). Because this candidate is the dispreferrred interpretation when the parser reaches the

---

[12] OBHD is a more general form of the SUBJ constraint referenced in the first example in this section.

word *committees*, we can conclude that OBHD is ranked higher than both ASSIGN-θ and LOCALITY, as demonstrated in the following tableau:

Table 4: Tableau of candidate interpretations for *I told the department committees [x]*

I told the department committees [x]	OBHD	ASSIGN-θ	LOCALITY
☞ a. … that budgets were cut.		*	
b. … would be formed.	*		*

These examples have demonstrated the essential architecture of an OT model for linguistic processing. The successful extensions of OT as a model of phonological competence to one of syntactic competence, and recently to a model of syntactic processing, constitutes significant progress for the field of linguistics in its objective to understand and characterize the language faculty as a whole. Extending this endeavor to another cognitive modality, music processing, may offer further insight and will help to reconcile evidence that language and music seem to share processing resources.

3.5   ADVANTAGES OF AN OT FOR MUSICAL SYNTAX

The current inquiry is driven by Jackendoff's (2011) criterion of *graceful integration*. If an optimality theoretic approach to music is successful, then positing OT as a model of general cognitive syntactic processes will have the advantage of being gracefully integrated with what we already know or assume about cognition in two respects:

(6)   *Integration with Empirical Evidence:* Using OT to model both linguistic and musical syntax supports the premise that they are processed by a more general syntactic optimization mechanism. This notion of a shared processing mechanism allows us to account for an empirically established cognitive link between linguistic and musical syntax,

whereas other syntactic theories fall short of providing a plausible explanation for the same phenomenon.

(7)   *Integration with Neuroscientific Theory*: According to Prince and Smolensky (1997), OT has its roots in more general theories and studies about human cognition, notably the use of a neural network concept to model cognitive processes. The success of neural networks as models for other cognitive processes lends OT a distinct advantage over theories that are more isolated from other areas of cognitive study.

In contrast to more abstract theories of grammar, the principles of OT as proposed by Prince and Smolensky can be gracefully integrated with more general theories of higher cognition. The researchers focus on the interconnectedness and integration of cognitive machinery: "[N]eural computation ought to explain fundamental principles of the higher level theory by deriving them as large-scale consequences of interactions at a much lower level" (Prince & Smolensky, 1997: 1606). In their model neural network, Prince and Smolensky describe constraints as being implicit in the connections between neural units, and activation selections are based on the combination of connections that violates the fewest constraints. For example, a constraint dictating the relationship between two particular neural units may require that when one neural unit is active, the other must be inactive. Another constraint may require that when one unit is active, the other must also be active. When these conflicting constraints compete in a complex neural network made up of many units and many other constraints, activations between units will inevitably differ in the patterns of activations and constraint violation and adherence.

The reality of hierarchical structure in syntactic representations is not at odds with the neural network processing model; in fact, the neural network model provides a deeper explanation for the combinatorial principles that characterize linguistic theory. Each linguistic unit represents a pattern of harmony maximization. As units are combined, new and existing constraints interact to maximize harmony in increasingly complex networks, producing theoretically larger and larger linguistic units. This approach links the well-formedness of linguistic structures to the differential harmony of the underlying neural network, but it alone does not account for the notion of strict domination evident in theoretical

analyses of linguistic constraints. Prince and Smolensky maintain that the current inability of neural networking models to integrate the characteristic of strict domination in linguistic constraint interaction does not undermine the proposition that grammar is cognitively based on processes of neural optimization.

CHAPTER 4:

TOWARD AN OPTIMALITY THEORETIC APPROACH FOR MUSICAL SYNTAX

4.1 SCOPE OF THE CURRENT DISCUSSION

The formalization of an OT for musical syntax is an enormous undertaking. The objective of this section

will not be to elucidate a complete OT analysis of musical syntax but rather to explore the principles of

constraint adherence and optimization that are at play in Western tonal music. As is the case for language,

enculturation in a musical system results in the development of cognitive processes and representations

that are complex and dynamic and, similarly to language, invoke violable constraint systems. The goal of

the current discussion is to support the hypothesis that language and music share syntactic processing

resources, and that these resources involve the evaluation of constraint violations to determine a preferred,

or optimal, structural interpretation of a linguistic or musical formation. In the following chapters I will

also address concerns and possible directions for future research into the optimization-based character of

musical syntax and what this means in a broader cognitive context.

    Because the focus here will be on musical syntax as a cognitive domain similar to language, there

are many issues commonly associated with musical analysis that will be outside the scope of the current

discussion. Crucially, the analyses will be concerned with the musical knowledge possessed by an

individual who has not necessarily received any formal musical training but has acquired implicit

knowledge of a music-syntactic system through regular exposure to the musical products of our culture

(through commercial jingles, pop music, folk songs, and the like). This type of individual will be referred

to as an *enculturated* listener and is analogous to a speaker who is fluent in a language. Additionally, the

function of music as a form of art, aesthetics, or affective expression will not be directly addressed, as

these are somewhat peripheral to the notion of the tonal system as a grammar. There is a distinction

between the cognitive processes that allow a listener to understand a piece of music and the affective

processes that allow him to enjoy it (or not). Sloboda (1985) makes a germane analogy between music

and humor; in order to understand a joke, the listener must invoke a series of cognitive processes that

allow him to understand the language of the joke and various other semantic and pragmatic factors to

determine what aspect is intended to be humorous. The listener then may have an affective response to

results of those cognitive processes, but the processes used to arrive at the affective stage are not

inherently concerned with affect themselves.[13]

My analysis will integrate two theoretical frameworks into a single OT approach to musical

syntax. The first framework is the OT model of linguistic syntax elucidated in the previous chapter, which

will then be somewhat modified once we delve deeper into the notion of syntactic parsing. The second

framework is the model of musical interpretation devised by Fred Lerdahl and Ray Jackendoff in their

1983 work *A Generative Theory of Tonal Music*. Chapter 2 discussed how GTTM develops a system of

rules that dictate how a listener attributes structure to a piece of music, representing hierarchies and

relationships among the musical events. The theory advances the notion that musical understanding, like

language, involves the construction of abstract structures according to innate and learned syntactic

principles. The authors maintain that this process of conceiving musical structure is what makes hearing

music such a richly cognitive activity in that a listener's understanding and appreciation of music clearly

involve far more than just hearing a sequence of pitches.

A central tenet of music theory, and GTTM in particular, is that a musical composition can be

reduced and simplified to a more abstract structure that corresponds to a listener's mental representation

of the musical surface. This property of music cognition explains why nearly every rendition of "The Star

Spangled Banner" is different but still identifiable, why Bach's 30 dramatically different *Goldberg*

Variations are still recognizable as variations on one simple theme, and why the recently popular

composition trend of the "mashup" allows a listener to recognize (portions of) a popular song even after

some aspects of it have been modified to fit with another. The phenomenon also comes into play in

---

[13] Temperley (2001) interestingly refers to musical syntax as "infrastructure," analogizing metrical and harmonic structure to the transportation and communication structures that are important to the functioning of society.

improvisational performances, such as jazz, wherein the performer must understand the basic structure of a musical theme or phrase and maintain its integrity while creating interesting modifications and variations. The perception of an abstract musical structure is what differentiates musical listening from merely acoustic perception. The listener perceives this underlying structure more or less unconsciously by making judgments about the structural importance of musical events in relation to one another. Given that the sound units that make up a musical composition are inherently meaningless, the listener employs a number of cognitive and syntactic organizational principles to parse the musical surface and determine its underlying structure.

Like language, the process of determining a syntactic structural interpretation of a musical passage requires that the listener already have a number of music-syntactic competencies. Linguistic competencies, which include knowledge of the lexicon, theta roles, phrase projection and movement, and grammatical relations, are a necessary basis for the linguistic theoretic models discussed in the previous chapter. For the purposes of the current discussion on musical syntax, it will be necessary to familiarize the reader with some of the basic competencies acquired by an enculturated listener of Western tonal music.[14] Borrowing terminology from Patel (2008), the musical counterparts[15] of these competencies fall into two categories: knowledge of pitch hierarchy and knowledge of event hierarchy. Knowledge of pitch hierarchy refers to the atemporal aspects of tonal organization, namely the differing stabilities of the notes of the diatonic scale and how these notes can be combined to form differentially stable chords. In an abstract sense (see ftn. 15), knowledge of pitch hierarchies may be thought of as functionally similar to knowledge of a lexicon, which includes the understanding of how the words relate to one another and how they can and should be combined. Knowledge of event hierarchy, on the other hand, involves the principles that dictate how musical events relate to one another temporally, so that a musical sequence is

---

[14] Describing these competencies as "acquired" may not be entirely accurate, since they may be connected to or derived from innate components of the human perceptual mechanisms. The investigation of innate musical competence is an important avenue of future research.

[15] I am careful not to suggest that any individual components of linguistic competence have direct counterparts in musical competence; i.e., I am not making the claim that the mental machinery responsible for lexical knowledge is also responsible for knowledge of the chromatic or diatonic scale hierarchies, although these comparisons are interesting in their own right.

perceived as having a hierarchical structure in which some events are heard as more important and others as more ornamental. A key component of event hierarchy competence is the division of the musical surface into temporal domains, called **time-spans** (following GTTM), over which decisions of structural dominance are made. This knowledge follows the intuition that a listener interprets musical structure as belonging to a hierarchy of time-spans; for example, a listener hears small phrases as being embedded in larger phrases, which are embedded within an entire piece. Both pitch hierarchy and event hierarchy competencies will be implicated in the OT approach to musical syntax to follow, so I will begin with a discussion of each of these components.

4.2 PITCH HIERARCHIES

The property of a musical sound that makes it sound higher or lower in pitch is its frequency. Sound frequencies exist in the natural world on a continuous and theoretically infinite spectrum; the human voice can produce a limited segment of the full sound continuum by starting on a low note and, supplying a continuous stream of breath, gliding the voice upwards. Compare this exercise with humming a tune, wherein the voice hovers on discrete pitches rather than sliding all over the continuum. Musical systems all over the world are based on a division of the sound continuum into a set of discrete pitches. This measured segmentation is not uncommon in other human organizational systems; consider the division of continuous time into seconds, minutes, and hours, or the categorization of colors from the full color spectrum. While an individual is certainly capable of perceiving significantly finer divisions of time, color, or, in the case of musical systems, sound, limiting the processing units of each of these domains to a subset of discrete constituents is felicitous for cognitive organization and meaning.

Out of this subset of discrete sound units emerges a hierarchy in which some pitches are perceived as more stable or important than others within a given key. Implicit knowledge of these differing stabilities is an essential factor in a listener's perception of music as a meaningful and complex temporal system of tensions and releases, rather than just noise. Consider a musical composition that has never before been heard by a particular listener. Although the sound sequences are completely novel, the

listener knows which sonic events sound tense and unstable and which ones sound more resolute, as if the harmonic movement has come to a resting point. Should the composition end right before presenting the final note, the listener is likely to know something is not quite right, as if the music never fully came to a close, because the second-to-last note sounds unstable. A basic example of this phenomenon in Western tonal music is the notion of the tonic, which is the one pitch that is perceived as the most structural, central, and stable. Atemporal, hierarchical systems of pitch organization are found in musical traditions worldwide, a phenomenon that is likely due at least in part to a preference for psychological reference points in the organization of mental entities. Pitch stability hierarchies in the Western tonal tradition are essentially manifested in two dimensions, scales and chords. These concepts were discussed in Chapter 1, but the following sections will address in more detail how they contribute to an enculturated listener's implicit music-syntactic knowledge.

4.2.1 Scales

As noted in Chapter 2, the inventory of pitches used in Western tonal music exists as a mental representation organized in terms of musical scales, specifically the chromatic and diatonic scales. The chromatic scale, which comprises all of the possible notes that may be used in Western tonal music, divides the octave into 12 logarithmically equal intervals resulting in 12 discrete notes. The diatonic scale is a subset of the chromatic scale; unlike the chromatic scale, the notes of the diatonic scale are not equidistant but are separated by one or two chromatic steps in the following pattern: 2 2 1 2 2 2 1. For example, beginning this pattern on D would produce a major diatonic scale that contains only the underlined notes in the following chromatic scale:
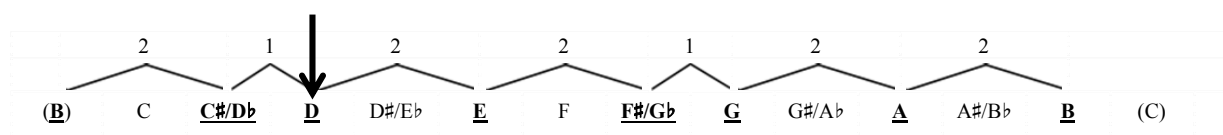
Figure 7: Major diatonic scale (underlined notes) beginning on D

The asymmetrical patterning of intervals is instrumental for establishing a stability hierarchy among the pitches of the diatonic scale. The note on which the interval pattern of 2 2 1 2 2 2 1 begins is perceived as being the most stable and important note of the entire scale, and the other notes are perceived as being less stable to different extents. For this reason, the scale is named after the most stable note (e.g., a scale that begins the pattern on D is called a D scale), and this note is also called the *tonic* or *tonal center.* When hearing a piece of music, an implicit knowledge of the asymmetrical patterning of intervals allows the listener to establish a point of reference for the presented pitches within the context of the diatonic scale. If the patterning of intervals were symmetrical, as is the case in the chromatic scale, the listener may have difficulty identifying a tonal center because it could be any note. Ball (2010) illustrates this concept by analogizing the diatonic scale to a staircase. Say an individual uses the same staircase every single day, and one day he finds himself blindfolded somewhere on the staircase. If the steps are all the same height, he might not know whether he is near the bottom, middle, or top of the staircase. By contrast, if the steps are of unequal heights, it may only take a few steps before he is able to make a good guess as to where he is on the staircase. The intervallic patterns of the diatonic scale can orient a listener to the tonal center in part because an enculturated listener has internalized the structure of the diatonic scale and instinctively employs this knowledge when hearing Western tonal music.

4.2.2   Chords

Also noted in Chapter 2, another level of syntactic organization in Western tonal music concerns harmony, or chordal structure. The concept of the "chord" (or triad) generally refers to notes that are sounded simultaneously in music, but a chord can also be implied by notes that are sounded individually. A chord sometimes occurs as a homogenous block of sound, in which all notes of the chord begin and end at the same time; traditional church hymns and Bach chorales are often composed of a single block chord per syllable of the sung line. Another way chords are created is by stacking melodic lines so that the notes that occur simultaneously each form a chord. For example, when "Row, Row, Row Your Boat" is sung in a canon, with different individuals starting at different times, the resulting polyphonic texture is

45

essentially a series of chords. In much of Western tonal music, chords are created through some combination or degree of these two methods, interweaving block chord textures with melodic lines to different extents. Transitions between chords can be fluid and involve overlap among the chord members. What emerges, sometimes more clearly than others, is a sense that the music comprises a sequence of polyphonic sonorities wherein a single chord is in focus at any given time.

The basic chordal framework in Western tonal music is the major or minor triad. Triads are formed by selecting any note of the diatonic scale (which is called the *root* and is often perceived as the most important note of the chord), then skipping the note following the root in the diatonic scale and selecting the next one, and then skipping the next note in the scale and selecting the next one. Using this procedure, the middle note of the chord is always either three or four chromatic steps away from the root, and the top note is always seven chromatic steps away from the root. The middle note is called the *third* of the chord and the top note is called the *fifth.* In music theory, roman numerals are conventionally used to label a triad corresponding to the degree of the diatonic scale on which it is built; e.g., a iii chord is built on the third note of the scale, a V chord on the fifth note of the scale, etc. Major chords, which we perceive as sounding bright and happy, are characterized by a distance of four chromatic steps between the root and the middle note of the chord and are represented by an uppercase roman numeral. Minor chords, which are often considered to sound darker and more foreboding or sad, are characterized by a distance of three chromatic steps between the root and middle note and are represented by a lowercase roman numeral. Figure 8 shows again the possible triads that can be used in a Western tonal composition, presented in *root position*:
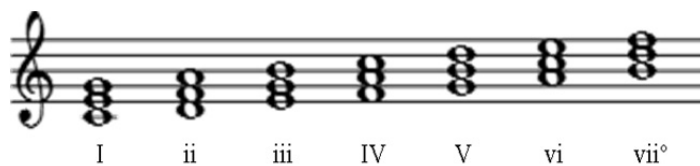
Figure 8 (reprint of Figure 2): Chords built on each note of the diatonic scale

In actual composition, the notes of the chords may be distributed over the entire staff, with any of the three notes occupying the lowest, middle, and highest positions. A chord with the third in the lowest position, or *bass*, is said to be in *first inversion*, and a chord with the fifth in the bass is in *second inversion*.

Of all the possible note combinations, it may seem arbitrarily specific for the triad to serve as a basic polyphonic building block of Western tonal music, but a number of hypotheses have been proposed to account for its perseverance. The triad seems to have particular qualities that allow it to serve as a cognitive anchor for the internalization and processing of principles of stability in Western tonal music. These qualities are related to notions of consonance and dissonance, which generally refer to how pleasing two or more notes are perceived to be when sounded at the same time. Dissonant intervals are those that seem to clash and produce a displeasing sound, a phenomenon that is at least in part related to how the sound waves of the two notes relate. Small intervals, such as the minor or major second, are considered dissonant because the wavelengths of the two notes are similar enough for the ear to attempt to fuse them together but different enough to clash against each other, producing an arguably unpleasant psychoacoustic sensation of rapid beating or roughness. By contrast, intervals that have simple relationships between wavelengths are thought to sound more pleasant, because there is more alignment among the frequencies. Triadic harmonies may be prevalent in Western tonal music because they use highly consonant intervals.[16]

---

[16] When discussing notions of consonance, it is important to clarify that what a listener considers consonant may be learned through exposure to his culture's music. However, many argue that the consonance of the triad is actually based in psychoacoustic properties rather than learned preferences, because triadic harmony mimics the naturally-occurring harmonic overtone series. Whenever a pitched instrument plays a single tone, the sound that is produced contains multiple frequencies, called overtones. The lowest overtone is called the fundamental and is heard as louder than the other, and the rest of the overtones are whole-number multiples of the fundamental frequency. The first overtone is an octave above the fundamental, the second overtone another interval of a fifth (or seven chromatic steps) above the first overtone. The third overtone is yet another octave, and the fourth is the interval of a third above the fundamental (four chromatic steps). The overtones continue upward but become significantly less perceivable. The ear "converts the 'chord' produced by these overtones into a perception not of several simultaneous notes but of timbre: the blend of harmonics gives a characteristic sonic quality to what seems to be a single pitch" (Ball, 2010: 67). While not always perceptible, the loudest overtones of the harmonic series create a major triad, an observation that may lend credence to the argument that the ear may have a natural preference for the tonal triad.

Just as the notes of the diatonic scale are in a hierarchy of stability, tonal triads are perceived as differentially stable in relation to one another. Chords that are more stable within the tonal hierarchy are determined to be more structural. The stability of a chord within the tonal hierarchy is intuited somewhat easily by any individual familiar with Western tonal music, even more so for a trained musician or music theorist. However, in the context of a theoretical approach to musical syntax, it is important to understand that these different degrees of stability were not arbitrarily assigned to each of the seven harmonies and then memorized by enculturated listeners; rather, the stability relationships emerge as a result of the organizational principles of tonality. A number of theorists have attempted to model these relationships (see Shepard, 1982: 351-378 and Lerdahl, 2001: 42-45 for a review of these approaches), one of the most heralded being Lerdahl's *Tonal Pitch Space* (TPS) theory (2001). TPS quantifies the level of stability of any given musical event within a tonal context in terms of that event's *tonal distance* from the tonal center.[17] Tonal distance, in this context, is not acoustic (i.e., it is not directly related to differences in frequency) but rather cognitive, and it corresponds to an internalized knowledge of tonal relationships. The algorithms for determining tonal distance (and stability) of a chord according to TPS are quite complex, but the basic architecture of the theory can provide important insight for an account of musical syntax. The theory quantifies the interaction of several factors as a prediction and measure of the tonal distance between two chords; the two most significant factors are the distance of the two chords on the theoretical construct called the *circle of fifths* and the number of common tones between the two chords. The circle of fifths (Figure 9) is used often in music theory to depict the most basic and fundamental relationships among pitches and chords in the tonal system. Pitches (and their corresponding triads) that are next to each other on the circle when moving clockwise are separated by the highly consonant interval of a fifth (and simultaneously the inverse interval of a fourth when moving counterclockwise[18]) and are

---

[17] TPS also suggests ways of measuring tonal distance between neighboring non-tonic chords, but these considerations are somewhat peripheral to the current discussion.

[18] Because the notes of the chromatic scale are cyclic (reflecting a natural property of the sound continuum that pitches with a frequency differing by a factor of 2 sound parallel), intervals between notes have inversions that are characterized by going the other direction in the cycle. So the distance from C to G is the interval of a fifth (seven chromatic steps), but the distance from G to C is the interval of a fourth (five chromatic steps).

therefore perceived as tonally similar. The number of clockwise or counterclockwise (whichever is a shorter distance) shifts on the circle needed to get from one chord to another is a contributing component to the overall tonal distance between those chords as calculated by TPS.
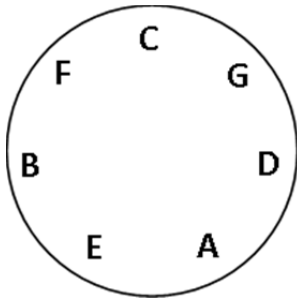
Figure 9: Diatonic circle of fifths[19]

The other major component of tonal distance according to TPS, number of common tones, is calculated so that it not only takes into account the number of shared pitches (assuming octave equivalence) between the two chords but also gives more significance to individual pitches that have more salience and importance within the established tonal hierarchy. Lerdahl proposes the concept of a *basic space* shown in Figure 10, which demonstrates the five levels of tonal hierarchy in Western tonal music. The hierarchy abstracts away from all twelve chromatic notes available to Western tonal music in level *e,* to only the members of the diatonic collection in level *d,* the basic tonic triad (the most stable triadic harmony) in level *c,* to the most consonant interval of a fifth rooted on the tonic in level *b*. The highest level of abstraction, level *a,* contains only the tonic and all of its octave equivalents. The basic space schematic provides a way of conceptualizing how pitches have varying weights or importance within the context of a key or chordal harmony; the more times a pitch is represented by a letter in the basic space, the more tonally important it is in relation to the other pitches.

---

[19] It should be noted that the standard circle of fifths (as opposed to the diatonic circle of fifths used by TPS) includes all chromatic notes and avoids the diminished fifth between B and F.

	C	C♯/D♭	D	D♯/E♭	E	F	F♯/G♭	G	G♯/A♭	A	A♯/B♭	B	(C)
Level *a*	C												(C)
Level *b*	C							G					(C)
Level *c*	C				E			G					(C)
Level *d*	**C**		**D**		**E**	**F**		**G**		**A**		**B**	**(C)**
Level *e*	C	C♯/D♭	D	D♯/E♭	E	F	F♯/G♭	G	G♯/A♭	A	A♯/B♭	B	(C)

Figure 10: TPS basic space (Lerdahl, 2001: 47)

To quantify the influence of shared pitches on perceived tonal distance, the basic space is manipulated so that the chromatic and diatonic levels of the basic space (levels *e* and *d*) remain the same, but the levels of the triad, fifth, and octave shift to reflect the pitch content of the chord in question. Figure 11 shows the basic space of the V chord (which contains the pitches G, B, and D) in the key of C:

	C	C♯/D♭	D	D♯/E♭	E	F	F♯/G♭	G	G♯/A♭	A	A♯/B♭	B	(C)
Level *a*								G					
Level *b*			D					G					
Level *c*			D					G				B	
Level *d*	**C**		**D**		**E**	**F**		**G**		**A**		**B**	**(C)**
Level *e*	C	C♯/D♭	D	D♯/E♭	E	F	F♯/G♭	G	G♯/A♭	A	A♯/B♭	B	(C)

Figure 11: TPS basic space for V chord in the key of C major

To calculate the tonal pitch space distance between the tonic C and the G chords, the pitches in the G chord that are not represented at the same levels of hierarchy in the C chord are tallied (circled in figure). This number is added to the number of shifts required on the circle of fifths to produce an overall measurement of tonal distance and stability. The resulting calculations are consistent with intuition and empirical findings (e.g. Krumhansl & Kessler, 1982) that within the tonal hierarchy, the IV and V chords are closest to the I chord (and therefore more stable and structural), followed by the vi and iii chords, followed by the ii and vii° chords[20]:

---

[20] The stability calculations performed by the TPS algorithm involve several other factors; most notably, the perceived stability of a chord is also influenced by the perceived stabilities of neighboring chords. For the sake of

Table 5: TPS distance scores for diatonic chords in relation to tonic (I) chord

	I	ii	iii	IV	V	Vi	viiº
Fifths Distance	0	2	3	1	1	3	2
Common Tones	0	6	4	4	4	4	6
Total TPS Distance	**0**	**8**	**7**	**5**	**5**	**7**	**8**

Like the hierarchies of pitches in the diatonic scale, chordal hierarchies are an important part of the cognitive structure a listener must have in place in order to "understand" and interpret the differential stabilities among the sonic events in a composition. The differing functions and points of stability that emerge as properties of a listener's atemporal knowledge of the tonal system help the listener construct a mental representation of a musical composition and to perceive cycles of tension and resolution among what would otherwise be combinations of meaningless sound segments.

4.3   EVENT HIERARCHIES

Because the musical parser makes decisions of structural importance primarily within local contexts, the parser must have a way of segmenting the musical surface into temporal domains, or time-spans. GTTM characterizes time-span segmentation as sensitive to two organizational factors: grouping structure and metrical structure. Grouping structure refers to how musical events are perceived as related to one another according to intuitions of similarity or coherence to form whole motifs and phrases. Perhaps coincidentally,[21] grouping seems to be guided by principles of constraint optimization, an observation which led a team of researchers to develop an OT model of grouping structure (van der Werf & Hendriks, 2004). While the grouping process is not the focus of my analysis, we will see that the OT model is an interesting way of understanding this somewhat rudimentary component of musical parsing. Metrical structure describes the relationship between the musical events and the emergent rhythmic patterning of

---

explanatory simplicity, only the most significant factors are described here. The approximations they provide are precise enough to be consistent with basic musical intuitions.

[21] The optimality-theoretic character of grouping may not be coincidental at all if we are to find that optimization is a property of more general cognitive process.

strong and weak beats in a composition. In the following sections I will elucidate how grouping and metrical structure interact to form time-spans, which will become a fundamental component of the OT model to follow.

4.3.1 Grouping Structure

Like language and many other areas of human cognition, the perception of a musical composition involves segmenting the stream of sound into groups according to organizational principles. These groups are hierarchical; a two- or three-note motif may constitute a group that is embedded in a theme, which is embedded in a phrase, which is embedded in a larger phrase, and so on, with the largest group being the entire musical composition. As is the case for language, this intuitive process of segmenting the sound signal into meaningful groups is necessary for syntactic processing because it provides structure and context for interpretation of how the individual elements relate to one another.

The principled segmentation of a musical surface into groups seems to proceed in an optimality theoretic manner. GTTM proposes a series of preference rules that describe the process of grouping in music. These rules are "soft," meaning they define a preferred or likely force in the grouping process, but in cases where grouping rules conflict, one rule may be violated in favor of another. For example, consider the phrase in Figure 12:

Figure 12: Example exhibiting multiple possible grouping analyses

Using the Gestalt principles of grouping for visual perception, it is easy to see in the musical notation very similar principles that affect the perception of the auditory signal. Because they are separated by a rest, or a silent pause, the first three notes are easily heard as a group and the second set of three notes is heard as a second group. However, considering the articulation markings, it may be feasible for the first

two notes to be heard as a group because they are slurred together (meaning the notes are connected and the transition between them is as smooth as possible, like slurred speech), while the last four notes may be heard as belonging to a group because they are all staccato (very short, abrupt bursts of sound). Although both of these factors influence the perception of groups, virtually all listeners will hear Figure 12 as comprising two groups of three notes each, suggesting that the force of temporal proximity is stronger than articulatory similarity.

Recognizing that grouping preference rules have variable strengths and often conflict with one another to result in an optimal interpretation, van der Werf and Hendriks (2004) develop an Optimality Theoretic approach to GTTM's grouping principles. The researchers reformulate GTTM's grouping preference rules as violable constraints and propose a ranking of these constraints. These reformulated constraints are:

(8) GPR 1 SINGLES: Groups never contain a single element.

(9) GPR 2a PROXIMITY SLUR/REST: No group contains a contiguous sequence of three notes, such that the interval of time from the end of the second to the beginning of the third is greater than that from the end of the first note to the beginning of the second.

(10) GPR 2b PROXIMITY ATTACKPOINTS: No group contains a contiguous sequence of three notes, such that the interval of time between the attackpoints of the second and the third note is greater than that between the attackpoints of the first and the second note.

(11) GPR 3a CHANGE REGISTER: No group contains a contiguous sequence of three notes, such that the interval from the second note to the third is bigger than that from the first note to the second.

(12) GPR 3b CHANGE DYNAMICS: No group contains a contiguous sequence of three notes, such that the first two share the same dynamics, different from the third.

(13) GPR 3c CHANGE ARTICULATION: No group contains a contiguous sequence of three notes, such that the first two share the same articulation, different from the third.

(14) GPR 3d CHANGE LENGTH: No group contains a contiguous sequence of three notes, such that the first two share the same length, different from the third (van der Werf & Hendriks, 2004: 4)

To exhibit how some of these constraints may interact in the determination of grouping structure of a musical sequence, van der Werf and Hendriks identify several possible interpretations wherein a musical

line consisting of five notes may be perceived as grouping in different ways, indicated with brackets in

the following tableau:

Table 6: Tableau of grouping interpretations (Adapted from van der Werf & Hendricks, 2004: 6)

	GPR 1: SINGLES	GPR 2a: PROXIMITY SLUR/REST	GPR 3b: CHANGE DYNAMICS
a.	*		
b.		*	
c.			*

Grouping (a) violates constraint GPR 1 because it allows a single note to comprise its own group.

Grouping (b) violates GPR 2a because the amount of time between the end of the F and beginning of the

E (which is essentially zero seconds) is much less than the amount of time between the end of the E and

the beginning of the E-flat. Finally, grouping (c) violates GPR 3b because it does not align with the *piano*

and *forte* dynamic markings. Van der Werf and Hendriks determine an overall constraint ranking by

asking a number of individuals to make grouping judgments on musical stimuli that violate the grouping

constraints in different ways. Using a computer program developed by the researchers, the following

hierarchy of grouping constraints is proposed to reflect the relative strengths of each constraint in

determining group boundaries in musical sequences (2004: 7):[22]

---

[22] The researchers did not include GPR 3c CHANGE ARTICULATION because of methodological difficulties.

(15) GPR 2a PROXIMITY SLUR/REST
        >> GPR 3d CHANGE LENGTH
            >> GPR 3b CHANGE DYNAMICS
                >> GPR 1 SINGLES
                    >> GPR 2b PROXIMITY ATTACKPOINTS
                        >> GPR 3a CHANGE REGISTER

In addition to the rules explored here, GTTM gives several more grouping preference rules that

apply to higher-level groups; converting these rules into constraints may constitute an interesting

extension of van der Werf and Hendriks' analysis. While exposing the optimality theoretic character of

the grouping component of auditory perception is not the focus of my current hypothesis, van der Werf

and Hendriks' successful application of OT to GTTM'S grouping preference rules may support the claim

that music-syntactic processing proceeds in an optimality theoretic manner. Regardless of their OT

nature, grouping principles are an important aspect of musical listening, and we will see in the following

sections how these principles interact with other components of musical structure during parsing.

4.3.2 Metrical Structure

Metrical structure refers to a regular framework of strong and weak beats underlying a musical

composition. The meter is usually easily intuited by listeners, who may clap their hands or tap their feet

along to the beat. The perception of meter is dependent on an expectation that the music will continue to

conform to the established pulse, a phenomenon which allows us to sense a rhythmic beat even in brief

moments of complete silence. Meter is perceived as being organized hierarchically into patterns of strong

and weak beats wherein some beats may be heard at higher and higher levels of abstraction as more

prominent or stronger than others. For example, in "The Star Spangled Banner", almost every word can

be heard as its own beat; the longer notes that occur on "see" and "light" are sustained over two beats. At

a higher level of metrical organization, the beats of "say," "see," "dawn's" and "light" are easily

perceivable as stronger than the beats at the first level. An even higher level of metrical structure may

only contain the beats of *say* and *dawn's*, and so forth. This pattern is reflected in the conventions of

musical notation; the beginning of a measure indicates the beginning of a metrical group, and the first

beat of the measure is perceived as strong. On a larger scale, measures may be grouped together, so that the first beat of every four measures is stronger.

Like grouping structure, the perception of metrical structure is a cognitive process that the listener instinctively employs to organize and "make sense" of acoustic, temporal material. The hierarchical patterning of strong and weak beats can be depicted as a metrical grid; for example Figure 13 shows a metrical grid for a composition in 3/4 meter (a), such as "The Star Spangled Banner", and the common 4/4 meter (b).

Figure 13: Metrical grids for (a) 3/4 time and (b) 4/4 time

Although the metrical grids here only show a few levels of hierarchy, it should be easy to see that the span of time between each main beat (tactus) can be divided into increasingly smaller intervals, such as eighth notes (shown in the grid), sixteenth notes, thirty-second notes, and so on. Every attack point is associated with a beat at some level of metrical structure, the beats at each level are equally spaced (throughout time), and every beat at a given level is a beat at every smaller level. Beats at each successive level are perceived as stronger than those at lower levels that are not projected to a higher level. The metrical framework is a key component of musical parsing, because it provides a context of strong and weak beats within which the pitch material can be interpreted.

The perception and establishment of a sense of meter is an interesting and complex cognitive process in and of itself; while nearly every listener can easily clap his hands or tap his foot along to a composition, it is more difficult to identify what precisely it is about the musical events that let us do so. While we often associate metrical beats with the onset of pitch events, it is clear that this is not always the

56

case. For example, we hear long, sustained notes as spanning over several beats (and we can sense the regularity of those beats even if there is no actual physical instantiation of them while the note is sustained), or we might sense a metrical beat in a moment of complete silence in the music. Consider the swinging melody from the popular song by The Rolling Stones, "I Can't Get No Satisfaction" (1965), wherein most of the notes occur *between* metrical beats (at the main "foot tapping" level) rather than on them.[23] A composition's meter is often so robust that even deviations such as temporary uneven spacing or tempo inflections (such as *rubato*) do not dramatically disrupt the organization. While the cognitive process responsible for the discernment of a metrical structure will be taken for granted in the current discussion, it is important to recognize that these processes are an inextricable part of musical parsing.

4.3.3   Time-Span Segmentation

Through the interaction of grouping structure and metrical structure, we can conceive of the notion of time-spans, which represent the segmentation of a musical surface into domains over which the parser may make judgments of stability and structural importance. GTTM makes the following claims about time-spans: 1) Every time-span has a musical event that is its structural head, and all other events in that time-span are subordinate to the head; 2) Each time-span is immediately contained within a larger time-span; and 3) The heads of lower time-spans compete at each successive level for headedness in larger time spans. At each level of time-span, every musical event is a potential candidate for structural head; the parser assigns structural importance to one of the candidates according to a set of principles, which GTTM refers to as preference rules. These rules determine which of the structural interpretations are most likely to correspond to an enculturated listener's hearing of the piece.

Time-span organization relies heavily on the established sense of meter, but grouping considerations also influence the time-span framework. At local levels, time-spans are divided so that they each represent a more-or-less equally sized segment of music, corresponding to the established

---

[23] Since meter is hierarchical, we would find that these in-between notes are actually occurring on a metrical beat at a more finely divided level of meter.

metrical organization. The lowest level of time-span segmentation is at the level of the smallest note duration. For example, the lowest level of time-span in Figure 14 is actually the eighth note (although this level of time-span is only indicated where the eighth note divisions actually occur, for simplicity and clarity). The next larger level is the quarter note, followed by the half note, and so forth.

Figure 14: Example of a time-span hierarchy in the theme from Beethoven's Symphony No. 9 in D minor, Op. 125 ("Ode to Joy;" Lerdahl & Jackendoff, 1983: 125)

This meter-based segmentation may continue for several levels, but at higher levels of abstraction we begin to see a need for another method of segmentation. Consider Figure 15, an excerpt from the well-known "Ode to Joy" theme from Beethoven's Symphony No. 9 in D minor, Op. 125:

Figure 15: Example of a time-span hierarchy with out-of-phase grouping and metrical structure (Theme from Beethoven's Symphony No. 9 in D minor, Op. 125; Lerdahl & Jackendoff, 1983: 126)

The F-sharp on beat 4 of the second measure should be considered and analyzed intuitively as part of the phrase in the third and fourth measures instead of that in the first and second, because it is an early presentation of the first note of the main motive. In situations such as this, considerations of meter in

time-span segmentation yield to considerations of grouping. Oftentimes these perceptions align with metrical structure (i.e., the beginning of a group falls on a metrically strong beat), but sometimes they do not. In Figure 15, the F-sharp on beat 4 of the second measure is grouped with the following measure because it shares registral and articulatory similarity. This grouping is reflected in the time-span segmentation with an intermediary level in which the F# is incorporated into the following half-note level time-span,[24] an adjustment that percolates up each successive level in the hierarchy.

The process of musical parsing involves the selection of the structurally most important element in each time-span (in terms of optimal constraint satisfaction). The selection begins in the second-lowest level (since the very lowest level contains only one pitch-event per time-span and therefore involves no competition) and cycles upward to the highest level of time-span (containing the entire composition). At higher levels of time-span, the structural heads of lower-levels contained within that time-span compete for structural head of the higher-level time-span. In terms of modeling and notation, these structural hierarchies may be represented here in different ways, depending on the specifics of the musical example and analysis. Figure 16 shows two different ways of conceiving of and formalizing the structural hierarchy of a musical phrase. The tree notation in (a) is similar to that used by generative linguistic theory. Each upward extension of a branch indicates that the corresponding pitch event is the head of its local time-span. Branches that do not terminate but continue upward are heads of larger time-spans. The multiple-system notation in (b) shows the same information that is depicted by the tree notation but in a way that may be more intuitive. Each successive level of abstraction should sound like a simplified, pared-down version of the previous level. By the same token, the events that are omitted at each successive level of abstraction are heard more or less as elaborations of the more structural events. Both methods of representation convey the same information about structural interpretations.

---

[24] Lerdahl and Jackendoff describe a methodology for the demarcation of time-spans, including well-formedness rules for truncating or augmenting meter-based time-spans to accommodate groups that span over metrical boundaries (1983: 146-178).

Figure 16: Methods of modeling structural hierarchies in music (Bach's "O Haupt voll Blut und Wunden" from St. Matthew Passion, BWV 244; Lerdahl & Jackendoff, 1983: 115)

4.4 CONSTRAINTS ON STRUCTURAL INTERPRETATIONS

Following standard OT methodology, I will propose a number of constraints that dictate which of the possible structural interpretations corresponds to an enculturated listener's hearing of a piece. Many of these constraints will be adaptations of GTTM's preference rules for time-span reductions. Like the rules for grouping, GTTM's preference rules conflict with one another, making it impossible for all rules to be satisfied all of the time. Lerdahl and Jackendoff claim that these preference rules "have no counterpart in standard linguistic theory" (1983: 9), but we will see that these preference rules do actually serve a

theoretical and formal function similar to the violable constraints of linguistic OT. Faced with a paucity of

empirical data on the types of grammatical judgments needed for such a sophisticated account of musical

syntax, these constraints on Western tonal syntax are largely motivated here by informed intuition (my

own and that of the esteemed linguist/music theorist duo Ray Jackendoff and Fred Lerdahl). However,

they should be testable as tonal constraints on grammaticality by assessing the judgments of listeners

without any formal musical training.

Each time-span typically has several possible structural interpretations that assign headedness to

different musical events, and each of these interpretations adheres to constraints to different degrees. This

means that constraint conflict and evaluation occur on inherently different levels of analysis. We will see

that the lower levels of time-span contain the most possible constraint violations and interactions, so these

levels will be the focus of analysis here. An important direction of future research will be to further

investigate constraint interactions and violations across the different levels of hierarchy in the time-span.

4.4.1 Metrical Position

The meter of a musical composition has a fundamental influence on its structural interpretation. Once the

meter is established, musical events that occur in strong positions within the metrical hierarchy will be

preferentially treated as contributing to the structure of the phrase. This constraint on structural

interpretations can be stated as follows:

> (16) METPOS (Metrical Position): The head of a time-span is the pitch event that occupies the
>
>   metrically strongest position in that time-span.

This phenomenon has a "bootstrapping" effect in both composition and the discernment of metrical

structure: The perception of pitch events as structurally important contributes to the establishment of a

metrical framework that optimally puts these events in strong positions. Once this framework is in place,

the metrical position of certain pitch events then contributes to judgments of their structural importance.[25]

Pitch events that occur in metrically stable positions are more likely to be considered structural heads; by the same token, structural heads are likely to be found in stable metrical positions.

Because the metrical hierarchy includes several levels of strength or stability, a beat that is metrically stable in a local time-span may be relatively less metrically stable in a higher level time-span that also includes that beat. Consider Figure 17 in which, like previous examples, the dots represent beats in successive levels of the metrical grid and the brackets indicate time-spans. (Note that time-spans do not always perfectly align with the metrical grid, as we saw previously in Figure 15.) At time-span level *b*, beats 1 and 3 are metrically stronger than beats 2 and 4 respectively. Assuming beats 1 and 3 are selected as structural heads of their respective time-spans, they are then compared at time-span level *a*, where we see that beat 1 is stronger than beat 3 and will therefore be preferred (according to METPOS) as structural head of that time-span.
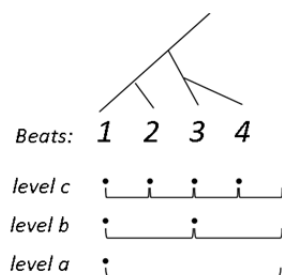
Figure 17: Metrical hierarchy

Metrical position is a property of a musical event independent of its tonal characteristics. Because tonal characteristics will almost always influence judgments of structural importance, it can be difficult to isolate the effects of metrical position on these judgments. However, these effects may be observed by comparing the legitimacy of candidate interpretations of a passage that is somewhat tonally

---

[25] Like Grouping structure, the perception and establishment of a metrical structure may proceed in an optimality-theoretic manner (see Katz & Pesetsky, 2011) and may actually occur in parallel with (or as part of) the structural interpretation process.

static, mitigating the effects that issues of tonality may have on structural judgments. Consider the melodic line in Figure 18, which is reduced to structural head candidates at the quarter note level of time-span and then the half note level of time-span. The reduction in (a) assigns a hierarchy of structural importance to pitch-events that occur in increasingly metrically stable positions, while reduction (b) favors pitch events in metrically weaker positions. A listener who hears the eighth note, quarter note, and half note versions in succession will unequivocally agree that the quarter note and half note reductions in (a) are more representative of the musical structure than those in (b), showing that the pitch events that occur in metrically stable positions are favored as structural heads.
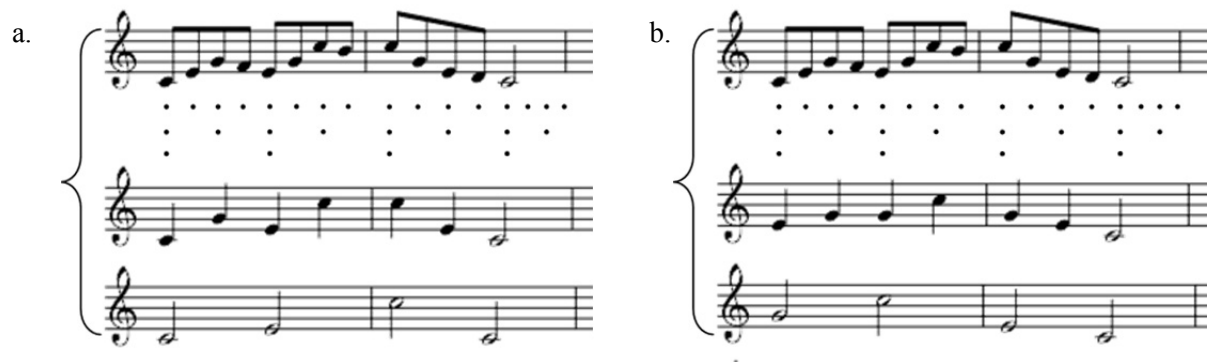
Figure 18: Structural reductions of phrase favoring pitches in (a) metrically strong positions and (b) metrically weak positions

4.4.2 Consonance

Another musical property central to Western tonality is the notion of consonance, which was described previously as the degree to which a note or group of notes fit together and into the surrounding context. A consonant chord fits quite well with surrounding elements and is pleasing to the ear, while a dissonant chord does not fit and results in a jarring or displeasing effect. While dissonance can be found to different degrees in Western tonal compositions, it is the consonant sonorities that allow a listener to find a tonal center. These sonorities are most meaningful and structural within the tonal context, suggesting the following constraint:

(17)   CONSON (Consonance): The head of a time-span is the most consonant pitch event in that

time-span.

Chordal consonance is manifested in two dimensions, as intrinsic consonance and tonal

consonance. Intrinsic consonance refers to the consonance of the interval content of a given chord. To

some extent, intrinsic consonance is easy to intuit. Imagine sitting at a piano and pressing a cluster of

white and black keys, all very close together; these notes likely sound very dissonant, while playing notes

that are two or three keys apart creates a much more pleasing sound. This is one reason why the triad

(notes that are approximately two white or black keys apart on the piano) is the basic chordal unit in

Western tonal music. Seventh chords, which comprise a triad with an added third, are also frequently

found in tonal music; these chords are considered somewhat unstable, because they create a dissonant

half- or whole-step interval between the seventh of the chord and the root. Major and minor triads in the

root position are most consonant, followed by triads in first inversion and then second inversion.

Tonal consonance refers to the tonal distance between a chord and the tonic chord of the local

key; one way of conceiving of and measuring this distance is through Lerdahl's Tonal Pitch Space theory

(2001), discussed previously in section 4.2.2. Chords that are tonally close to the tonic (i.e., V and IV)

sound more consonant than chords that are tonally distant and as such are considered more stable and

more structural. Chords that are tonally distant from the tonic, such as the iii and vii$^{\text{o}}$ chords, are

perceived to be unstable and require resolution to a more stable sonority; because these chords are

dependent on their resolutions, they are often perceived as more ornamental than structural.

Intrinsic dissonance in Western tonal music creates a sense of sonic tension for the listener. For

example, the opening theme to *The Simpsons* uses chord clusters to create a bizarre, almost sinister sound

reminiscent of a train horn. The consonance differential of chords in root position versus their inversions

is somewhat more subtle and difficult to observe, but consider the following simple example from the

traditional folk tune "The Farmer in the Dell" (Fig. 19). The very last chord is typically a I chord in root

position, but if this chord is changed to be a I chord in first inversion, the result is a dramatically more

unstable final chord. The final chord is made even more unstable if it is modified to be in second

inversion. A primary reason for this difference in perceived stability is the chords' respective inversions: A second inversion chord is less stable (and less intrinsically consonant) than the same chord in root position.

a.

Most stable:
Root position

b.

Less stable:
1ˢᵗ inversion

c.

Least stable:
2ⁿᵈ inversion

Figure 19: Variable stability of the final chord of folk tune "The Farmer in the Dell" (Traditional; harmonization adapted from *Ultimate Children's Songbook*, 2000)

Another particularly interesting example from popular music exploits the sense of stability created by the consonance of the tonic note. In the opening phrases of Adele's Grammy-winning song "Someone like You" (Adkins & Wilson, 2011, Fig. 20), the key of A major is set up unequivocally with a typical, functional chord progression and a vocal line that outlines the tonic triad (E, C-sharp, A). The A occurs on the word "you're," but most listeners will agree that while there is a moment of weight or stability on that note, the effect is confounded by the note's unconventional placement in an odd metrical position (a musical score places the note onset on the second sixteenth note division of the fourth beat),

65

and by the "resolution" (although not a resolution in the traditional sense) of this stable pitch to a highly unstable F-sharp on the downbeat of the last measure in the example below. This same pattern repeats four words later on the word "you," and again five words later on "you're," underscoring the composer's exploitation of syntactic phenomena (the stability of the tonic note juxtaposed with its unstable positioning and resolution) to mirror a sentiment in the lyrics, namely that the subject of the song is an individual about whom the singer has conflicting emotions.

Figure 20: Excerpt from Adele's "Someone like You" (Adapted from Adkins & Wilson, 2011)

4.4.3 Registral Extremes

Musical events that are particularly salient because they include registral extremes (i.e., pitches that are significantly higher or lower than surrounding pitches) are more likely to be interpreted as structurally important. This phenomenon can be articulated as the following constraint:

> (18) REGEXT (Registral Extremes): If a time-span contains a pitch event that has a registral extreme (high or low), that pitch event is the head of that time-span.

Lerdahl and Jackendoff (1983) point out that while the effect of a registral extreme is rarely strong enough to dictate structural judgments, it can have a significant strengthening or weakening influence on those judgments. For example, we may consider the candidates for structural head of the passage in

Figure 21, an excerpt from Mozart's Symphony No. 40 in G minor, K. 550, to be beat 1 of the first measure and beat 1 of the second measure (indicated with arrows). The candidates have the same pitch content (D and G) and are both in metrically strong positions. Beat 1 of measure 1 is in a stronger metrical position than beat 1 of measure 2, so we may first analyze this event as the structural head. However, by listening to the passage it is clear that the low G on beat 1 of measure 2 has an anchoring effect, adding perceptual salience to the event and promoting it to a position of structural importance. In the context of this registral extreme, the preceding material is interpreted as a precursor to the structural head on beat 1 of measure 2, despite the event's slightly weaker metrical position.
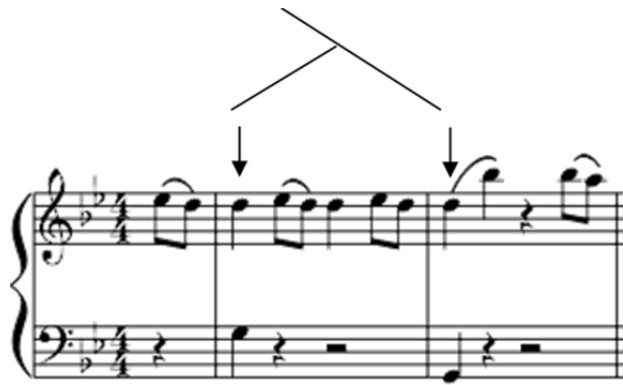
Figure 21: Theme from Mozart's Symphony No. 40 in G minor, K. 550 (Lerdahl & Jackendoff, 1983: 163)

It is not unusual for composers to exploit the effects of registral extremes on perceptions of structure to manipulate a listener's interpretation of a composition. MacKay (2003) describes Haydn's technique of using registral extremes to mark significant formal events, sometimes in the midst of contrary metrical or tonal cues. For example, in Figure 22 the second movement of Haydn's Keyboard Sonata in D major, Hob. XVI: 37, registral extremes are used to draw attention to and emphasize the F's at the end of this opening phrase. At this point, the composition is squarely in D minor, but the composition eventually shifts its tonal center from D minor to F major, and Haydn's emphasis of F in this phrase helps to bring it to the structural forefront in preparation for the shift.

Figure 22: Haydn's Keyboard Sonata in D major, Hob. XVI: 37, 2nd mvmt (MacKay, 2003: 131)

Still another example of the effect of registral extremes on the perception of structural importance can be seen in the well-known folk tune "Londonderry Air" (also known as "Danny Boy"), given in Figure 23. Consider the jump in the melody in the second half of the first full measure, from the second-space A to fourth-line D. This pronounced jump is part of what lends the melody its characteristic quality and as such makes the D significant to the overall structure of the phrase. If the D were omitted or changed to a passing tone B-flat, the average listener would likely agree that the structural integrity of the phrase had been compromised more dramatically than if, say, the G of the same measure were omitted or changed.

Figure 23: Beginning of theme from "Londonderry Air" (Traditional)

Because registral extremes may occur in either the upper- or lowermost voices, it may be helpful to conceive of the REGEXT constraint as two separate constraints, one for registral extremes in the melody, and one for registral extremes in the bass:

(19) REGEXT[MEL] (Registral Extremes - Melody): If a time-span contains a pitch event that has an upper registral extreme, that pitch event is the head of that time-span.

(20) REGEXT[BASS] (Registral Extreme - Bass): If a time-span contains a pitch event that has a lower registral extreme, that pitch event is the head of that time-span.

We can imagine situations where these constraints might conflict, i.e., when there is a registral extreme in the melody and one in the bass but they do not coincide. The excerpt from Mozart's Symphony K. 550 (Figure 21) is one such example – the lower registral extreme on the first beat of the second full measure does not align with an upper registral extreme on the second beat. In this case, we can intuit that the chord on the first beat of the measure is more structural, but it would be hasty to conclude from this interpretation that REGEXT[BASS] is more powerful than REGEXT[MEL] since there are other factors that may be at play. For example, it could be hypothesized that REGEXT[MEL] and REGEXT[BASS] are approximately equal in influence but also gradient, giving the greater intervallic distance between the bass notes dominance over the smaller leap in the melody. Perhaps the interpretation is influenced by the fact that the notes on beat one are in a stronger metrical position and constitute a more robust musical event (containing two, consonant sounded pitches instead of just a single pitch). While the particulars of the influence of registral extremes in the melody versus the bass should eventually be examined in detail, it will be sufficient for the current discussion to consider registral extremes in the melody and bass to have approximately equal influence on judgments of structure. As such, it will be useful to differentiate between REGEXT[MEL] and REGEXT[BASS] in analyses of musical examples, but potential rankings will group these together into a general REGEXT constraint.

4.4.4 Linear Stability

If a musical event is meaningful and well-formed within a melodic context, the parser is more likely to attribute structural importance to that event. This preference can be articulated as the constraint in (21):

    (21) LINSTABLE (Linear Stability): The head of a time-span makes an optimal linear connection with events in adjacent time-spans.

What makes a melodic line meaningful or well-formed? The answer to this question is likely complex and difficult to ascertain, but similarly to other types of sensory perception, the human mind seems to favor certain constructions and qualities such as symmetry, coherence, and good continuation. Examples of well-formed melodic lines include those that unfold in step-wise scalar motion, outline a triad (a common

compositional technique referred to as arpeggiation), are decidedly symmetrical, or follow an established

intervallic pattern (sometimes referred to as a melodic theme or motif). In his book *Musical Forces*,

Larson (2012) characterizes this tendency as a type of "musical inertia," analogizing a listener's possibly

innate preference for musical patterns to a physical force governing movement. The theory does not assert

that melodic lines that exhibit no meaningful patterns are necessarily particularly ill-formed, but rather

that melodic lines that do follow a particular pattern will be perceived as more perceptually salient and

robust.

Because well-formed constructions are more meaningful to the musical parser, the parser prefers

to assign structure to those elements that create a well-formed (or stable) melodic line with events from

neighboring time-spans. These well-formed lines can be local or abstracted away from the musical

surface so that notes that are quite distant from each other are picked out by the parser as forming a

musical unit. In theory, the well-formedness of a melodic line is likely gradient, with the *optimal* well-

formed line being most preferred by the parser as structurally important and less well-formed lines

incurring proportionately more constraint violations. In actual excerpts from Western tonal compositions,

however, we find that the identification of well-formed lines is relatively straightforward. It will be

sufficient for the current analysis to consider this constraint violation to be binary: If a well-formed line

can be derived from the musical surface, the parser prefers to assign structure to its constituents, and a

candidate interpretation that does not assign structure to one of the constituents incurs a violation.

Another simplification that I will make in the current analysis is to focus on the effects of linear stability

in only the (typically) most salient voices of a musical composition: the melodic line (usually the

uppermost voice) and the bass line. For clarity, like the REGEXT constraint, it will sometimes be helpful

to specify whether LINSTABLE is affecting the melody or the bass:

(22) LINSTABLE[MEL] (Linear Stability - Melody): The head of a time-span contains a melodic
    note that makes an optimal linear connection with melodic notes in adjacent time-spans.

(23) LINSTABLE[BASS] (Linear Stability - Bass): The head of a time-span contains a bass note
    that makes an optimal linear connection with bass notes in adjacent time-spans.

70

An example of the structural importance of higher-level melodic lines is the well-known tune "Greensleeves," discussed previously in Chapter 1. The local-level melody is quite contoured, turning upwards and downwards again every measure, but it is easy (and even automatic for most listeners) to pick out the higher level descending melodic line that begins on A in Figure 24 and ends on E. This perceptual effect of a well-formed descending line is that of a frame or skeleton for the entire phrase; the beamed notes are perceived as structural pillars around which the other more ornamental notes are organized.

Figure 24 (reprint of Figure 5): Higher-level linear connection in "Greensleeves" (Traditional; Cadwallader & Gagne, 1998: 23)

4.4.5 Cadential Retention

The primary motivation for positing a constraint that assigns structural importance to cadences follows the intuition that the ends of musical phrases or compositions are perceived as highly significant structural events. Phrases and compositions typically end on consonant notes or chords, but consonance alone cannot account for the elevation of phrase endings in the structural hierarchy, because consonant musical events also occur with regularity in the middle of phrases. A constraint on structural interpretation that reflects the parser's preference of assigning structure to cadential events is given in (24):

(24) CADENCE (Cadential Retention): A cadence is the structural head of its time span. Within a two-part cadence, the first element (usually a dominant chord) is subordinate to the resolution.

A cadence is a sequence of chords that conventionally informs the listener that the end of a musical phrase has arrived. The chord content of a cadence is highly constrained; most cadences comprise a V chord followed by a I chord (called an *authentic* cadence), while some proceed from V to vi (called a *deceptive* cadence because the listener expects the I chord instead of the vi). Cadences that end in a V chord are called *half* cadences, because the phrase ending sounds unstable and in need of resolution or continuation. The cadence as a marker of a phrase ending is a powerful force for enculturated listeners of Western tonal music. Within an authentic cadential progression, the V chord instills in the listener a need for resolution to the I chord that is so potent one might wonder if there truly are psychoacoustic forces at work rather than purely syntactic ones. A piece of music that ends with the V chord but does not resolve to the final I chord feels like it has been "left hanging." Even a listener completely untrained in music theory is likely to feel compelled to sing the missing final note, to alleviate the nearly psychosomatic discomfort of such a syntactically incomplete form.

A theory of structural interpretation must have a way of accounting for the fact that cadences are often made up of two chords that function together as a unit, and a chord's function as part of a cadence affects its structural role. However, not all V-I progressions are perceived to be cadential. GTTM proposes a number of conditions that must be met in order for the parser to recognize a cadential formula. First, the sequence of chords must form an authentic cadence (V to I), a half cadence (any chord to V), or a deceptive cadence (V to vi). The last chord of the cadence must also be at the end of a time-span that contains both elements. Finally the cadence must be part of a larger group for which it can serve as a structural ending; in other words, a V-I sequence that in itself makes up a group must be part of a larger group for which another group may serve as a structural beginning. If these conditions are met, then the cadence is preferred as a structural element.

Cadences are ubiquitous in Western tonal music, but what is most important to demonstrate in terms of syntactic constraints is the importance of interpreting the cadence as structural, even if other factors conspire against it. Consider, for example, the theme from Mozart's Piano Sonata in A major, K. 331 (Fig. 25). When the first phase ends in measure 4, the most structural element of the measure

according to the constraints discussed thus far seems to be the highly consonant tonic chord on beat 1. However, listening to the theme gives the listener the impression that the phrase clearly proceeds from tonic in measure 1 to the dominant (V) in measure 4, constituting a half cadence. The V chord does not occur until the very last eighth note of the measure, but its use as a cadential gesture elevates its structural importance.

Figure 25: Mozart's Piano Sonata in A major, K. 331 (Adapted from Lerdahl & Jackendoff, 1983: 32)

4.4.6   Attachment

From a theoretical and philosophical standpoint, it is important for a model of parsing to acknowledge that the parser does *something* with incoming material, and it may be argued that the following constraint follows naturally from the contextual nature of parsing and the notion of efficacy of the parser:

> (25)   ATTACH (Attachment): Every incoming element is subordinate to the preceding element within a group.

We take as given the well-formedness condition that every time-span requires a head but does not require elements that are subordinate to the head; given that a time-span can potentially comprise only one element, it is most efficacious for the parser to assign headedness to the first element it encounters in the time-span and consider subsequent elements as subordinate. The preference for optimal well-formedness of a structural interpretation is reflected in situations involving case ambiguity in language, as shown by

73

Fanselow (2005). Consider the German sentence *Welche Frau hat gestern der/den Mann angerufen?*

Until the parser receives the determiner of *Mann*, *der* or *den,* the case of *welche Frau* may be nominative

or accusative. When faced with this ambiguity, the parser seems to prefer a subject analysis of the wh-

phrase, because it fills the most essential (and most required) theta role slot for a potential predicate. The

Minimal Attachment constraint also reflects the efficacy of the parser; Frazier and Rayner describe

Minimal Attachment as follows: "Attach incoming material into the phrase-marker being constructed

using the fewest nodes consistent with the well-formedness rules of the language" (as cited in Gibson &

Broihier, 1998: 160). This preference also describes the conservative tendency of the parser to interpret

incoming material in the context of and subordinate to the material that has already been received.

ATTACH may not have visible effects at the local level of parsing, because an incoming event can

quickly be reinterpreted as locally structural if its constraint adherence makes it an optimal head.

However, the ATTACH constraint can account for higher-level phenomena, namely that structural

beginnings of phrases are perceived as hierarchically more important than other musical events. Consider

the overall structural representation (as analyzed in Lerdahl & Jackendoff, 1983) of the Bach Chorale "O

Haupt voll Blut und Wunden" from St. Matthew Passion, BWV 244, given in Figure 26. The most highly

structural events in this passage are the cadences (which we expect, given the CADENCE constraint),

depicted at (*a*); however, also occupying high levels of structural importance are the left-most elements of

many of the time-spans at the 2-bar, 4-bar, 8-bar, and 16-bar levels, depicted at (*b*). The preservation of

structural importance of each of these musical events cannot be fully accounted for by the constraints

presented thus far: All occur on metrically weak beat 4, few are linearly significant or contain registral

extremes, and, while all are highly consonant, there are other highly consonant musical events in the piece

that *do not* receive the same structural interpretation. By proposing the ATTACH constraint, the OT model

accounts for the intuitively empirical phenomenon that the beginnings of phrases are perceived as highly

structural and stable musical events.
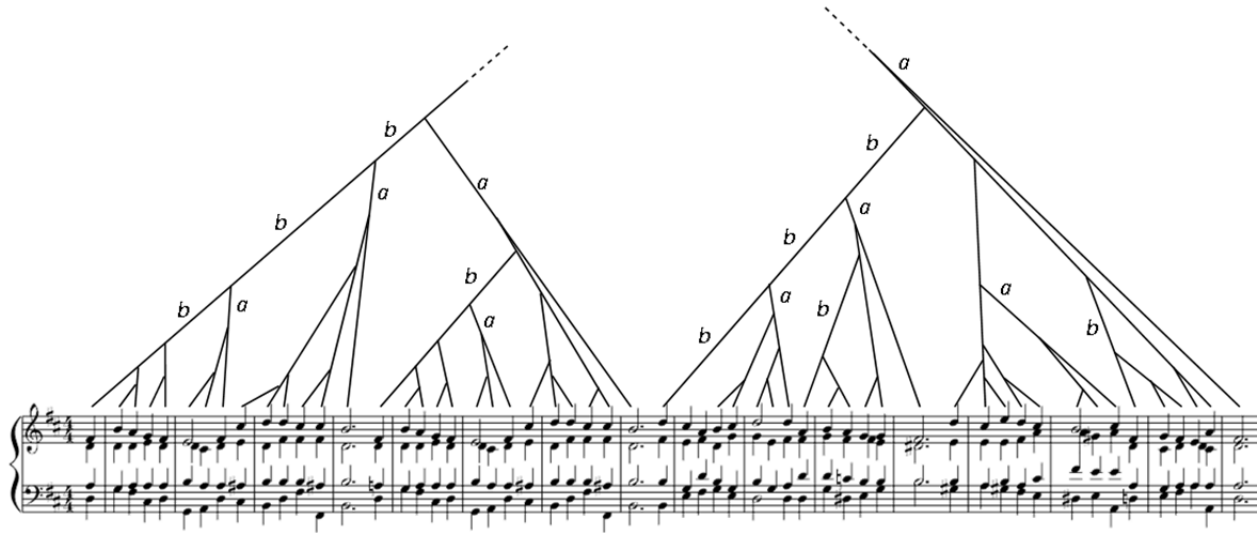
Figure 26: Hierarchical structure of Bach's "O Haupt voll Blut und Wunden" from St. Matthew Passion, BWV 244 (Adapted from Lerdahl & Jackendoff, 1983: 144)

4.5   CONSTRAINT INTERACTION

The proposed constraints are motivated primarily by observed patterns in Western tonal music, and the argument that a structural interpretation optimally adheres to these constraints follows naturally from an enculturated listener's introspection. In the following section, I will explore some of the ways these constraints conflict and interact in the selection of an optimal structural interpretation of a musical passage. In a typical linguistic OT analysis, the researcher attempts to isolate constraint conflicts so that a strict dominance ranking of constraints can be proposed. However, we will find in the following examples that identifying a traditional constraint ranking may not be a feasible or realistic way of developing an OT for musical syntax. Chapter 5 will explore the notion of weighted (rather than strictly ranked) constraints and whether this variation of OT can elegantly accommodate musical parsing phenomena. These sections are not intended to constitute a thorough OT analysis for Western tonal music but rather to explore the constraint interaction and optimization processes that seem to be at play in anticipation of and preparation for future research on the subject.

It is important to note that structural reductions of a musical surface, such as those used by music theorist Heinrich Schenker and more recently GTTM, are ultimately idealizations. Any number of

75

extragrammatical factors, such as inattention or memory limitations, may prevent a listener from interpreting a musical surface in precise concordance with the structure predicted by the model presented here. However, following similar arguments made by linguistic theorists, the idealistic quality of the model should not detract from its significance or importance for predicting a *likely* interpretation. Moreover, it should not detract from the assertion that structural interpretations are selected because they are the optimal candidates in terms of constraint violation.

A discussion of how the parser assigns structure to the opening phrase of Gotye's "Somebody That I Used to Know" (de Backer, 2011) is a good starting point for an analysis of constraint interaction in a musical setting, because this passage is relatively harmonically, texturally, and rhythmically simple. The first task at hand is to determine the structural heads of each level of time-span (Figure 27):

Figure 27: First excerpt from Gotye's "Somebody That I Used to Know" (de Backer, 2011)

Head selection at the quarter note level of time-span is fairly straightforward in the first measure, since each event essentially comprises, and is therefore head of, its own time-span.[26] We can easily identify structural heads of the eighth note figure in the second measure by eliminating each of the candidates and assessing how similar each modified passage sounds to the original; the candidate that seems to cause the most significant change to the passage is the more structural of the two. Some possible modifications are shown in Figure 28; in passage (a), the B-flat has been changed to A to force selection of the A as

---

[26] The eighth note figures in the bass clef are really just rearticulations of chord tones, so they are more textural than unique musical events.

structural head of time-span 1, and in passage (b) the A is changed to B-flat to force selection of the B-flat as structural head. Likewise, passages (c) and (d) reduce the candidates for head of time-span 2 to C and A respectively. Most, if not all, listeners will agree that modifying the A in time-span 1 and the C in time-span 2 more dramatically alter the passage than their respective competitors, so these events should be deemed structural heads of quarter-note level of time-span in our analysis.

Figure 28: Indications of structural importance in Gotye's "Somebody That I Used to Know"

The next step is to identify the structural heads at the next level of time-span, the half-note level. Table 7 (a) and (b) give possible structural heads (i.e., candidates) for the half-note level of time-span; while the first measure essentially only has one candidate head per time-span, the first half-note time-span in the second measure can possibly be headed by either the A or the C (which were both structural heads at the quarter-note level). Although reduction (b) is not a completely unsatisfactory interpretation, I argue that most listeners will judge reduction (a), which assigns structure to the pitch A, as more representative of the original passage:

Table 7: Tableau of candidate interpretations of first excerpt from Gotye's "Somebody That I Used to Know"

	LINSTABLE	METPOS ⌇ REGEXT
☞ a. (music notation)		*
b. (music notation)	*	*

The tableau also summarizes what these candidates tell us about possible constraint rankings. Winning candidate (a) obeys METPOS because the head pitch A is in a metrically stronger position than losing candidate pitch C; since candidate (b) assigns structure to an event in a metrically weaker position, it incurs a violation of METPOS. Candidate (a) also obeys LINSTABLE[MEL] because the A creates a strong linear constituent with the preceding structural heads F and G; since candidate (b) does not follow this optimal melodic line, it incurs a violation of LINSTABLE[MEL]. However, the pitch C is a registral extreme, so selection of the pitch A as structural head incurs a violation of REGEXT[MEL]. At this point in the analysis, we only know that satisfaction of *either* METPOS or LINSTABLE has resulted in the selection of candidate (a), so we can conjecture that at least one of these constraints outranks REGEXT. I have notated this in the tableau using a dotted line between LINSTABLE and METPOS, which conventionally signals that we are unsure of the ranking between these constraints, and a "squiggle" line between METPOS and REGEXT, which is my own method of indicating that either one or the other constraints to the left of the squiggle line could possibly still be ranked lower than REGEXT.

Because the winning candidate obeys both METPOS and LINSTABLE, it is difficult to know from this analysis which of the two constraints outranks REGEXT (or whether both do) and is responsible for

this structural interpretation. However, consider the passage in Figure 29, a variation of the opening

theme that occurs just several measures later in the introductory material:

Figure 29: Second excerpt from Gotye's "Somebody That I Used to Know"

Interestingly, although the passage is motivically and melodically parallel to the opening theme, adjusting

the first note from F to C changes the overall structure: Now the C in the second measure can more

reasonably be heard as structural head of its time-span (at the quarter-note level) than the A. Because C is

in a metrically weaker position than the A, this structural interpretation violates METPOS. However, we

are again in a situation where it is unclear whether the selection of C is motivated by LINSTABLE[MEL]

(since the line from C to G to C and back to G outlines a highly salient, symmetrical fifth-interval pattern)

or REGEXT (since the C is a registral extreme, especially now that the lower bound of the phrase is an

octave below it). [27] These observations are conveyed in the tableau in Table 8. Like the previous

example, we see here that one constraint seems to be outranked by one or both of the other constraints,

but we do not yet have enough information to determine a precise ranking.

---

[27] This analysis may be debatable, since there are still strong factors supporting the interpretation of the A as
structural head, such as its metrical position and consonance. It is important to emphasize that this structural
interpretation is idealized, i.e. the structure assigned without other interfering factors. In the context of the actual
song, having heard the opening figure first followed by this figure may actually affect listeners' interpretations, as
they would likely assign parallel structures to the passages. This "priming" effect would be an interesting extension
of the current analysis.

Table 8: Tableau of candidate interpretations of second excerpt from Gotye's "Somebody That I Used to Know"

		LINSTABLE	REGEXT	METPOS
☞	a.			*
	b.	*	*	

We might use logical reasoning to deduce possible constraint rankings of the three implicated constraints thus far. It is important to note that a strict ranking of constraints with respect to one another may not even be possible, but the process of attempting such a ranking helps to elucidate how the constraints interact to influence structural interpretations. To summarize, the first example showed that either METPOS or LINSTABLE (or both) likely outranks REGEXT, and the second showed that either LINSTABLE or REGEXT (or both) likely outranks METPOS. Figure 30 shows how an overall ranking of these three constraints can be narrowed down to two possibilities, LINSTABLE >> METPOS >> REGEXT *or* LINSTABLE >> REGEXT >> METPOS:

Possible Rankings:

First Example
- METPOS >> REGEXT
- LINSTABLE >> REGEXT

Second Example
- LINSTABLE >> METPOS
- REGEXT >> METPOS

Overall
- LINSTABLE >> METPOS >> REGEXT
- LINSTABLE >> REGEXT >> METPOS
- x
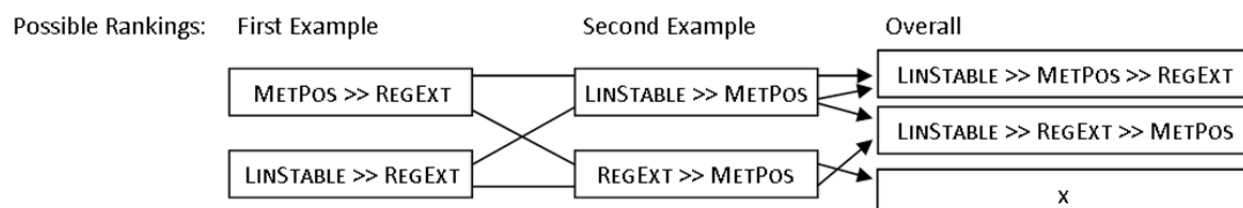
Figure 30: Synthesis of constraint rankings from first two examples

Since we may conclude that LɪɴSᴛᴀʙʟᴇ likely outranks both MᴇᴛPᴏs and RᴇɢExᴛ, the next step for a complete OT analysis would be to determine an ordering of MᴇᴛPᴏs and RᴇɢExᴛ (or to determine if such an ordering is even possible). Examples such as Mozart's Serenade No. 13 for strings in G major, K. 525 ("Eine Kleine Nachtmusik") suggest that RᴇɢExᴛ, can be more powerful than MᴇᴛPᴏs, somewhat contrary to Lerdahl and Jackendoff's assertion that RᴇɢExᴛ is rarely a deciding factor; the reduction in Figure 31 reflects the structural prominence of the Bs and Cs, despite their relatively weak metrical positions:

Figure 31: Reduction of Mozart's Serenade No. 13 for strings in G major, K. 525 ("Eine Kleine Nachtmusik")

The plausibility of this judgment is due to the height position of these notes at the apex of the melodic contour; that the notes are approached by a leap emphasizes their salience.[28] However, I will not seriously attempt an ordering of RᴇɢExᴛ and MᴇᴛPᴏs here for a number of reasons regarding the nature of these constraints. At this point in my explorations of structural interpretations, the RᴇɢExᴛ and MᴇᴛPᴏs constraints are largely motivated by observed patterns and informed intuition, but without computational analyses of musical passages and listeners' interpretations we can only approximate their influence on

---

[28] This salience is further emphasized by the juxtaposition of the melodic line created by the registral extremes (B-C-B-C) and a concurrent melodic line that mirrors it (G-F#-G-F#).

structure. The influence of REGEXT especially is relative to the complex characteristics of the musical surface. For example, within the context of a melody that has many leaps and spans an octave or more, a leap of a fifth or an octave may not be extreme at all; on the other hand, in a more conservative passage that moves primarily by stepwise motion and stays within a smaller register, a leap of a third or a fourth can significantly increase the structural importance of a musical event. A more precise characterization of these constraints is an interesting and important subject for future investigations of the optimizing nature of the musical parser, but for the current discussion I will consider these constraints to be dominated by LINSTABLE but unranked in respect to each other.

As another example of constraint interaction, consider how the parser proceeds upon hearing beats 3 and 4 of the first full measure of Bach's "O Haupt voll Blut und Wunden," boxed in Figure 32:

Figure 32: Excerpt from Bach's "O Haupt voll Blut und Wunden" from St. Matthew Passion, BWV 244

This figure indicates the chord content as well as a metrical grid denoting the hierarchy of beat strength. (The superscript x in the chord label I^x indicates that this chord contains a non-chord tone [E], i.e. a note that does not belong to the triad.) The parser has several options for interpreting the structural hierarchy of the selected segment, three of which are depicted in Figure 33. For simplicity of demonstration, these descriptions will select only the first and second most structural pitch-events in the hierarchy. The structural head of the entire segment is indicated by an upward-extending branch. A secondary (and

subordinate) head is indicated by a branch that terminates at a relatively high point on the branch of the primary head. The remaining two pitch-events, indicated by a dotted line, will be considered elaborations that are approximately equal in structural importance (subordinate to the primary and secondary heads).
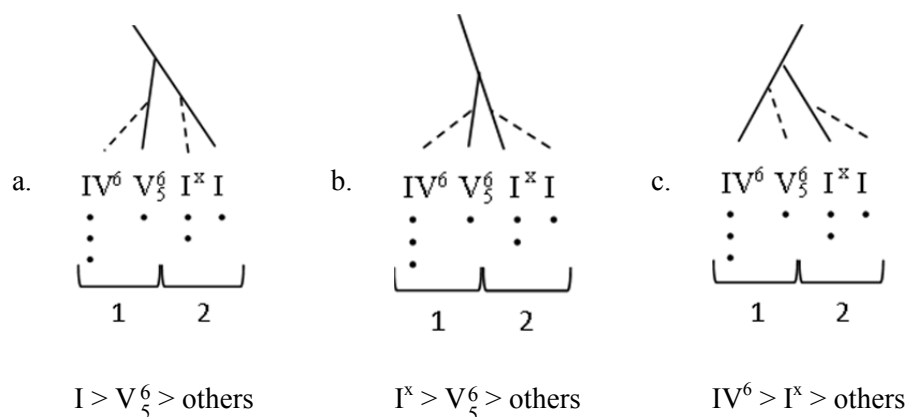
Figure 33: Possible structural interpretations for measure 1, beats 3 & 4 of Bach's "O Haupt voll Blut und Wunden" from St. Matthew Passion, BWV 244

While all possible permutations of a stability hierarchy are theoretically subject to *EVAL*, I have only listed candidate structures that seem reasonably plausible according to musical intuition. Parse (a) is plausible because it gives structural preference to the consonant chords in the highly salient and well-formed V-I progression. While the parser soon learns that this progression is not an actual cadence (because it does not articulate a group ending), it is certainly cadence-like and is therefore a reasonable candidate for structural headedness. For the same reason, parse (b) is intuitively plausible; however, this parse assigns structure to the less consonant I^x chord, which represents the onset of the tonic sonority, rather than the consonant I chord on the last half-beat of the measure. Finally, parse (c) is plausible because it gives structural importance to the pitch-events that occur on what has, up to this point, been established as the main beat, or tactus.

Lerdahl and Jackendoff (1983) judge the parse given in (a) of Figure 33 as the correct interpretation of the structural hierarchy of the selected segment. This judgment is consistent with my

own intuitions about the musical structure, and may be reinforced by listening to the entire phrase reduced at increasingly higher levels of the structural hierarchy, given in Figure 34. The listener will find that the reductions produced from the interpretation shown in Figure 33 (a) sound the most similar to the original phrase and retain its structural skeleton at all levels, confirming that this analysis reflects the enculturated listener's intuitions about and interpretation of the musical structure.[29]

Figure 34: Structural reductions of measures 1 & 2 of Bach's "O Haupt voll Blut und Wunden" from St. Matthew Passion, BWV 244 (Adapted from Lerdahl & Jackendoff, 1983: 115)

Two of the structural interpretation candidates, winning candidate (a) and losing candidate (b), are compared in the following tableau:

---

[29] It should also be noted that interpretation (c) is not entirely unacceptable, for reasons related to expectations of chord progressions. The most common and fundamental harmonic progression in Western tonal music (which characterizes countless pop songs and folk tunes) is I–IV–V–I; this sequence of harmonies is likely so prevalent because it uses the most consonant chords and easily orients the listener to a diatonic tonal center (the chords together use all and only the notes of the diatonic scale). A listener may hear the IV^6 chord as structurally important because it is part of a larger, functional progression of I (the opening chord), IV (beats 1 and 3 of the first full measure), V (beat 2 of the second measure), and I (the final chord). The salience of this type of well-formed harmonic progression is akin to the LINSTABLE constraint on melodic lines and may be considered a constraint of its own in future studies of musical syntax.

Table 9: Tableau of candidate interpretations (a) and (b) from Bach's "O Haupt voll Blut und Wunden" from St. Matthew Passion, BWV 244

	CONSON	METPOS
☞ a. IV^6 V^6_5 I^x I ... 1 2	*	**
b. IV^6 V^6_5 I^x I ... 1 2	**!	*

Because candidate (a) assigns structure to the musical events occurring in the two metrically weakest

positions (indicated by the metrical grid), the candidate incurs two violations of METPOS. Candidate (b)

assigns structure to one metrically weak musical event, incurring one violation of METPOS. Both

candidates also incur violations of CONSON. Of the four chords that comprise the selected musical

statement, the final I chord is most consonant, followed by the more dissonant I^x (because an intrinsic

dissonance is created between the E and the F and between the D and the E). The tonic chords are

followed in consonance by the IV^6 chord and then the $V^{6/5}$ chord. Typically the IV and V sonorities are

approximately equal in tonal consonance (both have a TPS score of 5 as calculated in section 4.2.2); here

these chords are both in first inversion, but the $V^{6/5}$ chord in this excerpt has an added intrinsic dissonance

of a seventh, which makes it less consonant than the IV^6 chord.[30] Because both candidates assign structure

---

[30] Lerdahl and Jackendoff (1983) argue that the $V^{6/5}$ chord and the IV^6 chord are approximately equal in terms of consonance; however, according to basic principles of consonance, adding a seventh to a triad should actually make the chord inherently more dissonant. The V^7 chord contains a tritone between the seventh scale degree and the fourth scale degree, an interval that has traditionally been considered highly dissonant (and long ago was associated with evil and the devil). The fourth and seventh scale degrees beg resolution to the closest stable notes in the diatonic scale, the third and first scale degrees respectively. This is most likely because these resolutions require the notes to move only one chromatic step in the scale (whereas most notes in the diatonic scale are separated by two chromatic

to the V$^{6/5}$ over the more consonant chord in time-span 1 (the IV^6), both candidates incur one violation of CONSON. However, candidate (b) also assigns structure to the more dissonant I^x chord, so this candidate incurs a second violation of CONSON. Because we know that the constraint ranking must result in the selection of candidate (a) as the optimal interpretation, we can posit that CONSON seems to outrank METPOS.

The hypothesis that CONSON outranks METPOS is corroborated by countless examples in Western tonal music. While metrical position is certainly important to structural interpretations, consonance and stable relationships to a tonal center prove to be much more crucial, as these relationships are the very basis of the tonal system. Consider for example the structure of the first half of the first full measure of the passage in Figure 35 also from "O Haupt voll Blut und Wunden;" while the tonic I^6 chord is not in a metrically strong position, it is decidedly structural, especially in the context of the surrounding unstable minor (and diminished) chords:

Figure 35: Structural hierarchy of measure 1, beats 1 & 2 of Bach's "O Haupt voll Blut und Wunden" from St. Matthew Passion, BWV 244 (Lerdahl & Jackendoff, 1983)

---

steps). This force is analogous to magnetism: The shorter the distance between an unstable note and a stable note, the more the ear may crave that resolution. On the other hand, the V^7 chord is conventionally used almost interchangeably with the V chord in Western tonal music, and it can be argued that enculturated listeners have become quite tolerant of the V^7 dissonance, giving weight to the analysis of the V$^{6/5}$ and IV chords as approximately equal in consonance.

We see CONSON outrank METPOS again in the well-loved song from *The Muppets*, "The Rainbow Connection" (Williams & Ascher, 1979). This song consistently employs appoggiaturas and other non-chord tones on strong beats to delay the consonant (and structural) chords until the second (metrically weak) beat of the measure. Despite their metrically weak position, the consonance of these second-beat events causes them to be judged as unequivocally more structural:

Figure 36: Structural hierarchies in "The Rainbow Connection" (Williams & Ascher, 1979)

Returning to the "O Haupt voll Blut und Wunden" example, Figure 37 repeats the passage and Table 10 gives a tableau showing a comparison of the third candidate (candidate [c] in Figure 33) to winning candidate (a).
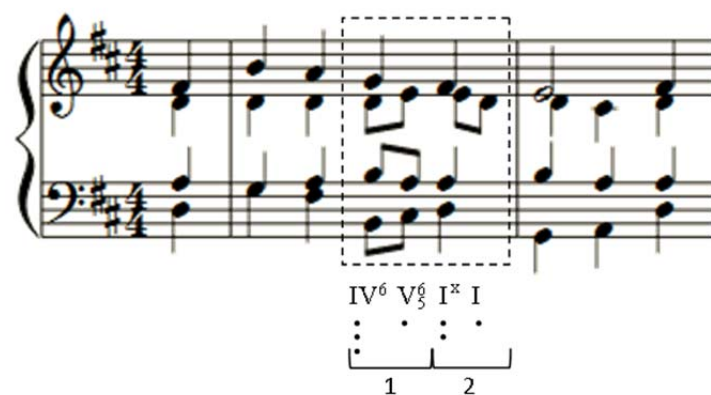

Figure 37 (reprint of Figure 32): Excerpt from Bach's "O Haupt voll Blut und Wunden" from St. Matthew Passion, BWV 244

Table 10: Tableau of candidate interpretations (a) and (c) from Bach's "O Haupt voll Blut und Wunden" from St. Matthew Passion, BWV 244

	LINSTABLE	CONSON	METPOS
☞ a. IV⁶ V₅⁶ Iˣ I 1 2		*	**
c. IV⁶ V₅⁶ Iˣ I 1 2	*!	*	

We saw previously that candidate (a) incurs a violation of CONSON; candidate (c) also violates CONSON because it assigns structure to the I^x over the more consonant I in time-span 2, but since every structural head is in a metrically stable position, this candidate does not incur any violations of METPOS. However, here we might also consider the influence of a third constraint, LINSTABLE. At this point it is important to reiterate that evaluation of LINSTABLE only occurs when one head candidate creates a particularly salient and well-formed line with heads of neighboring time-spans and another does not. LINSTABLE was not implicated in the previous comparison of candidates (a) and (b) because both candidates contained the same melodic and bass notes. At this most local level of analysis, time-span 1 has two options for structural head: the IV^6 or the $V^{6/5}$. Of these two, the C# in the bass of the V^6 chord forms a more stable linear bass progression for two reasons: 1) The movement mirrors and "answers" the preceding bass notes, which move from G to F-sharp, creating symmetry and a call-and-response effect, and 2) The half-step movement from the leading tone to the tonic creates a satisfying and aesthetically pleasing tension-

resolution pattern. Because the C-sharp and D are part of a linearly stable bass line, the candidate that does not assign structure to a pitch-event containing the C-sharp, candidate (c), incurs a violation of LINSTABLE[BASS], and candidate (a) does not.

Based on these constraint interactions, we can at this point posit a partial ranking where LINSTABLE and CONSON both outrank METPOS, but we are still unsure as to the ranking of LINSTABLE and CONSON in relation to each other. The success of candidate (a) over candidate (b) in the first comparison indicated that CONSON must outrank METPOS, a ranking that is corroborated by (or at least not contradicted by) the success of candidate (a) over candidate (c). In the second comparison we saw that candidates (a) and (c) each incurred one violation of CONSON; since these violations are approximately equal, this constraint does not play a role in the selection of an optimal interpretation. We do know from this comparison that LINSTABLE seems to outrank METPOS, because candidate (c)'s violation of LINSTABLE results in the selection of candidate (a) as optimal. However, both rankings CONSON >> LINSTABLE and LINSTABLE >> CONSON allow this selection to be possible.

Figure 38, an excerpt from Mozart's Sonata K. 331, shows a competition between candidates that may tell us more about the ranking of CONSON and LINSTABLE, especially when we consider how the parser selects the structural heads at the measure-level of time-span:
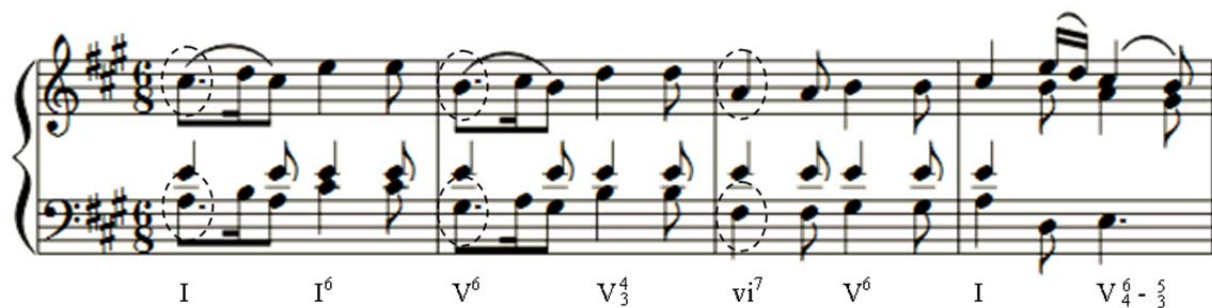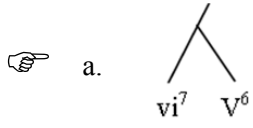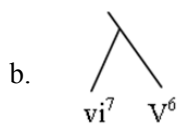

Figure 38: Excerpt from Mozart's Piano Sonata in A major, K. 331 (Adapted from Lerdahl & Jackendoff, 1983: 32)

In the third measure of this excerpt, the first chord of the measure, vi^7, is very clearly heard as most structural even though it is significantly less consonant than the second chord, V^6. Here we seem to have evidence that LINSTABLE outranks CONSON, but we also must acknowledge that both LINSTABLE[MEL] and LINSTABLE[BASS] are implicated here. The A on beat 1 of the third measure completes a very stable and rhythmic melodic descent from the C-sharp at the beginning of the passage; this line, a descent from the third to the second to the first scale degrees (circled in figure), is particularly well-formed and salient in Western tonal music.[31] Another stable linear sequence is created by the bass notes of the structural heads of these three measures, from A to G-sharp to F-sharp.

Table 11: Tableau of candidate interpretations of excerpt from Mozart's Piano Sonata in A major, K. 331

	LINSTABLE [MEL]	LINSTABLE [BASS]	CONSON
☞ a. vi^7 V^6			*
b. vi^7 V^6	*!	*!	

From this analysis, we may tentatively conclude that the LINSTABLE family of constraints outranks CONSON. However, another possibility may be that LINSTABLE[MEL] and LINSTABLE[BASS] have a cumulative effect. While conventional linguistic OT selects candidates based on strict constraint dominance and not gradient or cumulative effects, it seems reasonable to consider that the influence of

---

[31] One of the prevailing forms of music analysis, Schenkerian analysis, purports that this $\hat{3}$ - $\hat{2}$ - $\hat{1}$ melodic line is so fundamental that it can be found at the most abstracted, reduced level of all tonal compositions. This line is part of what Schenker calls the *Ursatz,* or "fundamental structure" of tonal music.

linear stability on structural interpretation will be greater when this stability occurs in both the melodic and bass lines than just one or the other.[32]

Clearly more research will be needed to elucidate a ranking between CONSON and LINSTABLE, as the relationship between these forces in interpreting musical structure is complex and multi-faceted. One complicating factor in this comparison is the gradient nature of both consonance and linear stability. We may conjecture that to an extent, consonance is a more important factor in structural interpretations than linear stability, if only because a certain degree of consonance is necessary for a composition to be understood and interpreted within the tonal system. The primacy of relatively consonant, functional harmony in Western tonal music contrasts with the optionality of melodic lines that are particularly well-formed or salient (in terms of symmetry, step-wise motion, or thematic significance). As an example, consider the excerpt from Webern's atonal piece, Movements for String Quartet, Op. 5, No. 3 given in Figure 39. I have added beams and changed the circled pitches from D and C to A and G respectively to emphasize a higher-level melodic line that could be heard as linearly salient (and therefore preferentially structural) in a tonal context. However, because this excerpt lacks any sense of tonal organization, it is difficult to argue that *any* musical event seems particularly structural.

Figure 39: Piano reduction of Webern's Movements for String Quartet, Op. 5, No. 2 (Adapted from Straus, 2005; circled notes have been modified from original score)

---

[32] Indeed, similar proposals have been made for linguistic OT to accommodate situations where two constraints can be locally conjoined as a more complex (and powerful) constraint (Smolensky, as cited in Crowhurst, 2011).

Another indication that consonance may ultimately dominate linear stability can be observed in many compositions written in the *theme and variation* format. Consider Mozart's Twelve Variations on "Ah, vous, dirais-je Maman," K. 265. It is clear that each of the variations maintains the overall structure of the original theme, even though some of the variations obscure the linear connections among the notes of the original melody (especially variations III and VIII and parts of variations VI and VII). As an example, consider the first eight measures of Variation III given in Figure 40:

Figure 40: Excerpt from Variation III of Mozart's Twelve Variations on "Ah, vous, dirais-je Maman," K. 265 (1778/1986)

While the constituents of the familiar melodic line can somewhat be identified in some measure of this excerpt, it is difficult to imagine that they would necessarily be selected as structural by a listener who is completely unfamiliar with the theme. Rather, the connection with the original theme seems to have more to do with the overall melodic contour (which is not necessarily linked to the LINSTABLE constraint) and the relative consonances of the successive harmonies comprising a familiar hierarchy of stability that parallels the original theme.

As a final note, we can see how constraint interaction results in different structural interpretations of almost identical melodic material using a partial analysis given by Lerdahl and Jackendoff (1983) of the *rondo* theme from Beethoven's Piano Sonata No. 8 in C minor, Op. 13 (*Sonata Pathetique*). Consider

the motivic phrases circled in Figure 41. In this passage, several constraints are implicated to determine the higher-level structural interpretation of parallel motivic elements. Ultimately the E-flat on beat 1 of measure 1 is determined to be the structural head of the passage; while the C on beat 1 of measure 2 is slightly more consonant, the E-flat is both in a significantly stronger metrical position (as the first strong beat of the entire passage) and it incurs fewer violations of ATTACH since it is the left-most viable head option. The D in this opening motive, while more structural than the eighth-note F and E-flat of the same measure by virtue of its metrical and linear stability, is less structural than either the E-flat on beat 1 of the same measure or the C on beat 1 of the following measure. Almost an opposite interpretation holds true for the parallel closing material, in which the CADENCE constraint takes over to elevate the structural positions of the D (which now comprises the highly structural cadential V sonority) and the C (which now comprises the highly consonant and structural resolution of the cadential V).
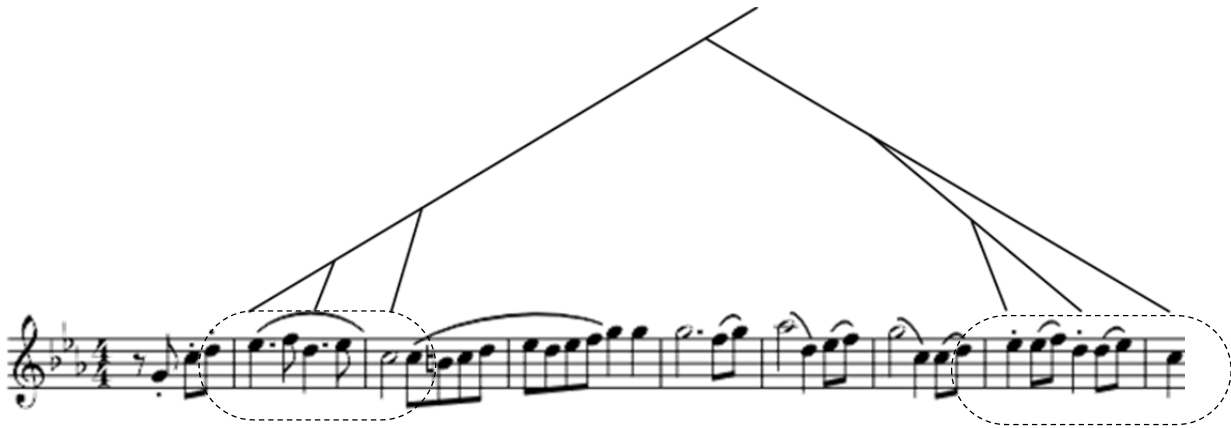
Figure 41: Partial structural hierarchy of *rondo* theme from Beethoven's Piano Sonata No. 8 in C minor, Op. 13 (*Sonata Pathetique*; Lerdahl & Jackendoff, 1983: 253)

4.6 CONCLUSIONS

The analyses presented thus far demonstrate some of the principles of constraint interaction and optimization in Western tonal music. In accordance with standard OT, I have attempted to characterize the ranking relationships among some of these constraints. The purpose of these analyses is to lay the groundwork for a formal OT account of musical syntax. Such an account will require further research and

data on the musical intuitions of enculturated listeners for the purposes of refining and better understanding the constraints on tonal syntax. This account will also require a decisive characterization of constraints to allow for optimization of candidate selection. While a partial constraint ranking has been suggested here, these rankings seem rather ad hoc and do not necessarily align with what we intuitively know to be true about musical listening. The discernment of musical structure as an automatic part of listening to music is a cognitively substantiated process, and this process appears to be influenced by a number of different constraints that can have gradient effects and may not always result in a single "correct" interpretation. As an alternative to the strictly ranked constraint system associated with traditional linguistic OT, the following chapter will explore the viability of application of a weighted constraint system for musical syntax that may better help us understand the constraint optimization process during musical parsing.

CHAPTER 5:

OPTIMALITY THEORY USING WEIGHTED CONSTRAINTS

5.1 RANKED CONSTRAINTS VS. WEIGHTED CONSTRAINTS

The analyses discussed in the previous chapter support the hypothesis that the perception and processing of Western tonal music proceeds in an optimality-theoretic manner. The grammar that an enculturated listener acquires merely through exposure to (but not necessarily formal instruction in) the tonal idiom allows the listener to formulate a hierarchical mental representation of the musical surface through syntactic processes of integration of incoming elements. The formation of these representations is constrained by violable principles, and the representation that emerges optimally adheres to these principles. As in language, the variable strength effects of the constraints are a significant factor in the optimization of the structural interpretation. For example, while it is undeniable that metrical position is important for assigning structural importance to a musical event, it is frequently overridden by issues of consonance or linear stability.

In language OT models, the variable strength effects are manifested in a definitive ranking of the proposed constraints in a hierarchy of dominance. In such a hierarchy, violation of a higher-ranked constraint disqualifies an output candidate if another candidate does not have violations of that or any other higher-ranked constraint. So if candidate A violates a highly ranked constraint but no lower-ranked constraints, and candidate B violates many lower-ranked constraints but not a highly-ranked constraint, then candidate B will win as the output form regardless of its total number of violations. The previous chapter explored how this winner-take-all approach to constraint evaluation applies to music-syntactic processing and found that, although a tentative ranking can be proposed for some of the constraints, such a ranking is tenuous and inelegant.

Under the proposal that language and music share a syntactic processing mechanism characterized by an optimality-theoretic parser, it is natural to question what implications the difficulty of ranking music-syntactic constraints presents for this hypothesis. Following the standard OT model, the musical analyses given in the previous chapter show that the musical parser selects a structural interpretation that optimally adheres to (and minimally violates) a set of constraints. However, unlike standard OT, the constraints do not consistently operate in a winner-take-all style of competition; instead, they seem to exhibit variable weighted influence on the selection of structural heads. The output is still optimal in the sense that the selected (or probable) interpretation is the one that incurs the least severe violations.

This deviation from the standard linguistic OT model may seem to challenge the current hypothesis for a shared syntactic processing mechanism; however, certain linguistic parsing phenomena suggest that language processing, like music, actually favors a system of weighted constraint competition rather than a strict dominance hierarchy. Supporting this hypothesis are gradient acceptability judgments of perfectly grammatical forms, which reveal that some low-ranked constraints remain visible to the human sentence parser even though they have no overt influence on the selection of the optimal surface form. Fanselow (2005) observes this phenomenon in wh-extraction from embedded clauses in English and German. English has a high-ranking constraint on wh-subject extraction from complement clauses (where the complementizer is expressed) that renders (26) ungrammatical; in contrast, this constraint is not highly ranked in German, so (27) is grammatical:

    (26)   *who do you think that ___ loves Mary

    (27)   wer   denkst du   dass   ___   die Wahl   gewinnen   wird
             who   think   you   that   ___   the election win   will
             "who do you think will win the election"

While (27) is certainly a grammatical construction in German that arguably presents little to no extra-grammatical processing challenges (e.g., working memory), Featherston (as cited in Fanselow, 2005) finds that native speakers consider this type of subject extraction to be *less acceptable* than similar object

extractions. This finding suggests that the extraction constraint restricting output forms in English but not in German is still influential in the parsing of German structures, even though it is technically outranked.

5.2   WEIGHTED CONSTRAINTS FOR LANGUAGE

Gibson and Broihier (1998) explore a number of other situations in which the constraint ranking component of Optimality Theoretic syntax falls short of accounting for sentence parsing phenomena. Consider, for example, the interaction of three well-established constraints on serial word integration (as occurs in sentence processing) that are responsible for *garden path* processing difficulties. The first two constraints are articulations of Chomsky's Theta Criterion: *Thematic Reception* (THEMREC) requires that every argument be assigned a theta-role, and *Lexical Requirement* (LEXREQ) refers to the obligatory fulfillment of lexically required nodes (i.e., ones that are hypothesized by the parser but not yet filled).[33] The third constraint, *Node Locality* (NODELOC), requires incoming lexical items to be attached in the current clause or phrase and incurs a violation for each phrase projection separating the point of attachment from the point of occurrence in the surface articulation.

Gibson and Broihier give a number of examples that suggest both of the theta constraints (THEMREC and LEXREQ) outrank NODELOC. Consider the sentence *Ron gave the letter to Nancy to the postman*. When the parser receives the first *to*, it has two options for attachment: within a PP argument of *gave* or within the NP containing *letter*. At this point in the parse, the PP argument is the preferred interpretation. Neither interpretation incurs violations of THEMREC, and both incur one violation of LEXREQ because the obligatory NP complement of *to* is empty. The within-NP interpretation incurs an additional LEXREQ violation, because at this point in the parse it leaves the required goal argument of *gave* unfilled. The PP argument interpretation incurs one NODELOC violation, because it passes over the local NP to attach higher in the structure to the VP. These violations are summarized in the following tableau:

---

[33] LEXREQ is a rearticulation of the Obligatory Head (OBHD) principle introduced in section 3.4.

Table 12: Tableau of candidate interpretations for *Ron gave the letter to [x]*

Ron gave the letter <u>to</u> *Nancy to the postman*	LEXREQ	NODELOC
☞ [IP [NP Ron] [VP [V gave] [NP [DET the] [N letter]] [PP to]]]	*	*
[IP [NP Ron] [VP [V gave] [NP [DET the] [N' [N letter] [PP to]]]]]	**!	

We can see in another example that THEMREC also outranks NODELOC. Consider the sentence *Arthur brought the dog the ball hit a bone,* wherein *the ball* can attach either at the VP as an argument or it can attach as the subject of a relative clause complement to *dog*. The preferred interpretation of *the ball* as an argument of the verb (which we know is preferred because the parser encounters a garden path effect at the word *hit*) incurs no THEMREC or LEXREQ violations but does incur one NODELOC violation because it bypasses the current phrase to attach higher at the VP. The relative clause interpretation, on the other hand, incurs no NODELOC violations but two THEMREC violations, because *the ball* and the non-lexical operator Op_i do not yet have a theta-role assigning element. These violations are summarized in the following tableau:

Table 13: Tableau of candidate interpretations for *Arthur brought the dog <u>the ball</u> [x]* (Gibson & Broihier, 1998)

Arthur brought the dog <u>the ball</u> *hit a bone*	LEXREQ	THEMREC	NODELOC
☞ [IP [NP Arthur] [VP [V brought] [NP the dog] [NP the ball]]]			*
[IP [NP Arthur] [VP [V brought] [NP [Det the] [N' [N dog$_i$] [CP [NP Op$_i$] [IP [NP the ball]]]]]]]		*!*	

While these examples do not elucidate a ranking of THEMREC and LEXREQ with regard to each other, they do show that both constraints outrank NODELOC.[34] However, this ranking is problematic when

---

[34] Although NODELOC has been consistently outranked at this point, we can see that it is still a necessary and relevant constraint. Consider the sentence *Bill said John left yesterday*. The parser attaches *yesterday* within the most local phrase, modifying *left*. This analysis incurs fewer NODELOC violations than if *yesterday* were attached to the main clause, modifying *said*. This effect is emphasized by changing the tense of the main clause and adverb to produce the somewhat unacceptable but still interpretable *Bill will say John left tomorrow*.

we consider situations in which NODELOC seems to outrank both of these theta-criterion constraints.

Consider the sentence *While I talked with the woman John was ignoring me at the party* in which the

parser initially interprets *John* as the subject of a relative clause modifying *woman* (as if the sentence

were to proceed *While I talked with the woman John was ignoring at the party I came to like her*), instead

of the correct interpretation of *John* as the subject of the main clause. This constraint ranking of

THEMREC above NODELOC incorrectly predicts that the parser will prefer the main clause subject

interpretation of *John.*[35]

Table 14: Tableau of candidate interpretations for *While I talked with the woman <u>John</u> [x]* (Gibson & Broihier, 1998)

While I talked with the woman <u>John</u> *was ignoring me at the party*	THEMREC	NODELOC
✂ [IP [CP [C While] [IP [NP I] [VP [V talked] [PP [P with [NP DET the] [N woman]]]]]] [IP [NP John]]]	*	**
☞ [IP [CP [C While] [IP [NP I] [VP [V talked] [PP [P with [NP DET the] [N' [N woman$_i$] [CP [NP Op$_i$] [C e] [IP [NP John]]]]]]]]]]	**	

An example from Spanish shows again the inadequacy of a winner-take-all approach. Spanish

differs from English in the preferred attachment site of a relative clause when there are two prospective

NPs the clause could modify. Consider the sentence *The journalist interviewed the daughter of the*

colonel [who had had the accident] (and its Spanish version *El periodista entrevisto a la hija del coronel*

[que tuvo el accidente]). Adhering to the NODELOC constraint, English prefers for the CP to attach to the

most local NP, *the colonel*; however, Spanish prefers attachment as close as possible to the head of the

predicate. This well-established constraint, which is relatively highly ranked in Spanish, is *Predicate*

Proximity (PREDPROX). According to a winner-take-all constraint ranking, it seems that in English

NODELOC outranks PREDPROX, while in Spanish PREDPROX outranks NODELOC. However, this ranking

---

[35] Gibson and Broihier do not address the effect that a prosodic pause (or a comma) would have on the interpretation of this sentence. It may be argued that the reason the parser has trouble with the relative clause interpretation of *John* is because a pause or comma is typically expected in this situation, so the violation of expectation is what causes the processing trouble. On the other hand, we might conclude that the pause or comma is added in this and similar situations to aid the parser, which would otherwise proceed with a ranking of THEMREC above NODELOC.

becomes problematic when we consider Spanish sentences that have three potential NP attachment sites. Gibson, Perlmutter, Canseco-Gonzalez, and Hickok (1996) measured processing difficulties for phrases similar to those in Table 15, wherein the intended attachment site is disambiguated using number agreement, by collecting data on reading times of the different types of attachments:

Table 15: Reading time comparisons for varied NP attachment sites in Spanish constructions like *the lamp(s) near the painting(s) of the house(s) that was damaged in the flood* (Gibson & Broihier, 1998)

Attachment	Sentence	Reading Time
a. Low (most local)	[NP$_1$ las lámparas cerca de [NP$_2$ las pinturas de [NP$_3$ la casa] [CP que fue dañada en la inundación]]]	Fastest
b. Middle	[NP$_1$ las lámparas cerca de [NP$_2$ la pintura de [NP$_3$ las casas] [CP que fue dañada en la inundación]]]	Slowest
c. High	[NP$_1$ la lámpara cerca de [NP$_2$ las pinturas de [NP$_3$ las casas] [CP que fue dañada en la inundación]]]	Middle

Contrary to expectation, the reading times showed that the most local attachment site was the easiest to process for Spanish speakers, followed by the highest attachment site and then the middle attachment site. These findings suggests that in ambiguous cases wherein the parser might select any one of three possible attachment sites, the most local site may be favored contrary to what would be predicted by the winner-take-all constraint ranking system.

The examples presented here model just a few situations wherein a strict constraint dominance ranking does not appear sufficient to explain linguistic grammatical phenomena. A number of other modifications of Optimality Theory have been proposed to explain similar phenomena. Sorace and Keller (as cited in Keller, 2006) identify a need to classify syntactic constraints as hard and soft; violations of soft constraints may only trigger mild unacceptability judgments, have context-dependent effects, and are subject to optionality during language acquisition, whereas violations of hard constraints are associated with absolute ungrammaticality, are context-immune, and are developmentally stable during acquisition. These characterizations are extended from Linear Optimality Theory, devised by Keller (2001), which proposes that gradient degrees of acceptability of syntactic structures follow from cumulative effects of

constraint violations. Yet another theoretical approach to reconciling the apparent gradient effects of constraints in syntactic processing is Stochastic Optimality Theory (Boersma, Escudero, & Hayes, 2003) which assumes each constraint has a range of possible ranking values along a continuous scale, and the precise ranking of the constraint at the time of evaluation (*EVAL*) is probabilistic rather than definitive. This approach allows constraints to be *likely* to outrank other constraints but does not require them to do so. While Stochastic OT is methodologically and philosophically different from a weighted constraint system, the approach helps support the notion that strict constraint ranking does not always seem to appropriately characterize linguistic phenomena.

Gibson and Broihier (1998) propose a relative weighting system to account for these problematic constraint interactions. In the model proposed, the precise weight value proposed for each constraint is not as important as the consistency and accuracy of the resulting structural selections. This accuracy can be achieved for the given examples by associating a weight of one PLU (Processing Load Unit, terminology used by Gibson & Broihier) with each violation of THEMREC and LEXREQ, and a slightly higher PLU of 1.5 with each violation of NODELOC. Consider the first example we examined, repeated in Table 16 with the added constraint of PREDPROX (which, for reasons of simplicity, was not considered in the initial example since it is outranked by NODELOC and does factor into a winner-take-all analysis). As expected, the preferred interpretation incurs a weight of 2.5 PLUs, while the other candidate incurs a weight of 3 PLUs.

Table 16: Weighted constraint tableau of candidate interpretations for *Ron gave the letter to [x]* (including PREDPROX) (Gibson & Broihier, 1998)

Ron gave the letter to *Nancy to the postman*	LEXREQ	NODELOC	THEMREC	PREDPROX	TOTAL
☞ [IP [NP Ron] [VP [V gave] [NP [DET the] [N letter]] [PP to]]]	1	1.5			2.5
[IP [NP Ron] [VP [V gave] [NP [DET the] [N' [N letter] [PP to]]]]]	2			1	3

The second example *Arthur brought the dog the ball hit a bone*, repeated here in Table 17 with the PREDPROX constraint included, can also be understood in terms of constraint weights. Interestingly, in this case the difference in PLUs between the two candidates is significantly greater than in the previous example, a phenomenon that reflects the difference in the garden path effects between the two. In the previous example, the reanalysis required by the parser once the phrase *to the postman* is presented seems intuitively less severe than the reanalysis required by the parser in the following example when it must incorporate *hit a bone*. The second reanalysis may even require two or three readings, or the speaker/writer may be likely to rephrase the sentence entirely to avoid the reanalysis effects.

Table 17: Weighted constraint tableau of candidate interpretations for *Arthur brought the dog the ball [x]* (including PREDPROX) (Gibson & Broihier, 1998)

Arthur brought the dog the ball *hit a bone*	LEXREQ	NODELOC	THEMREC	PREDPROX	TOTAL
☞ [IP [NP Arthur] [VP [V brought] [NP the dog] [NP the ball]]]		1.5			1.5
[IP [NP Arthur] [VP [V brought] [NP [Det the] [N' [N dog$_i$] [CP [NP Op$_i$] [IP [NP the ball]]]]]]]			2	1	3

We saw previously that the garden path sentence *While I talked with the woman John was ignoring me at the party* presents a problem for a winner-take-all constraint ranking, but under the weighted constraint system the analysis is not problematic. The preferred interpretation of *John* as the subject of a relative clause modifying *woman* has a significantly lower PLU score than the second candidate parse. This analysis is exhibited in Table 18.

Table 18: Weighted constraint tableau of candidate interpretations for *While I talked with the woman John [x]* (Gibson & Broihier, 1998)

While I talked with the woman John *was ignoring me at the party*	THEMREC	NODELOC	TOTAL
☞ [IP [CP [C While] [IP [NP I] [VP [V talked] [PP [P with [NP DET the] [N' [N woman$_i$] [CP [NP Op$_i$] [C e] [IP [NP John]]]]]]]]]]	2		2
[IP [CP [C While] [IP [NP I] [VP [V talked] [PP [P with [NP DET the] [N woman]]]]] [IP [NP John]]]	1	3	4

Finally, a weighted constraint system can also account for the problematic Spanish example of three-site attachment ambiguities. Assuming that the PREDPROX constraint has a higher weight in Spanish than in English, we might associate violation of PREDPROX in Spanish with 1.5 PLUs and violation of NODELOC with 1 PLU. Unlike the other constraints discussed here, PREDPROX violation does not seem to be cumulative with increasing distance from the node that is maximally proximate to the predicate. In other words, attachment to a site that is not as close as possible to the predicate incurs 1.5 PLUs, regardless of how many other potential attachment sites intervene.[36]

Table 19: Reading time comparisons using weighted constraints for varied NP attachment sites in Spanish constructions like *the lamp(s) near the painting(s) of the house(s) that was damaged in the flood* (Gibson & Broihier, 1998)

Attachment Site	PREDPROX	NODELOC	TOTAL	Reading Time of Disambiguated Version
a. Low (N_3)	1.5		1.5	Fastest
b. Middle (N_2)	1.5	1	2.5	Slowest
c. High (N_1)		2	2	Middle

In conjunction with the NODELOC violations, these constraint weights correctly predict the preferences of the parser, assigning a low processing score for the low attachment and the highest processing score for the middle attachment.

5.3   WEIGHTED CONSTRAINTS FOR MUSIC

If a weighted constraint OT more elegantly accounts for syntactic parsing phenomena than the traditional winner-take-all approach, it stands to reason that musical parsing may also be explained using weighted constraints. In the previous chapter we saw that CONSON violations seem to have significant sway over the selection of structural heads, but we also saw situations in which the effects of CONSON could be overpowered by violations of constraints that might otherwise seem subordinate. On the other hand, we

---

[36] Gibson et al. (1996) give more details about this conception of the PREDPROX constraint.

saw that while REGEXT and METPOS certainly impact judgments of structural importance, these effects are rather weak compared to other constraints. LINSTABLE seems to fall somewhere in between, serving as a fairly significant predictor of structural importance, but likely not quite as powerful as CONSON. Following these tendencies, I propose the constraint weights listed in Table 20. As is the case for linguistic models, the precise values given to the weights here are not as important as their success in capturing the differential effects each constraint has on the selection of an optimal structural interpretation. Gibson and Broihier's "Processing Load Unit (PLU)" term does not seem appropriate here, so I will refer to these weights as Structural Preference Units (SPUs).[37] Unlike Gibson and Broihier's PLUs, music-syntactic SPUs are often calculated based on how a musical event compares to the other events in its time-span; therefore, the SPU for an event in one time-span may be different than its SPU in a higher time-span.

Table 20: Proposed music-syntactic constraint weights

Constraint	SPU
CONSON	2
LINSTABLE	1.5
REGEXT	1
METPOS	0.5
ATTACH	Negligible

A structural interpretation of the segment from "O Haupt voll Blut und Wunden" that was analyzed in the previous chapter (reprinted here in Figure 42) can be better derived and understood in terms of weighted constraints. The tableau in Table 21 shows how each constraint violation is calculated as an SPU value contributing to the candidate's overall violation score. While this constraint weight analysis selects the same candidate structure as our initial constraint ranking, it also reflects the intuition

---

[37] A better term may be Structural *Dispreference* Units, since an interpretation candidate that incurs more of these units is less preferred as a structural head than a candidate that incurs fewer units, but I will use SPU in the interest of avoiding a neologism.

that the other candidate structures are gradiently less acceptable interpretations of the input: Candidate (c) is still an acceptable interpretation, most likely because the IV^6 can also easily be heard as structural (as mentioned in section 4.5), followed by (b), which is a decidedly unlikely interpretation.
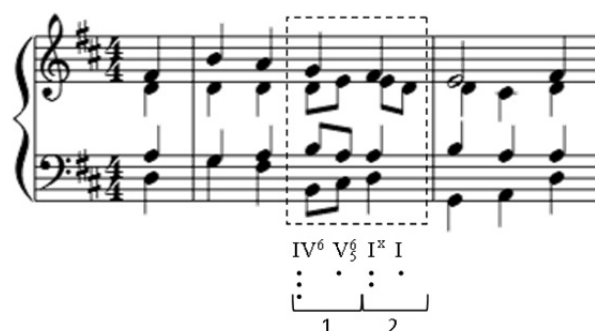
Figure 42 (reprint of Figure 32): Excerpt from Bach's "O Haupt voll Blut und Wunden" from St. Matthew Passion, BWV 244

Table 21: Weighted constraint tableau of candidate interpretations from Bach's "O Haupt voll Blut und Wunden" from St. Matthew Passion, BWV 244

		CONSON	LINSTABLE	METPOS	TOTAL
☞	a.	2*1=2		0.5*2 = 1	3
	b.	2*2=4		0.5*2=1	5
	c.	2*1=2	1.5*1=1.5		3.5

Consider now the excerpt from Mozart's Sonata K. 331, in which violation of the powerful CONSON constraint is permitted by the satisfaction of many lower-ranked constraints. According to a winner-take-all OT model, violation of CONSON would likely require candidate (b) to be selected as optimal, regardless of the number or severity of its lower-ranked constraint violations. However, in this case it seems the violations of several lower-ranked constraints have conspired to outweigh the violation of a stronger constraint, resulting in the selection of candidate (a) as the optimal interpretation.
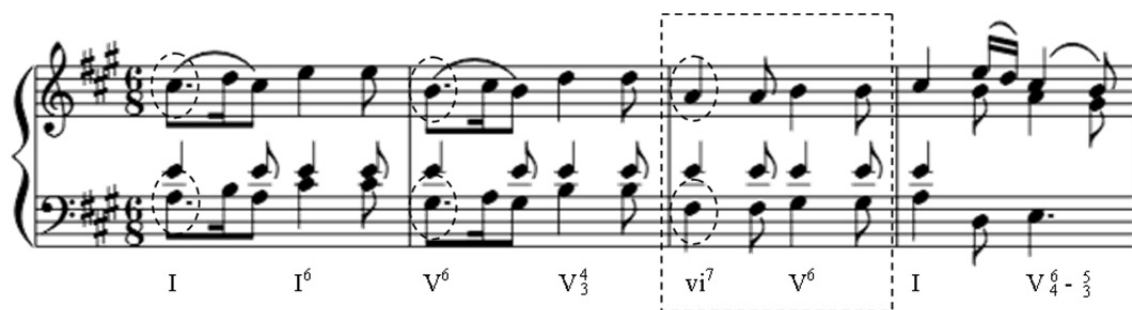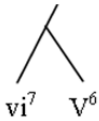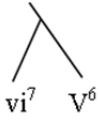
Figure 43 (reprint of Figure 25): Mozart's Piano Sonata in A major, K. 331 (Adapted from Lerdahl & Jackendoff, 1983: 32)

Table 22: Weighted constraint tableau of candidate interpretations for segment of Mozart's Piano Sonata in A major, K. 331

	CONSON	LINSTABLE [MEL]	LINSTABLE [BASS]	METPOS	TOTAL
☞ a.	2	2	2	2	2
b.		1.5	1.5	0.5	3.5

The notion of constraint weights also helps to elucidate and fortify the proposed ATTACH constraint. As is made clear by the copious examples that do not assign headedness to every left-most element of a time-

106

span, the effect of ATTACH is easily overridden by other constraints during local parsing. However, these effects do surface at higher levels of time-span, evidenced by the tendency of a listener to hear the beginnings of overall phrases as increasingly structural as these phrases are perceived in grander contexts, until, at the highest level, the very first event of a piece is likely to be interpreted as one of the most structural events, second only to the final cadence. Adopting the notion of weighted constraints allows violations at local levels to be minimal, but as many of the constraints are neutralized at higher levels of abstraction, we see more of a tendency for left-most elements to be interpreted as structurally important.

Consider the opening chords of our familiar example, "O Haupt voll Blut und Wunden," given in Figure 44:
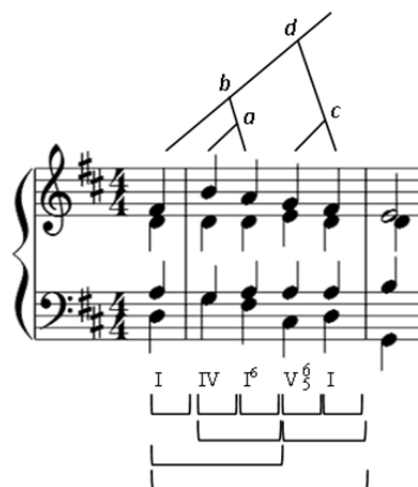
Figure 44: Structural hierarchy of measure 1 of Bach's "O Haupt voll Blut und Wunden" from St. Matthew Passion, BWV 244

The first time-span requires a head selection between the IV chord on beat 1 and the I^6 chord on beat 2 (point *a*). The I^6 chord wins the structural head assignment with an SPU score of 0.5 for violation of METPOS, while the selection of the less consonant IV chord would have incurred an SPU score of 2 for a CONSON violation. At the next highest level of the augmented time span (point *b*), the I^6 loses to the opening I chord with a CONSON violation. Point *c* selects a head of the time-span immediately containing beats 3 and 4 (which have been reduced here per previous analysis), selecting the I chord on beat 4 as

head despite its violation of METPOS because it is more consonant than the $V^{6/5}$ on beat 3. At this point in the analysis we have not seen ATTACH have much of an influence on structural head assignment, but consider what happens at the next level of time-span (point *d*), when the opening I chord and the I chord on beat 4 of measure 1 compete for structural head. The chords are in identical metrical positions, have identical pitch content, are equally good linearly, and do not have registral extremes. At this point in the parse, the fact that the left-most element was encountered first by the parser elevates it to a higher point of perceptual salience, which in this case influences judgment of structure, and the opening I chord is selected by ATTACH as more structural.

5.4 CONCLUSIONS

While it is clear that more analysis will be necessary to formalize an OT for musical syntax that uses weighted rather than strictly ranked constraints, the discussions presented in this chapter are important because they provide support for what may be a fundamental feature of the syntactic integration processes for both language and music. While there is considerable research showing that constraints likely exist in a strict dominance hierarchy in OT models of linguistic competence, there is a growing body of evidence suggesting parsing processes may translate these rankings into weights, allowing even low-ranked constraints to have an influence on linguistic performance. Such a proposal is reasonable considering the fact that the online language processing mechanism likely uses whatever resources it can to quickly resolve ambiguities and conflicts it encounters. Given the phenomena discussed in this chapter, I assert that future investigations of optimality-theoretic parsing in language and music will be most fruitful if both domains are presumed to use weighted rather than strictly ranked constraints during online processing.

CHAPTER 6:

CONCLUSIONS & FUTURE AVENUES OF RESEARCH

6.1   SUMMARY AND CONTRIBUTIONS OF THE CURRENT APPROACH

The body of research on music-syntactic processing has shown that musical processing tasks that are

specifically syntactic activate areas of the brain that have been previously supposed to be exclusively

dedicated to language. These findings are corroborated by behavioral studies showing a distinct overlap in

processing resources between syntax-specific linguistic and musical tasks, warranting for many

researchers a reevaluation of the current conception of the language faculty as encapsulated from other

cognitive domains. Patel (2008) emphasizes the distinction between the representational knowledge base

of a domain (i.e. linguistic and musical competence) and the syntactic processes that are employed during

the access and integration of this knowledge, proposing that the processing resources are shared between

the two domains while the knowledge bases remain separate. While Patel's hypothesis has received much

attention and praise for capturing an empirically valid claim about the cognitive relationship between

language and music, there have been very few attempts to directly characterize the syntactic integration

resource shared by the two domains.[38] While this inquiry may be outside the realm of linguistic

competence, I endorse Jackendoff's (2011) desideratum for linguistic theory to be gracefully integrated

with what we know about the rest of the mind/brain and argue that a theoretical account of the

connections between music and language is an important contribution to our understanding of language as

a cognitive domain. I propose that the shared processing resource for linguistic and musical syntax is

---

[38] I am aware of only two other attempts. Katz and Pesetsky (2011) propose the *Identity Thesis for Language and Music,* hypothesizing that language and music are identical (and characterized by binary Merge) in all respects except for their fundamental building blocks. Slevc and Okada (2014) suggest that music- and linguistic-syntactic processes both invoke *cognitive control* (see section 6.2).

characterized predominantly by the constraint evaluation and optimization mechanisms of Optimality Theory.

OT was originally formulated for phonology, and its successful adaptation for both syntactic competence and performance phenomena suggests that the notions of violable constraints and optimization may be a more general characteristic of the language faculty, or even general cognition. The notion of extending OT to other cognitive domains is not at all at odds with the theory's origination as a model of general neural networking and computations. Following models of OT parsing for linguistic syntax, I argue that the process of musical parsing involves the integration of new material into an optimal structural interpretation conditioned by violable constraints. The analysis presented in this paper is based on Lerdahl and Jackendoff's *Generative Theory of Tonal Music* (1983), which formalizes a number of rules for constructing a tree representation of the hierarchy of structural importance assigned by a listener to the events of a musical composition.

While the OT approach for musical syntax presented here is far from complete, it is important to delineate the ways in which my approach may constitute an advancement in our conceptions of the theoretical and cognitive links between linguistic and musical syntax. This approach to musical syntax differs from that of its predecessor, GTTM, in a number of important ways. Lerdahl and Jackendoff explicitly and emphatically state that GTTM is *not* a processing theory; rather, GTTM's rules were formulated for the development of reductions that represent the final state of a listener's understanding of a musical composition. The rule system proposed in GTTM is intended to model linguistic generative theory, but the authors are extremely cautious about suggesting that these formal similarities can be indicative of any significant cognitive connection between the two domains. In contrast to this view, I support the notion that a theoretical approach is valuable to the extent that it models and represents a cognitive reality, be it representational knowledge (competence) or mental processes (performance). Optimality Theory, while originally proposed as a theory of linguistic competence, is well-suited for modeling linguistic processing, and my analyses show how an OT of musical syntax can also account for parsing phenomena during musical listening. Additionally, while GTTM gives examples of how the

preference rules interact and sometimes conflict with one another, the authors stop short of proposing a ranking or dominance hierarchy for the rules. By contrast, the weighted constraint system is a more realistic and comprehensive method of modeling the optimality theoretic processes involved in Western tonal musical syntax. In turn, the success of future refinements to an OT for musical syntax may provide valuable insight into the viability of a weighted constraint OT for linguistic syntax or even other cognitive domains.

Another way the current analysis advances the model proposed by GTTM is in its ability to better accommodate gradient judgments of grammaticality. GTTM states that their preference rules elucidate a likely structural interpretation, i.e., one that most listeners will attribute to a musical surface, but they do not address how or why other interpretations may be possible. Furthermore, the authors of GTTM cannot (and explicitly do not want to) account for issues of ungrammaticality, or what happens when the parser receives an input that it finds very difficult or impossible to parse. The weighted constraint system advanced in this paper may be a significant step toward better understanding notions of ungrammaticality in both music and language. This is a complicated issue for even linguists and one our understanding of musical syntax is certainly not advanced enough to explain at this time, but I contend that a weighted constraint approach to musical syntax is certainly a significant step in this direction. In the following section, I will suggest an adjustment to the current approach that may constitute a more elegant account of how an optimality theoretic parser handles notions of ungrammaticality.

6.2 EXTENSIONS AND FUTURE AVENUES OF RESEARCH

The current approach will clearly require elaborations and refinements to constitute a thorough and elegant theory of musical syntax. As mentioned previously, one such refinement will be to better account for processing difficulties encountered by the musical parser when, for instance, it encounters an out-of-key chord. Future research may conclude that, under an optimality-theoretic system of weighted constraints, ungrammaticality represents a threshold where the PLU (or SPU) violations are so severe they inhibit processing. This approach is certainly feasible in music, where most instances of

"ungrammaticality" involve a note or chord that is *dramatically* (not just somewhat) out-of-key. This hypothesis will require an adjustment to the weighted constraint approach proposed in Chapter 5, since the analysis asserts that each CONSON violations receives an SPU score of 2, regardless of the severity of the violation (or consonance of the losing candidate) in question. Such an adjustment, where the variable consonance of a losing candidate translates into higher or lower SPU scores, can be facilitated in part by incorporating actual Tonal Pitch Space (TPS) scores into the SPU score. According to the TPS theory, out-of-key chords will have a higher TPS score, and under this variation of a weighted constraint system, should result into a more severe violation of CONSON and higher SPU score. Many, possibly all, of the other constraints may also be interpreted to have variable weights consistent with the severity of the violation; for example, violation of METPOS may be more or less severe depending on how significantly the two competing candidates differ in the strength of their metrical position.

While this adjustment may be satisfactory in aligning online processing difficulties with severity of constraint violations for candidate structural events, there is one aspect of parsing that it does not directly address: *expectation*. The role of expectation and prediction in cognitive processes is quite significant; it is well-known that expectation and prediction condition every facet of our interactions with and understanding of the world around us. The field of linguistic study has recently seen a burgeoning of theoretical approaches that place more emphasis on stochastic knowledge and processes than the abstract rule systems of Chomskyan formalisms (Chater & Manning, 2006). As we saw in section 3.4, optimality-theoretic parsing models rely on listeners' expectations about incoming elements while parsing syntactically ambiguous structures. These expectations emerge from a language user's implicit knowledge of constraint rankings and, in turn, are what allow the OT researcher to deduce a ranking. However, it is less clear how expectations in tonal music can be incorporated into the OT model for musical syntax. Making a similar argument to the ones I have presented here, Slevc and Okada (2014) claim that listening to music "involves building up complex cognitive representations of musical structure over time. This involves not only the incremental processing and integration of musical elements as they occur, but also the incremental generation of musical predictions and expectations" (Ambiguity and Cognitive Control in

Musical Structure section, para. 1). Slevc and Okada identify the shared processing resources as mechanisms of *cognitive control*, proposing that these resources are invoked when new information cannot be integrated as expected and must be reinterpreted in the context of an alternate structure. An OT model is a reasonable way of conceiving of the processes that integrate elements and interpret structure in musical parsing, but clearly more research is needed to understand how the model can account for the influence of music-syntactic expectancies.

An absolutely critical next step in the development of an OT for musical syntax is the extension of the analyses conducted here to musical systems other than Western tonal music. An intermediary goal of such an analysis will be to refine (or rework) the proposed constraints and identify additional constraints that are implicated in other musical systems. The researcher may find that other systems do not lend themselves to the same notions of structural interpretation as does Western tonal music, and it will be important to determine what this means for the current hypothesis. At this point in its inception, the current proposal of an OT for musical syntax is extremely limited in its focus on Western tonal music; however, this decision was made out of necessity rather than ethnocentricity. At this time there is relatively little work that directly investigates enculturated listeners' perceptions of structure in Western tonal music and even less on those in other musical systems. The result is a limited body of literature in which ubiquitous assumptions are made about other musical systems based on knowledge about the Western tonal system. As researchers become increasingly interested in music cognition, investigations of many different types of musical systems will help advance our understanding of optimality-theoretic processes that may be at play in these systems. Mirroring the desiderata of linguistic theory, one of the ultimate goals of music cognitive theory should be to determine whether there are *musical universals*, or properties of tonal organization and processing that extend across all musical systems. If we do find such properties in musical systems across the world, we might conclude that they represent or reflect some innate truths about the human music faculty. Further comparisons between language and music may find universal properties shared across systems in both domains, providing us with a deeper understanding of how these domains fit into a greater cognitive context.

Another avenue of future research will be to explore other processes that may be involved in musical listening and whether these processes also use more generally cognitive (or perhaps specifically linguistic) mechanisms. The structural interpretation processes described in this paper are a key component of music cognition, but they are only one aspect of the musical experience. Furthermore, linguistic and musical knowledge and processes certainly do not exist in isolation from one another, although methodologically it is often helpful to conceive of them this way. The field of linguistic study is increasingly interested in understanding the interactions among the many different modules responsible for all linguistic knowledge and behavior, from the interface between phonology, syntax, and semantics, to the utilization of a language user's linguistic competence during actual, real-time language production and perception. Music cognition, in particular music listening, also seems to invoke parallel, interfacing processes. For example, although it was not discussed in detail in this paper, the perception of metrical structure may be an optimality theoretic process in which the listener uses melodic and harmonic information to interpret an optimal hierarchy of strong and weak beats. Metrical structure was posited here as a constraint on interpretations of musical structure, but it is likely that musical structure also acts as a constraint on metrical structure, creating an interface between the two systems. Similarly, the current discussions did not address in detail the automatic process an enculturated listener employs to identify a tonal center.[39] Like the identification of a metrical structure, this process occurs in conjunction with the interpretation of musical structure, and future research may elucidate how these processes interact.

6.3 CONCLUDING REMARKS

Strained comparisons and over-romanticized notions of "music as a universal language" may indeed be an old and futile game, but this concern has not stopped researchers from making serious and focused strides in our understanding of how music and language may be connected in the mind. Modern linguistic theory

---

[39] A number of different theories and algorithms have been proposed to model the process of key finding, but little consensus has been reached. For example, see the rare interval theory proposed by Browne (1981); intervallic rivalry theory proposed by Butler (1989); and the self-organizing map neural network model using iterative key-finding algorithms developed by Krumhansl and Toivanen, 2001.

has focused on issues of competence while pushing aside issues of performance as peripheral to characterizations of the language faculty. I argue that to the extent that research on linguistic performance contributes to our understanding of language as a human faculty, it should be considered valuable to and integrated with the enterprise of linguistic theory. The observed connections between our complex and uniquely human capacities for language and music provide a remarkable opportunity to better understand these capacities and how they fit into a broader cognitive context. To this end, this thesis has explored one way of modeling these connections using a theoretical framework, Optimality Theory, that can successfully account for other linguistic phenomena, and possibly other cognitive phenomena as well. While an optimality-theoretic approach to musical syntax is still in its infancy, it is my hope that this inquiry will inspire further investigation, grounded both in empirical research and sound theory, into the intriguing cognitive relationship between these two domains.

REFERENCES

Ball, P. (2010). *The music instinct: How music works and why we can't do without it*. London: Bodley Head.

Bharucha, J. and Krumhansl, C. L. (1983). The representation of harmonic structure in music: Hierarchies of stability as a function of context. *Cognition, 13,* 63-102.

Boersma, P., Escudero, P., and Hayes, R. (2003). Learning abstract phonological from auditory phonetic categories: An integrated model for the acquisition of language-specific sound categories. In *Proceedings of the 15th international congress of phonetic sciences*. Barcelona.

Brown, S. (2000). The "musilanguage" model of music evolution. In N. L. Wallin, B. Merker, and S. Brown (Eds), *The Origins of Music* (pp. 271-301). Cambridge, MA: The MIT Press.

Browne, R. (1981) Tonal implications of the diatonic set. *In Theory Only,* 5 (6-7), 3-21.

Burns, E. M. and Ward, W. D. (1982). Intervals, scales, and tuning. In D. Deutsch (Ed.), *Psychology of music* (pp. 241-269). New York, NY: Academic Press, Inc.

Butler, D. (1989). Describing the perception of tonality in music: A critique of the tonal hierarchy theory and a proposal for a theory of intervallic rivalry. *Music Perception: An Interdisciplinary Journal, 6* (3), 219-241.

Cadwallader, A. C. and Gagné, D. (1998). *Analysis of tonal music: A Schenkerian approach*. New York: Oxford University Press.

Chater, N. and Manning, C. D. (2006). Probabilistic models of language processing and acquisition. *TRENDS in Cognitive Sciences, 10* (7), 335-344.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge: The MIT Press.

Crowhurst, M. J. (2011). Constraint conjunction. In M. Oostendorp, C. J. Ewan, E. V. Hume, and K. Rice (Eds.), *The Blackwell Companion to Phonology,* (Vol. 3, pp. 1461-1490). Malden, MA: Wiley-Blackwell.

Darwin, C. (1871). *The descent of man and selection in relation to sex* (2nd ed.). New York, NY: A. L. Fowle.

De Backer, W. (2011). Somebody that I used to know. On *Making Mirrors,* Melbourne: Eleven Music.

Dowling, J. (1982). Melodic information processing and its development. In D. Deutsch (Ed.), *Psychology of Music* (pp. 413-429). New York, NY: Academic Press, Inc.

Fanselow, G. (2005). *Some predictions of Optimality Theory on Sentence Processing, Grammaticality, and Corpus Frequency*. Unpublished manuscript. Accessed at http://www.ling.uni-potsdam.de/cgi-fanselow/index.py?site=%20e_publications

Fitch, W. T. (2012). The biology and evolution of rhythm: Unravelling a paradox. In P. Rebuschat, M. Rohrmeier, J. A. Hawkins, and I. Cross (Eds)., *Language and Music as Cognitive Systems* (pp. 73-95). New York, NY: Oxford University Press.

Fodor, J. A. (1983). *Modularity of mind: An essay on faculty psychology.* Cambridge, MA: The MIT Press.

Gibson, E. and Broihier, K. (1998). Optimality Theory and human sentence processing. In P. Barbosa, D. Fox, P. Hagstrom, M. McGinnis, and D. Pesetsky (Eds.), *Is the best good enough?: Optimality and competition in syntax* (pp. 157-191). Cambridge, MA: The MIT Press.

Gibson, E., Pearlmutter, N., Canseco-Gonzalez, E., and Hickok, G. (1996). Recency preference in the human sentence processing mechanism. *Cognition, 59,* 23-59.

Hauser, M. D., Chomsky, N., and Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science, 298* (5598), 1569-1579.

Holdcroft, D. (1991). *Saussure: Signs, systems, and arbitrariness.* Cambridge, MA: Cambridge University Press.

Jackendoff, R. (2011). What is the human language faculty? Two views. *Language, 87* (3), 586-624.

Jagger, M. and Richards, K. (1965). I can't get no satisfaction. On *Out of Our Heads*. London: 45-LON 9766.

Jentschke, S., Koelsch, S., Sallat, S., and Friederici, A. D. (2008). Children with Specific Language Impairment also show impairment of music-syntactic processing. *Journal of Cognitive Neuroscience, 20* (11), 1940-1951.

Katz, J. and Pesetsky, D. (2011). The Identity Thesis for language and music. Retrieved from http://ling.auf.net/lingbuzz/000959

Keller, F. (2001). *Gradience in grammar: Experimental and computational aspects of degrees of grammaticality* (Doctoral dissertation). Retrieved from British Library EThOS. (ethos.016265255)

Keller, F. (2006) Linear Optimality Theory as a Model of Gradience in Grammar. In G. Fanselow, C. Fery, R. Vogel, and M. Schlesewsky (Eds.), *Gradience in Grammar: Generative Perspectives* (pp. 270-288). Oxford, NY: Oxford University Press.

Koelsch, S., Grossmann, T., Gunter, T. C., Hahne, A., Schröger, E., and Friederici, A. D. (2003). Children processing music: Electric brain responses reveal musical competence and gender differences. *Journal of Cognitive Neuroscience, 15* (5), 683-693.

Koelsch, S., Gunter, T. C., von Cramon, D. Y., Zysset, S., Lohmann, G., and Friederici, A. D. (2002). Bach speaks: A cortical "language-network" serves the processing of music. *NeuroImage, 17* (2)*, 956-966.

Koelsch, S., Gunter, T. C., Wittfoth, M., and Sammler, D. (2005). Interaction between syntax processing in language and in music: An ERP study. *Journal of Cognitive Science, 17* (10), 1565-1577.

Krumhansl, C. L., Bharucha, J. J., and Kessler, E. J. (1982). Perceived harmonic structure of chords in three related musical keys. *Journal of Experimental Psychology: Human Perception and Performance, 8* (1), 24-36.

Krumhansl, C. L. and Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological review, 89* (4), 334-368.

Krumhansl, C.L. and Toiviainen, P. (2001). Tonal cognition. In R.J. Zatorre, and I. Peretz (Eds.), *The Biological Foundations of Music*. Annals of the New York Academy of Sciences (Vol. 930, pp. 77–91).

Larson, S. (2012). *Musical forces: Motion, metaphor, and meaning in music*. Bloomington: Indiana University Press.

Lerdahl, F. (2001). *Tonal pitch space*. New York, NY: Oxford University Press.

Lerdahl, F. and Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: The MIT Press.

Lévi-Strauss, C. (1963). *Structural Anthropology*. (C. Jacobson and B. G. Schoepf, Trans.) New York, NY: Doubleday Anchor Books. (Original work published 1958)

MacKay, J. S. (2003). Franz Joseph Haydn and the five-octave classical keyboard: Registral extremes, formal emphases and tonal strategies. *Canadian Music Review, 23* (1-2), 126-144.

Maess, B., Koelsch, S., Gunter, T. C., and Friederici, A. D. (2001). Musical syntax is processed in Broca's area: An MEG study. *Nature Neuroscience, 4* (5), 540-545.

Mozart, W. A. (1986). *Twelve variations on "Ah, vous dirai-je, Maman,* K. 300e (265). Van Nuys, CA: Alfred Publishing Co., Inc. (Original work published 1778)

Osterhout, L. and Holcomb, P. J. (1992). Event-related potentials elicited by syntactic anomaly. *Journal of Memory and Language, 31* (6)*,* 785-806.

Osterhout, L. and Holcomb, P. J. (1993). Event-related potential and syntactic anomaly: Evidence of anomaly detection during the perception of continuous speech. *Language and Cognitive Processes, 8* (4)*,* 413-437.

Osterhout, L., McKinnon, R., Bersick, M., and Corey, V. (1996). On the language specificity of the brain response to syntactic anomalies: Is the syntactic positive shift a member of the P300 family? *Journal of Cognitive Neuroscience, 8* (6)*,* 507-526.

Patel, A. D. (2008). *Music, language, and the brain*. New York, NY: Oxford University Press.

Patel, A. D., Gibson, E., Ratner, J., Besson, M., and Holcomb, P. J. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience, 10* (6)*,* 717-733.

Picton, T. W. (1992). The P300 wave of the human event-related potential. *Journal of Clinical Neurophysiology, 9* (4)*,* 456-479.

Pinker, S. (1997). *How the mind works*. New York, NY: Norton.

Prince, A. and, Smolensky, P. (1993). *Optimality Theory: Constraint interaction in generative grammar* (Technical Report TR-2). Rutgers Center for Cognitive Science, Rutgers University, New Brunswick, N.J.

Prince, A. and Smolensky, P. (1997). Optimality: From neural networks to universal grammar. *Science*, *275,* (5306), 1604-1610. http://www.jstor.org/stable/2893709

Reimer, B. and Evans, E. G. (1972). *The experience of music.* Englewood Cliffs, NJ: Prentice-Hall.

Shepard, R. N. (1982). Structural representations of musical pitch. In D. Deutsch (Ed.), *Psychology of music* (pp. 344-390). New York, NY: Academic Press, Inc.

Slevc, L. R. and Okada, B. M. (2014). Processing structure in language and music: A case for shared reliance on cognitive control. *Psychonomic Bulletin & Review,* doi:10.3758/s13423-014-0712-4

Slevc, L. R., Rosenberg, J. C., and Patel, A. D. (2008). Language, music, and modularity: Evidence for shared processing of linguistic and musical syntax. Proceedings from the 10th International Conference on Music Perception and Cognition. Adelaide: Causal Productions.

Sloboda, J. A. (1985). *The musical mind: The cognitive psychology of music*. Oxford: Clarendon Press.

Sloboda, J. A. (2005). *Exploring the musical mind.* New York, NY: Oxford University Press, Inc.

Smolensky, P., Legendre, G., and Tesar, B. (2006). Optimality Theory: The structure, use, and acquisition of grammatical knowledge.In P. Smolensky and G. Legendre (Eds)., *The harmonic mind: From neural computation to Optimality-Theoretic grammar* (Vol. 1, pp. 453-544). Cambridge, MA: The MIT Press.

Stevenson, S. and Smolensky, P. (2006). Optimality in sentence processing. In P. Smolensky and G. Legendre (Eds)., *The harmonic mind: From neural computation to Optimality-Theoretic grammar* (Vol. 2, pp. 307-338). Cambridge, MA: The MIT Press.

Straus, J. N. (2005). *Introduction to post-tonal theory* (3rd ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Temperley, D. (2001). *The cognition of basic musical structures.* Cambridge, MA: The MIT Press.

Traditional (2000). *The Farmer in the Dell*. Milwaukee, WI: Hal Leonard Corporation.

Trainor, L. J. and Trehub, S. E. (1994). Key membership and implied harmony in Western tonal music: Developmental perspectives. *Perception and Psychophysics, 56,* 125-132.

Van der Werf, S. and Hendriks, P. (2004). A constraint-based approach to grouping in language and music. In R. Parncutt, A. Kessler, and F. Zimmer (Eds.). Proceedings from CIM '04: *The Conference on Interdisciplinary Musicology*. Graz, Austria.

Williams, P. and Ascher, K. (1979). Rainbow Connection. On *Fuzzy Muppet Songs*, US: Atlantic.